

UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Channel Coding Techniques for Network Communication

Permalink

<https://escholarship.org/uc/item/8219c43q>

Author

Wang, Lele

Publication Date

2015

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Channel Coding Techniques for Network Communication

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Electrical Engineering
(Communication Theory and Systems)

by

Lele Wang

Committee in charge:

Professor Young-Han Kim, Chair
Professor Philip E. Gill
Professor Alon Orlitsky
Professor Paul H. Siegel
Professor Alexander Vardy

2015

Copyright
Lele Wang, 2015
All rights reserved.

The dissertation of Lele Wang is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2015

DEDICATION

*To my parents
whose love and support
made this dissertation possible*

EPIGRAPH

*He that wrestles with us strengthens our nerves
and sharpens our skills. Our antagonist is our helper.*

—Edmund Burke

TABLE OF CONTENTS

	Signature Page	iii
	Dedication	iv
	Epigraph	v
	Table of Contents	vi
	List of Figures	ix
	List of Tables	xi
	Acknowledgements	xii
	Vita	xv
	Abstract of the Dissertation	xvii
Chapter 1	Introduction	1
Chapter 2	Universal Channel Coding	5
	2.1 Introduction	5
	2.2 Method	7
	2.2.1 Universal polar coding	15
	2.2.2 Rate	16
	2.2.3 Error probability	16
	2.2.4 Complexity	16
	2.3 Codes with Arbitrary Rates	17
	2.3.1 Polar coding	24
	2.3.2 Rate	25
	2.3.3 Error probability	25
	2.3.4 Complexity	25
	2.4 Discussion	25
Chapter 3	Interference Channels: Polar Codes	28
	3.1 Introduction	28
	3.2 Insufficiency of Single-Block Rate-Splitting	30
	3.3 Proof of Theorem 3.2.1	35
	3.4 Polar Coding Preliminaries	45
	3.4.1 Aligning Polarized Indices	45
	3.4.2 ‘Polar Splitting’ for MACs	49
	3.5 Two-user Compound MAC	51
	3.5.1 Uniform Independent Inputs	51
	3.5.2 Arbitrary Inputs	56
	3.5.3 Main Result	57

	3.6	Interference Networks	58
	3.6.1	‘Polar Splitting’ for K -User MAC	60
	3.6.2	Aligning Polarized Indices for K Users	64
	3.6.3	Han–Kobayashi Inner Bound	64
	3.7	Discussion	66
Chapter 4		Interference Channels: COTS Codes	68
	4.1	Sliding-Window Superposition Coding	69
	4.1.1	Corner Point: The First Illustration	69
	4.1.2	General Rate Point: Single Dimensional SWSC	76
	4.2	Variations for Practical Purpose	81
	4.2.1	Rate Loss	81
	4.2.2	Homogeneous Superposition Coding	82
	4.2.3	Decoding Orders	82
	4.2.4	Superposition Layers	84
	4.3	Sliding-Window Coded Modulation	87
	4.3.1	Theoretical Performance Comparison	88
	4.3.2	Implementation With LTE Turbo Codes	89
	4.3.3	Simulation Results	90
	4.4	Han–Kobayashi Inner Bound	90
	4.4.1	Two-dimensional SWSC	91
	4.4.2	SWSC achieves the Han–Kobayashi Inner Bound	98
	4.5	Discussion	103
Chapter 5		Broadcast Channels	105
	5.1	Introduction	105
	5.2	Rate Regions for the Two-Receiver BC	107
	5.2.1	Homogeneous Superposition Coding (UV Scheme)	108
	5.2.2	Heterogeneous Superposition Coding (UX Scheme)	110
	5.3	Main Result	112
	5.4	Polar Coding for Cover’s Superposition Coding	114
	5.5	Discussion	116
	5.6	Appendix: Optimality of the Rate Region in (5.4)	117
Chapter 6		Relay Channels	120
	6.1	Introduction	120
	6.2	Decode–Forward Relaying	123
	6.2.1	Polarization for Broadcasting	124
	6.2.2	Coding Scheme	128
	6.3	Compress–Forward Relaying	130
	6.3.1	Polarization for the Wyner–Ziv coding	131
	6.3.2	Coding Scheme	131
	6.4	Discussion	135

Chapter 7	Channel Coding and Slepian–Wolf Coding	136
	7.1 Introduction	136
	7.1.1 Channel Coding Problem	136
	7.1.2 Slepian–Wolf Problem	137
	7.1.3 Background	138
	7.1.4 Contributions	139
	7.2 Linear Code Duality	140
	7.2.1 Symmetric Channel Code to Symmetric Slepian–Wolf Code	140
	7.2.2 Symmetric Slepian–Wolf Code to General Slepian–Wolf Code	145
	7.2.3 Symmetric Slepian–Wolf Code to Symmetric Channel Code	148
	7.2.4 Symmetric Channel Code to General Channel Code . .	150
Chapter 8	Concluding Remarks	154
Appendix A	Polarization Preserves Less Noisy Ordering	156
Appendix B	A Sufficient Condition for Less Noisy Ordering	161
Bibliography	164

LIST OF FIGURES

Figure 2.1: When X_1 and X_2 are uniform and independent, the channel $U_1 \rightarrow Y_1^2$ is equivalent to $L_1 = W^-$ and $U_2 \rightarrow Y_1^2 U_1$ is equivalent to $R_1 = W^+$. 10

Figure 2.2: The channels $U_i \rightarrow Y_1^4 U_1^{i-1}$ are equivalent to those inside the parentheses. 10

Figure 2.3: Four pairs of level-1 channels are chained to create six level-2 and two level-1 channels. The channels $U_i \rightarrow Y_1^8 U_1^{i-1}$ are equivalent to the ones on the left. 10

Figure 2.4: A 3-level transform with $K = 4$ 11

Figure 2.5: The $(n + 1)$ -level construction. Here, $\ell = 2^{n-1} - 1$ 12

Figure 2.6: A one-level transform that combines $b+g = 6$ channels with $g = 2$ and $b = 4$. Only the labels of the channels at the corresponding locations are shown. 18

Figure 2.7: A rate-2/6 transform (left) and a rate-4/6 transform (right). Channels enter both transforms on the right-hand-side and produce the channels on the left-hand-side. 19

Figure 2.8: An example of polarization. 22

Figure 3.1: Two-user interference channel. 29

Figure 3.2: Symmetric Gaussian interference channel. 31

Figure 3.3: Alignment of the incompatible indices 48

Figure 3.4: Two MAC regions with equal sum-rate. 51

Figure 3.5: First recursion. 52

Figure 3.6: Second recursion. Here, $c' = 2n + c, d' = 2n + d, e' = n + e$, and $f' = n + f$ 53

Figure 3.7: Two MAC regions with unequal sum-rates. 55

Figure 3.8: K -sender L -receiver interference networks. 59

Figure 3.9: Han–Kobayashi coding scheme. 64

Figure 4.1: The simultaneous decoding inner bound and the corner point (4.1) (the one with the arrow) that will be illustrated to be achievable by SWSC. 69

Figure 4.2: Simultaneous decoding inner bound formed by two multiple access rate regions with unequal sum-rates. 80

Figure 4.3: The two-user Gaussian interference channel. 87

Figure 4.4: Performance comparison in the symmetric Gaussian interference channel. 91

Figure 4.5: Message scheduling for the two-dimensional SWSC. 93

Figure 4.6: The dependency in recovering all messages at receiver 1. 93

Figure 4.7: The dependency graph corresponds to layering order s_1 97

Figure 4.8: The dependency graph corresponds to layering order s_2 98

Figure 4.9: Han–Kobayashi coding scheme. 99

Figure 4.10: Two-dimensional message scheduling for L_2 and L_3 102

Figure 5.1:	Superposition codebooks for which (a) the structure within each cloud is identical and (b) the structure is nonidentical between clouds. Code-words (dots) are annotated by “ m_1, m_2 ”, where m_1 is the coarse layer message and m_2 is the fine layer message.	107
Figure 5.2:	Two-receiver broadcast channel.	108
Figure 5.3:	Rate regions for the proof of Theorem 5.3.1.	112
Figure 5.4:	Illustration of Cover’s superposition coding as coding for two-user interference channels.	115
Figure 6.1:	Three-node relay channel.	121
Figure 7.1:	A summary of the main results. SW is short for Slepian–Wolf coding and CC is short for channel coding.	139
Figure 7.2:	A channel code for symmetric BMC $p(y x)$	140
Figure 7.3:	The codify block. Left: The block diagram. Right: Illustration of a shift by $\bar{H}X^n$ in $\{0, 1\}^n$ space.	141
Figure 7.4:	The noisy codify block. Left: The block diagram. Right: Illustration of a shift by $\bar{H}X^n$ in \mathcal{Y}^n space.	141
Figure 7.5:	Relations of the random variables (X^n, Y^n, C^n, R^n)	144
Figure 7.6:	The construction of a symmetric Slepian–Wolf code from a symmetric channel code.	144
Figure 7.7:	A code for the symmetrized Slepian–Wolf problem $p(\tilde{x}, \tilde{y})$	146
Figure 7.8:	The construction of a general Slepian–Wolf code from a symmetric Slepian–Wolf code.	147
Figure 7.9:	A code for symmetric Slepian–Wolf problem $p(x, y)$	148
Figure 7.10:	The construction of a symmetric channel coding from a symmetric Slepian–Wolf code.	149
Figure 7.11:	Channel coding for symmetric BMC $p(\tilde{y}, \tilde{z} \tilde{x})$	151
Figure 7.12:	The construction of a general channel coding from a symmetric channel code.	151
Figure A.1:	Dependence graph of the random variables in (A.2).	157
Figure A.2:	Dependence graph of the random variables in (A.4).	160
Figure B.1:	Lorenz curves for a BSC, a BEC, and a generic BMS channel with the same value at $u = 1$	163

LIST OF TABLES

Table 2.1:	Bounds on $I(R_n)$ for $I(W) = 0.5$	15
Table 2.2:	Bounds on $I(R_n)$ for $I(W) = 0.8$	15
Table 4.1:	Sliding-window superposition coding scheme.	70
Table 4.2:	SWSC scheme with decoding rule 1.	78
Table 4.3:	Sliding-window superposition encoding for (U, X_1) and (V, X_2)	84
Table 5.1:	Rate regions for homogeneous and heterogeneous superposition coding.	110

ACKNOWLEDGEMENTS

Working towards a PhD with my advisor Young-Han Kim was quite an experience. As a teacher, he revealed the most beautiful insights in network information theory and magically smoothed out my transition from course work to research. As a researcher, he transformed seemingly unsolvable research problems into fun math puzzles, through which he exemplified how to think clearly, critically, creatively, and confidently. As a mentor, he constantly challenged me and forced me to confront my weaknesses. His desire for me to excel exceeded even mine. I will never forget the step-by-step derivations he walked me through in my first project, the numerous emails we exchanged late nights, the seven dry runs we practiced together for the Qualcomm Innovation Fellowship, and the two-month full time writing he led me through for a two-page research statement. Young-Han, thank you for being such a devoted teacher, brilliant thinker, and genuine mentor. You have propelled me to reach far beyond what I thought I was ever capable of accomplishing. For that, I am deeply grateful.

The past six years wouldn't have been so valuable, were it not for the help, encouragement, and support from many teachers and mentors. My sincere gratitude goes to Paul H. Siegel for his excellent lectures on coding theory, his strong support whenever I was in need, and his kind invitation to join his family thanksgiving gatherings. I would like to thank Alexander Vardy for sharing his insights on polar codes and his detailed suggestions on my work. I must thank Rüdiger Urbanke, who has been a nurturing mentor since the summer of 2013 when he kindly invited me to EPFL. The stimulating discussions on modern coding theory and the thought-provoking advices on personal growth are my most precious memories of Lausanne. I am indebted to Erdal Arıkan, for his elegant invention of polar codes that inspired the work presented here and for his interest and encouragement in my career. I am grateful to my Qualcomm Innovation Fellowship mentor Joseph Soriaga, who provided feedback to my work from an industry engineer's perspective. I would like to thank Chandra Nair, Igal Sason, Emina Soljanin,

and Emre Telatar for many interesting conversations during conferences and for their generous encouragement for an academic career. I would also like to thank Bruce Driver, Massimo Franceschetti, Philip E. Gill, Alon Orlitsky, the late Jack K. Wolf, and many other teachers in Math and ECE departments for their wonderful classroom lectures.

The interaction with past and present ITA postdocs have benefited me a great deal. Our discussions from mathematical puzzles to academic anecdotes to positive psychology have left me with many memorable moments in my grad life. I would like to thank Bernd Bandemer, Sreechakra Goparaju, Sudeep Kamath, Yonatan Kaspi, Anand Sarwate, Eren Şaşıoğlu, Ofer Shayevitz, Ido Tal, and Himanshu Tyagi for the intellectual stimulations they brought to Young-Han’s group meetings and for the everyday company and emotional support. I would particularly like to thank Eren for being a motivational collaborator and trustworthy friend. His style on academic presentation and writing has influenced me a lot.

Life would have been less cheerful without the company of many friends. I would like to thank Jayadev Acharya, Fatemeh Arbabjolfaei, Chiao-Yi Chen, Lorenzo Coviello, Hirakendu Das, Nicholas Genise, David Ho, Yuzhe Jin, Jing Li, Taehyung J. Lim, Enning Luo, Paolo Minero, Mohammad Naghshvar, Bāxtiyar Neyman, Shenjun Pan, Minghai Qin, Sankeerth Rao, Ananda Theertha Suresh, Yu Xiang, Eitan Yaakobi, and Jun Zhou for making UCSD a home away from home. I would like to thank all IPG friends in Lausanne—Mine Alsan, Marc Desgroseilliers, Rafah El-Khatib, Hamed S. Hassani, Mohammad Karzand, Wei Liu, Marco Mondelli, and Rajai Nasser—for making EPFL my second home. I would particularly like to thank Weina Wang and Yuhua Chen for participating in my joys and sorrows ever since we met ten years ago in Tsinghua.

Most importantly, for trying to understand my research purely out of their love for me, for attempting to learn English purely out of their care for my life, and for backing my journey regardless of the peaks and valleys, I thank my parents. It is to them I dedicate this dissertation.

Chapter 2 and Appendix A are, in part, a reprint of the material in the paper: Eren Şaşoğlu and Lele Wang, “Universal polarization,” accepted for publication in *IEEE Transactions on Information Theory*, 2015. Chapter 3 is, in part, a reprint of the material in the papers: Lele Wang and Eren Şaşoğlu, “Polar coding for interference networks,” submitted to *IEEE Transactions on Information Theory*, 2014; and Lele Wang, Eren Şaşoğlu, and Young-Han Kim, “Sliding-window superposition coding for interference networks,” to be submitted to *IEEE Transactions on Information Theory*. Chapter 4 is, in part, a reprint of the material in the papers: Lele Wang, Eren Şaşoğlu, and Young-Han Kim, “Sliding-window superposition coding for interference networks,” to be submitted to *IEEE Transactions on Information Theory*; and Hosung Park, Young-Han Kim, and Lele Wang, “Interference management via sliding-window superposition coding,” *Proceedings of IEEE GLOBECOM*, Austin, TX, December 2014. Chapter 5 is, in part, a reprint of the material in the papers: Lele Wang, Eren Şaşoğlu, Bernd Bandemer, and Young-Han Kim, “A comparison of superposition coding schemes,” *Proceedings of the IEEE International Symposium on Information Theory*, Istanbul, Turkey, July 2013; and Lele Wang and Eren Şaşoğlu, “Polar coding for interference networks,” submitted to *IEEE Transactions on Information Theory*, 2014. Chapter 6 is, in part, a reprint of the material in the paper: Lele Wang, “Polar coding for relay channels,” *Proceedings of the IEEE International Symposium on Information Theory*, Hong Kong, China, June 2015. Chapter 7 is, in part, a reprint of the material in the paper: Lele Wang and Young-Han Kim, “Linear code duality between channel coding and Slepian–Wolf coding,” *Proceedings of the 53th Annual Allerton Conference on Communication, Control, and Computation*, Monticello, Illinois, September 2015.

VITA

- 2009 B.S. in Electrical Engineering, Tsinghua University
- 2015 Ph.D. in Electrical Engineering (Communication Theory and Systems), University of California, San Diego

PUBLICATIONS

Eren Şaşođlu and Lele Wang, “Universal polarization,” accepted for publication in *IEEE Transactions on Information Theory*, 2015.

Lele Wang and Mohammad Naghshvar, “On the capacity of the noncausal relay channel,” accepted for publication in *IEEE Transactions on Information Theory*, 2015.

Lele Wang and Eren Şaşođlu, “Polar coding for interference networks,” submitted to *IEEE Transactions on Information Theory*, 2014.

Lele Wang, Eren Şaşođlu, and Young-Han Kim, “Sliding-window superposition coding for interference networks,” to be submitted to *IEEE Transactions on Information Theory*.

Chen Chi, Yu Zhang, and Lele Wang, “Joint power control and FEC unequal error protection for scalable H.264 video transmission over wireless fading channels,” *Proceedings of the 28th IEEE GLOBECOM*, Honolulu, Hawaii, December 2009.

Lele Wang and Young-Han Kim, “Sum-capacity of multiple-write noisy memory,” *Proceedings of the IEEE International Symposium on Information Theory*, Saint Petersburg, Russia, August 2011.

Lele Wang and Mohammad Naghshvar, “On the capacity of the noncausal relay channel,” *Proceedings of the IEEE International Symposium on Information Theory*, Saint Petersburg, Russia, August 2011.

Yu Xiang, Lele Wang, and Young-Han Kim, “Information flooding,” *Proceedings of the 49th Annual Allerton Conference on Communication, Control, and Computation*, Monticello, Illinois, September 2011.

Lele Wang, Minghai Qin, Eitan Yaakobi, Young-Han Kim, and Paul H. Siegel, “WOM with retained messages,” *Proceedings of the IEEE International Symposium on Information Theory*, Cambridge, MA, July 2012.

Lele Wang, Eren Şaşođlu, Bernd Bandemer, and Young-Han Kim, “A comparison of superposition coding schemes,” *Proceedings of the IEEE International Symposium on Information Theory*, Istanbul, Turkey, July 2013.

Fatemeh Arbabjolfaei, Bernd Bandemer, Young-Han Kim, Eren Şaşođlu, and Lele Wang, “On the capacity region for index coding,” *Proceedings of the IEEE International Symposium on Information Theory*, Istanbul, Turkey, July 2013.

Eren Şaşođlu and Lele Wang, “Universal polarization,” *Proceedings of the IEEE International Symposium on Information Theory*, Honolulu, HI, July 2014.

Lele Wang and Eren Şaşođlu, “Polar coding for interference networks,” *Proceedings of the IEEE International Symposium on Information Theory*, Honolulu, HI, July 2014.

Lele Wang, Eren Şaşođlu, and Young-Han Kim, “Sliding-window superposition coding for interference networks,” *Proceedings of the IEEE International Symposium on Information Theory*, Honolulu, HI, July 2014.

Hosung Park, Young-Han Kim, and Lele Wang, “Interference management via sliding-window superposition coding,” *Proceedings of IEEE GLOBECOM*, Austin, TX, December 2014.

Lele Wang, “Polar coding for relay channels,” *Proceedings of the IEEE International Symposium on Information Theory*, Hong Kong, China, June 2015.

Lele Wang and Young-Han Kim, “Linear code duality between channel coding and Slepian–Wolf coding,” *Proceedings of the 53th Annual Allerton Conference on Communication, Control, and Computation*, Monticello, Illinois, September 2015.

ABSTRACT OF THE DISSERTATION

Channel Coding Techniques for Network Communication

by

Lele Wang

Doctor of Philosophy in Electrical Engineering
(Communication Theory and Systems)

University of California, San Diego, 2015

Professor Young-Han Kim, Chair

Next-generation wireless networks aim to enable order-of-magnitude increases in connectivity, capacity, and speed. Such a goal can be achieved in part by utilizing larger frequency bandwidth or by deploying denser base stations. As the number of wireless devices is exploding, however, it is inevitable that multiple devices communicate over the same time and same spectrum. Consequently, improving the spectral efficiency in wireless networks with multiple senders and receivers becomes the key challenge. This dissertation investigates low-complexity channel coding techniques that implement canonical random coding schemes in network information theory, such as universal channel coding, superposition coding, rate-splitting, successive cancellation, simultaneous decod-

ing, decode-forward relaying, compress-forward relaying, and Slepian–Wolf coding. In representative communication scenarios, such as compound channels, interference channels, broadcast channels, and relay channels, the proposed channel coding techniques achieve the best known information theoretic performance, some utilizing the recently invented polar codes and some making use of the commercial off-the shelf codes, e.g., turbo and LDPC codes. These techniques have a potential to become important building blocks towards a general theory of channel coding techniques for the next-generation high-spectral-efficiency, low-power, broad-coverage wireless communication.

Chapter 1

Introduction

In a network communication system, multiple senders and multiple receivers communicate over shared noisy medium. Each sender wishes to transmit a message reliably to its desired receiver(s) by mapping the message into an input sequence. Each receiver recovers its either the whole or the part of the desired message(s) by mapping the output sequence to the message estimate(s). The goal for channel coding is to design encoding and decoding mappings such that the rates of the messages are high and the probability of error is small. For practical purposes, it is desired that the two mappings can be computed at low complexity.

Ever since Shannon's ground-breaking 1948 paper [69] that established the fundamental limit on reliable communication between one sender and one receiver, finding low-complexity capacity-achieving channel coding techniques has been the theme of coding theory. In the past sixty years, the point-to-point channel coding techniques evolved from algebraic codes (such as Hamming, Golay, Reed–Muller, Bose–Chaudhuri–Hocquenghem, and Reed–Solomon codes) to probabilistic codes (such as low-density parity-check, turbo, and fountain codes). In 2001, the capacity of the additive white Gaussian noise channel was shown to be achieved within 0.0045dB by low-density parity-check codes [20]. Most recently, the capacity of the binary input symmetric output channel was shown to be

achieved by polar codes [7] in 2009 and spatially coupled codes [47] in 2012. In other words, existing channel coding techniques have approached the capacity of point-to-point communication both in theory and in practice.

Unlike in the point-to-point case, existing low-complexity channel coding techniques are far from sufficient for network communication systems. Straightforward extensions of point-to-point coding techniques typically result in a large gap to the theoretically guaranteed performance, except for a few special cases such as successive cancellation in multiple access channels and superposition coding in single antenna Gaussian broadcast channels. When multiple messages are involved in communication, more advanced encoding and decoding schemes, which typically result in much higher computational complexity, are needed to achieve the best known theoretical performance.

Meanwhile, from the 2.4 kbps analog voice phone in 1980 to the 100,000 kbps fast mobile broadband in 2010, the mobile industry made a leap every ten years in the past four decades. In the next five to ten years (around 2020), the mobile industry aims to connect everything and to have a networked society. This fundamental overhaul in communication technology and infrastructure calls for order-of-magnitude increases in connectivity, speed, and mobile data volume. While part of the goal can be accomplished by exploiting broader frequency band or deploying smaller cells, there are limitations in scaling in the two directions. The key challenge in boosting the performance in the next generation communication systems lies in achieving higher spectral efficiency.

As a result of the theoretical curiosity and the practical need from the mobile industry, this dissertation investigates low-complexity channel coding techniques that implement canonical random coding schemes in network information theory, such as universal channel coding, superposition coding, rate-splitting, successive cancellation, simultaneous decoding, decode-forward relaying, compress-forward relaying, and Slepian–Wolf coding.

We first study the *universality* of code design, which requires a single code to perform reliably over multiple heterogeneous channels. Despite all the nice properties

such as achieving capacity, low encoding/decoding/construction complexities, no error floor, among others, the original polar code design by Arikan does not have universality. In Chapter 2, we develop a low-complexity polar coding technique that has universality. Our construction is independent of underlying channel statistics and achieves the optimal rate in compound channels.

As the ever-growing number of wireless devices is driving a denser and denser network deployment, efficient interference management among multiple parties communicating over the same spectrum becomes the key challenge for future communication networks. Existing interference management techniques either avoid interference by coordinating among the senders to transmit in orthogonal time/frequency/space dimensions, or ignore interference by treating it as part of the noise. However, the theoretically optimal approach is *simultaneous decoding*, which exploits the digital structures of both the desired and interfering signals. Due to its high computational complexity, simultaneous decoding has never been implemented in practice. In Chapters 3 and 4, we proposed two solutions to implement simultaneous decoding at low complexity, first with polar codes in Chapter 3 and then with commercial off-the-shelf (COTS) codes in Chapter 4. Using 4G LTE turbo codes, our simulation shows that the proposed scheme outperforms the existing scheme that treats interference as noise in both strong and weak interference regimes.

Superposition coding is one of the canonical coding schemes for broadcasting. There are two variants of superposition coding schemes. Cover's original superposition coding scheme [22] has code clouds of the identical shape, while Bergmans's superposition coding scheme [10] has code clouds of independently generated shapes. These two schemes yield identical achievable rate regions in several scenarios, such as the capacity region for degraded broadcast channels. In Chapter 5 we show that under the optimal maximum likelihood decoding, these two superposition coding schemes can result in different rate regions. In particular, it is shown that for the two-receiver broadcast channel, Cover's superposition coding scheme can achieve rates strictly larger than Bergmans's

scheme. Based on this fact, we then propose a polar coding scheme that achieves the rate region given by Cover's superposition coding.

Decode-forward and *compress-forward* are the two fundamental coding schemes for the relay channels. However, existing techniques in current wireless systems only allow multi-hop relaying that performs far worse than the decode-forward and compress-forward lower bounds. Chapter 6 investigates the low-complexity implementation of decode-forward and compress-forward with polar codes. Using the universal polarization technique in Chapter 2, the proposed scheme strictly improves upon existing methods and achieves the best known theoretical performance.

As a long term goal, we hope to develop a low-complexity version of any random coding scheme in network information theory. Instead of evaluating one particular code and its encoder and decoder at a time, we hope to develop a universal scheme that translates the performance of the COTS codes that are well studied and simulated in one communication scenario into the performance of codes for another communication scenario. Chapter 7 takes a first step in this direction and establishes the linear code duality between channel coding and Slepian-Wolf coding.

Chapter 8 makes concluding remarks and comments on future directions.

Chapter 2

Universal Channel Coding

A method to polarize channels universally is introduced. The method is based on combining channels of unequal capacities in each polarization step, as opposed to the standard method of combining identical channels. The locations of the good and bad channels that emerge upon polarization are only a function of the polar transform chosen, and are otherwise independent of the channel being polarized. This yields a simple method to design universal polar codes for discrete memoryless channels. It is also shown that the *less noisy* ordering of channels is preserved under polarization, and thus a good polar code for a given channel will perform well over a less noisy one.

2.1 Introduction

The compound channel models communication without perfect knowledge of the physical channel. The channel is assumed to belong to a certain class, and a code needs to be designed to perform well over all members of this class. The problem is relevant from a practical standpoint since one can rarely estimate the channel perfectly, and it is undesirable for small variations in the channel to impair the code performance dramatically.

Let \mathcal{W} be a set of binary-input memoryless channels $W: \{0,1\} \rightarrow \mathcal{Y}$. A rate

R is said to be achievable over \mathcal{W} if there exists a sequence of encoder–decoder pairs whose encoding rate converges to R and whose decoding error probability vanishes for all $W \in \mathcal{W}$. The highest achievable rate $C(\mathcal{W})$ is called the *compound capacity*, and is given by [11]

$$C(\mathcal{W}) = \sup_Q \inf_{W \in \mathcal{W}} I(W, Q).$$

Here, $I(W, Q)$ denotes the mutual information across channel W with input distribution Q . In this chapter, we are interested in the *symmetric* compound capacity $I(\mathcal{W})$, which is the highest achievable rate over \mathcal{W} by codebooks with an equal frequency of zeros and ones. Letting $I(W)$ denote the mutual information across W with uniform inputs, we have

$$I(\mathcal{W}) = \inf_{W \in \mathcal{W}} I(W).$$

We say that a code sequence of rate R achieves symmetric capacity *universally* if its error probability vanishes over all channels in the class $\{W : I(W) > R\}$.

In this chapter, we show that universal codes can be constructed by Arikan’s polarization methods [7]. We consider the setting where the channel is unknown only to the transmitter. This is an idealized version of the practical scenario where the receiver may estimate the channel prior to data transmission, for example through the use of training symbols. Polar coding for this setting was first considered by Hassani *et al.* [37], who concluded that Arikan’s original codes are not universal under successive cancellation (SC) decoding. It is worth noting, however, that under maximum likelihood decoding, *any* good code for the binary symmetric channel (BSC) is also good (up to a linear factor in its error probability) for *any* channel with the same capacity, and therefore a capacity-achieving polar code sequence for the BSC is in fact universal [65, pp. 87–89]. Unfortunately, no subexponential algorithm is known for maximum likelihood decoding of polar codes over arbitrary channels. It thus remains an open question whether one can construct polar codes that are universal under low-complexity decoders. This chapter answers this question in the affirmative.

In recent work, Kudekar, Richardson, and Urbanke [47] showed that spatially-coupled LDPC codes universally achieve the capacity of symmetric channels under low-complexity message-passing decoders, making them the first known class of codes to do so. Here, we show the same result for polar codes and for general channels (i.e., without symmetry assumptions). Hassani and Urbanke [36] have independently arrived at conclusions similar to the result in this chapter, and we compare the two approaches briefly in Section 6.4.

There are cases in which designing a polar code for multiple channels is easy. The most prominent of these is the *degraded* case: A polar code tailored to a given channel will also perform well over all upgraded versions of that channel [7], [43]. In Appendix A, we see that a similar statement holds for the more general class of *less noisy* comparable channels. In Appendix B, we will derive a sufficient condition to check the less noisy ordering for the class of binary-input symmetric output channels.

2.2 Method

Our aim here is to show a method to polarize channels universally. We will first discuss how to achieve rate 1/2, and in Section 2.3 show constructions that achieve arbitrary rates. As in Arkan's original method, we will polarize channels recursively. The construction will consist of two stages, which we will call the *slow* polarization and the *fast* polarization stages. Slow polarization will create only two types of channels after each recursion. Almost half of the polarized channels will be of the first type and become increasingly good, the other half will become increasingly bad. The indices of the good channels will be independent of the underlying channel, and thus universality will be attained at this stage. We will see, however, that this type of polarization is too slow to enable reliable SC decoding. In order to improve reliability, we will switch to the standard (fast) polarization method once sufficient universality is achieved.

Given two binary-input memoryless channels $W: \{0, 1\} \rightarrow \mathcal{Y}$ and $V: \{0, 1\} \rightarrow \mathcal{Z}$,

define the binary-input channels

$$(W, V)^-(y, z | x) = \sum_{u \in \{0,1\}} \frac{1}{2} W(y | u + x) V(z | u)$$

and

$$(W, V)^+(y, z, u | x) = \frac{1}{2} W(y | u + x) V(z | x).$$

Note that if $W \equiv V$, then these are equivalent to the standard polarized channels W^- and W^+ in [7]. We will let L_n and R_n denote the two channels that will emerge in the n th level of slow polarization. These are defined recursively through

$$\begin{aligned} L_0 &= R_0 = W \\ L_{n+1} &= (R_n, L_n)^- \\ R_{n+1} &= (R_n, L_n)^+ \end{aligned} \quad n = 0, 1, \dots \quad (2.1)$$

Observe that each recursion except the first combines two different channels to produce the channels of the next level. This is in contrast with the original polarization method, which combines identical channels to create 2^n polarized channels at the n th recursion,

$$\begin{aligned} W^{\mathbf{s}^-} &= (W^{\mathbf{s}}, W^{\mathbf{s}})^- \\ W^{\mathbf{s}^+} &= (W^{\mathbf{s}}, W^{\mathbf{s}})^+ \end{aligned}, \quad \mathbf{s} \in \{-, +\}^{n-1}.$$

It is readily seen that for all n we have

$$I(L_n) + I(R_n) = 2I(W).$$

Standard arguments also show that $I(L_n)$ is decreasing and $I(R_n)$ is increasing:

$$I(L_{n+1}) \leq I(L_n) \leq I(R_n) \leq I(R_{n+1}).$$

Since both $I(L_n)$ and $I(R_n)$ are monotone and bounded by 0 and 1, they have $[0, 1]$ -valued limits, which we respectively call $I(L_\infty)$ and $I(R_\infty)$. Further, it follows from [65, Lemma 2.1] that the inequalities above are strict for $n \geq 1$ unless $I(L_n) \in \{0, 1\}$ or $I(R_n) \in \{0, 1\}$. This implies the following polarization result.

Proposition 2.2.1.

(i) If $I(W) \geq 1/2$, then

$$I(L_\infty) = 2I(W) - 1, \quad I(R_\infty) = 1.$$

(ii) If $I(W) \leq 1/2$, then

$$I(L_\infty) = 0, \quad I(R_\infty) = 2I(W).$$

We now describe a transform that recursively produces the channels L_n and R_n . This is best done graphically; the claims will be evident from the figures. Note first that L_1 and R_1 are identical to W^- and W^+ , and thus can be obtained in the standard manner (Figure 2.1). In order to create L_2 and R_2 from these, one can take two independent (L_1, R_1) pairs, and combine an L_1 from one pair with an R_1 from the other, as in Figure 2.2. Following the notation of the figure, it can be easily checked that the channel $U_1 \rightarrow Y_1^4$ is equivalent to L_1 , channel $U_2 \rightarrow Y_1^4 U_1$ is equivalent to L_2 , channel $U_3 \rightarrow Y_1^4 U_1^2$ is equivalent to R_2 , and channel $U_4 \rightarrow Y_1^4 U_1^3$ is equivalent to R_1 . Inspecting the figure, one may be tempted to combine U_1 and U_4 to create another (L_2, R_2) pair, but some thought reveals that this would instead create channels with more complicated descriptions.

Instead, more (L_2, R_2) pairs can be obtained by combining more than two (L_1, R_1) pairs in a chain. This is shown in Figure 2.3, where four (L_1, R_1) pairs are chained. The resulting transform creates three (L_2, R_2) pairs. One can more generally

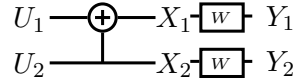


Figure 2.1. When X_1 and X_2 are uniform and independent, the channel $U_1 \rightarrow Y_1^2$ is equivalent to $L_1 = W^-$ and $U_2 \rightarrow Y_1^2 U_1$ is equivalent to $R_1 = W^+$.

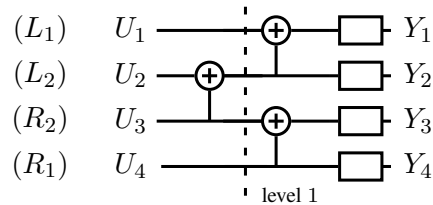


Figure 2.2. The channels $U_i \rightarrow Y_1^4 U_1^{i-1}$ are equivalent to those inside the parentheses.

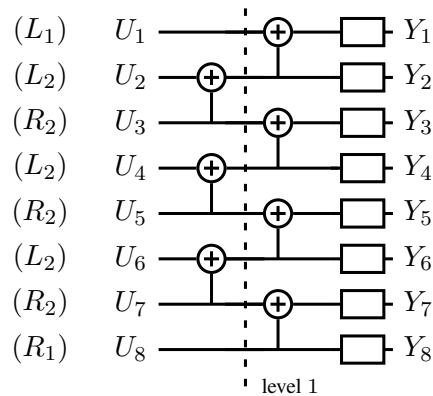


Figure 2.3. Four pairs of level-1 channels are chained to create six level-2 and two level-1 channels. The channels $U_i \rightarrow Y_1^8 U_1^{i-1}$ are equivalent to the ones on the left.

chain K channel pairs (L_1, R_1) to produce $K - 1$ channel pairs (L_2, R_2) . Thus, the fraction of (L_2, R_2) pairs can be made as close to 1 as desired by taking K sufficiently large. Observe also that the channels $U_i \rightarrow Y_1^{2K} U_1^{i-1}$ obtained by such a chain are equivalent to $U_i \rightarrow Y_{i-2}^{i+2} U_{i-2}^{i-1}$. That is, not all channel outputs are relevant to U_i .

There are several ways to continue this construction in order to polarize the channel beyond two levels. We describe here perhaps the simplest one, where chaining as in Figure 2.3 is used only at the second polarization level, as in the paragraph above. Each subsequent recursion combines only two blocks. The third level of this construction with $K = 4$ is shown in Figure 2.4. Here, only the level-2 channels L_2 and R_2 are

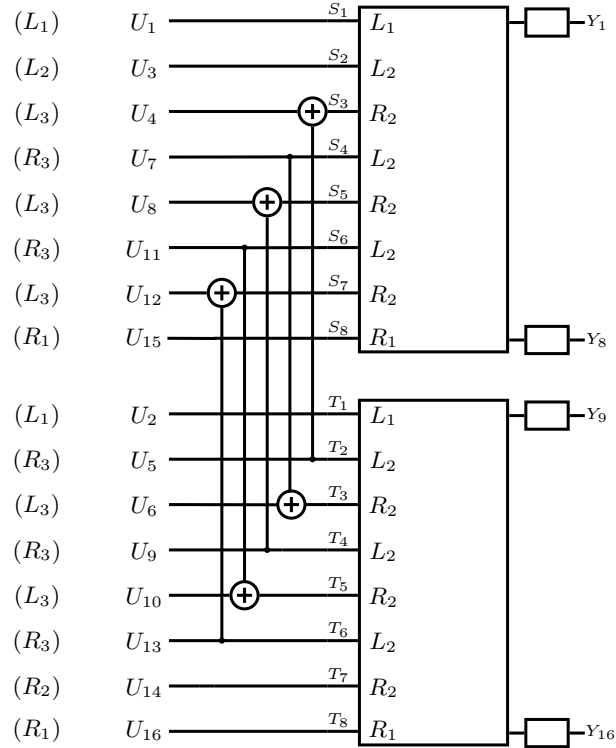


Figure 2.4. A 3-level transform with $K = 4$.

combined in the third recursion, L_1 and R_1 are not. Further, the first L_2 in the first block and the last R_2 in the second are also left unconnected, in order to ensure that the remaining channels polarize to the third level to produce L_3 and R_3 . This idea is easily extended to further levels: To obtain L_{n+1} and R_{n+1} in the $(n + 1)$ -th recursion, one only combines the L_n s from the first block with the R_n s from the second, and vice versa. The first L_n from the first block and the last R_n from the second block are left unconnected. This is shown in Figure 2.5. Each of the two blocks represents the transform in Figure 2.3. The channels written inside the respective blocks correspond to $S_i \rightarrow Y_1^8 S_1^{i-1}$ and $T_i \rightarrow Y_9^{16} T_1^{i-1}$. If one labels U_1 to U_{16} as above, then the channels $U_i \rightarrow Y_1^{16} U_1^{i-1}$ are equivalent to the ones on the left.

Recall that our initial goal was to ensure that all channels after the n th recursion become either L_n or R_n , but the procedure described above leaves some channels in lower levels of polarization. The number of these less polarized channels in fact increases

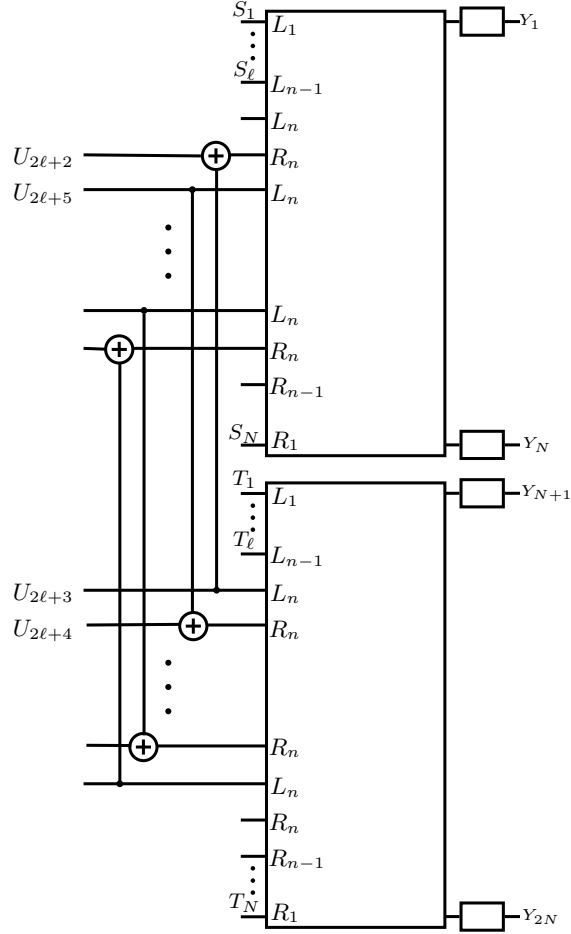


Figure 2.5. The $(n + 1)$ -level construction. Here, $\ell = 2^{n-1} - 1$.

with each recursion, but the loss is limited. One can indeed check that the blocklength is $N = 2^{n-1}K$ after the n th recursion, and the number of channels (L_i, R_i) at level $i = 1, \dots, n - 1$ is 2^{n-i} . Therefore the fraction of level- n channels can be lower bounded as

$$1 - \frac{\sum_{i=1}^{n-1} 2^{n-i}}{2^{n-1}K} \geq 1 - \frac{2^n}{2^{n-1}K} = 1 - \frac{2}{K},$$

which can be made arbitrarily close to 1 by picking a large K .

Observe that the construction described above is universal: The positions of the good channels, that is, R_n s, after the transformation is independent of the underlying channel W . One may therefore hope to use these channels to achieve rate $1/2$ over any

W with $I(W) \geq 1/2$. Unfortunately, however, the speed of polarization is too slow for an SC decoder to succeed. This is most easily seen by noting that the Bhattacharyya parameter of R_{n+1} is given by

$$Z(R_{n+1}) = Z(R_n)Z(L_n).$$

Since $Z(L_n)$ approaches a non-zero constant (in particular it approaches 1 if $I(W) = 1/2$) as n grows, the multiplicative improvement in $Z(R_n)$ gradually slows down (to a halt if $I(W) = 1/2$). This is in contrast with the squaring of the Bhattacharyya parameters in each ‘+’ transform in Arıkan’s standard method, which is necessary for the exponential decay of the error probability.

We now derive simple bounds on the speed of universal polarization using the extremal properties of the binary erasure channel (BEC) and the BSC under polarization. For this purpose, let $h: [0, 1/2] \rightarrow [0, 1]$ denote the binary entropy function, and let $a * b = a(1 - b) + b(1 - a)$ denote binary convolution. Define the functions

$$\begin{aligned} f(x, y) &= x(2y - x) \\ g(x, y) &= 2y - h(h^{-1}(x) * h^{-1}(2y - x)), \end{aligned}$$

over $y \in [0, 1]$ and $x \in [\max\{0, 2y - 1\}, y]$, and the functions

$$\begin{aligned} F_0(y) &= G_0(y) = y \\ F_n(y) &= f(F_{n-1}(y), y) \\ G_n(y) &= g(G_{n-1}(y), y) \end{aligned}, \quad n = 1, 2, \dots$$

Finally, let $H(W) = 1 - I(W)$ denote the entropy of the input to W given the output.

Proposition 2.2.2. $F_n(H(W)) \leq H(R_n) \leq G_n(H(W))$.

Proof. The claim holds trivially for $n = 0$. Suppose now that it holds for some $n \geq 1$.

Recall that among all pairs of channels V and W with given entropies $H(V) = h_1$ and $H(W) = h_2$, the entropy $H((V, W)^+)$ is minimized when V and W are both binary erasure channels (BECs) and maximized when both are binary symmetric channels (BSCs) [65, Lemma 2.1]. This implies that

$$f(H(R_n), H(W)) \leq H(R_{n+1}) \leq g(H(R_n), H(W)). \quad (2.2)$$

On the other hand, $f(x, y)$ and $g(x, y)$ are increasing in x . To see the latter, note that $h(x)$ is increasing for $x \in [0, 1/2]$, and thus it suffices to show that $k_x(t) = h^{-1}(t) * h^{-1}(2x - t)$ is decreasing in t . Defining $f = h^{-1}$, some algebra yields

$$\frac{d}{dt}k_x(t) = f'(t)[1 - 2f(2x - t)] - f'(2x - t)[1 - 2f(t)].$$

The right-hand-side of the above is at most zero, since $f'(t) \leq f'(2x - t)$ and $f(t) \leq f(2x - t) \leq 1/2$, which in turn are due to f being convex, increasing, and $[0, 1/2]$ -valued.

Thus, it follows from (2.2) that

$$\begin{aligned} F_{n+1}(H(W)) &= f(F_n(H(W)), H(W)) \\ &\leq f(H(R_n), H(W)) \end{aligned}$$

and

$$\begin{aligned} g(H(R_n), H(W)) &\leq g(G_n(H(W)), H(W)) \\ &= G_{n+1}(H(W)). \end{aligned}$$

Combining these with (2.2) implies the claim for $n + 1$, concluding the proof. \square

Observe that the above upper bound on $I(R_n)$ is obtained by replacing R_n and L_n with BECs with symmetric capacities $I(R_n)$ and $I(L_n)$ respectively before each polarization step. Similarly, the lower bound is obtained by replacing these channels with BSCs.

Recall that both descendants of a BEC are also BECs during polarization, whereas only one of BSCs descendant is a BSCs. This implies that while the upper bound is achieved by the BEC, the lower bound is loose. Tables 2.1 and 2.2 list the bounds for $I(W) = 0.5$ and $I(W) = 0.8$.

n	lower bound	upper bound
0	0.5	0.5
1	0.713	0.750
2	0.771	0.812
3	0.805	0.847
4	0.829	0.870
5	0.846	0.887
10	0.895	0.931
20	0.932	0.960
30	0.949	0.972
40	0.958	0.978

Table 2.1. Bounds on $I(R_n)$ for $I(W) = 0.5$.

n	lower bound	upper bound
0	0.8	0.8
1	0.928	0.960
2	0.957	0.986
3	0.972	0.994
4	0.981	0.997
5	0.986	0.999
10	0.996	0.999991
15	0.9990	0.99999991
20	0.9996	0.9999999991

Table 2.2. Bounds on $I(R_n)$ for $I(W) = 0.8$.

2.2.1 Universal polar coding

To obtain a good code, we can append Arikan’s fast (but not universal) polarization transform to the universal (but slow) polarization transform described above. That is, once n is sufficiently large so that $I(R_n) > 1 - \epsilon$ for all W , we may start polarizing R_n fast. The simplest way to do so is to take $M = 2^m$ copies of the slow polarization transform, and pass the M copies of each R_n through a length- M fast polarization transform.

Inputs to the remaining channels are frozen and the resulting code blocks are decoded in succession.

One may tailor the polar codes in the fast polarization stage to the channel that is least degraded with respect to all channels with $I(W) \geq 1 - \epsilon$. How to find such channels is shown in [51]. A computationally simpler alternative is to find a universal upper bound $Z(R_n) \leq \delta$ (as in Proposition 2.2.2) and tailor the code in the fast polarization stage to a BEC with erasure probability δ . This method is motivated by the fact that among all channels with a fixed $Z(W)$, the BEC's polarized descendants have the highest Bhattacharyya parameters, and the latter can be computed in linear time [7].

2.2.2 Rate

Since $I(R_n)$ is close to 1, both approaches mentioned in the previous paragraph will induce a negligible rate loss in the fast polarization stage. Recall also that the loss in the slow polarization stage is $O(1/K)$. Hence the rate of the code can be made as close to $1/2$ as desired.

2.2.3 Error probability

Recall that the reliabilities of the good channels after fast polarization is $o(2^{-M^\beta})$ for all $\beta < 1/2$ [6], and thus the block error probability of this code of length NM is upper bounded all $W \in \mathcal{W}(1/2)$ by

$$No(2^{-M^\beta}),$$

which for fixed N vanishes as M grows.

2.2.4 Complexity

To estimate the decoding complexity, it is useful to explain the decoding scheme in some detail:

The decoder can be thought of as SC decoders for the slow and fast polarization transforms operating in tandem. In each decoding stage, the SC decoders for all slow polarization blocks compute the likelihood ratio for the next bit-channel. If the latter is a frozen channel (i.e., any channel other than R_n), then the decoders simply declare the frozen bit values. Otherwise, they pass the likelihood ratios to the corresponding SC decoder for the fast polarization stage, which uses these likelihood ratios for decoding, and passes the decoded bit values back to the slow polarization decoder. A straightforward computation shows that the total complexity of this decoder is

$$O(N)\kappa_f(M) + M\kappa_s(N),$$

where $\kappa_f(M)$ and $\kappa_s(N)$ respectively are the decoding complexities of the fast polarization transform of length M and a slow polarization transform of length N . It is known [7] that $\kappa_f(M) = O(M \log M)$. Now, observe that the slow polarization transform is almost identical to the fast one; it only differs in the chaining operation in the second level and in the combination of non-identical channels at each step. It is easy to see that neither of these differences affects the complexity of computing the likelihood ratios of the polarized channels. That is, $\kappa_s(N) = O(N \log N)$. This implies that the total decoding complexity at blocklength MN is $O(MN \log MN)$, as in standard polar codes. Similar arguments show that the encoding complexity is also $O(MN \log MN)$. Note also that the chain length K affects encoding/decoding complexities only insofar as it appears as a linear factor in the blocklength.

2.3 Codes with Arbitrary Rates

We now discuss how to obtain universal polar codes with rates other than $1/2$. Recall that in the previous section we fixed the rate of the code by using only the universally good channel R_n for coding. When $I(\mathcal{W})$ is greater than $1/2$, the code rate can be increased by considering coding over L_n also, since Proposition 2.2.1 then implies

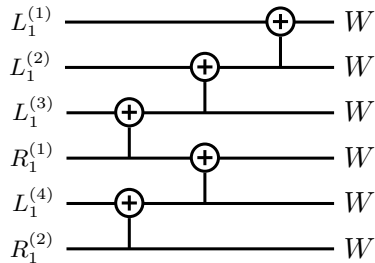


Figure 2.6. A one-level transform that combines $b + g = 6$ channels with $g = 2$ and $b = 4$. Only the labels of the channels at the corresponding locations are shown.

$I(L_n) > 0$. For example, once $I(R_n)$ is sufficiently close to 1, one may obtain more universally good channels by slow-polarizing L_n alone. When $I(W)$ is less than $1/2$, the same method can be used by slow-polarizing R_n further once $I(R_n)$ becomes sufficiently close to $2I(W)$. Each stage of this polarization method turns half of the remaining nonextremal channels to extremal ones. The resulting good channels can then be fast-polarized for coding. However, the blocklengths of such constructions can be very large, since, as we have seen in Tables 2.1 and 2.2, even a single stage of slow polarization requires a large number of recursions.

Instead, here we construct codes with rates $g/(b + g)$ for given positive integers g and b by generalizing the ideas in Section 2.2. Following the reasoning there, this can be done if one can (i) combine $b + g$ channels at a time to create only $b + g$ channel types after each level of slow polarization, and (ii) ensure that g of these become better in each step and the remaining b become worse. Once the good channels become nearly perfect, one can boost their reliabilities through fast polarization.

It thus suffices to describe a construction that has properties (i) and (ii). Again, the simplest description is through figures. Figure 2.6 shows an example of the type of transforms we will consider. In particular, the transform circuit consists of $b + g$ horizontal wires, each of which has a single modulo-2 addition that connects it to the wire below. Starting at the second wire from the top, one can place this connection to the right or to the left of the connection above.

The channels $U_i \rightarrow Y_1^{b+g} U_1^{i-1}$ produced by the transform are defined as usual,

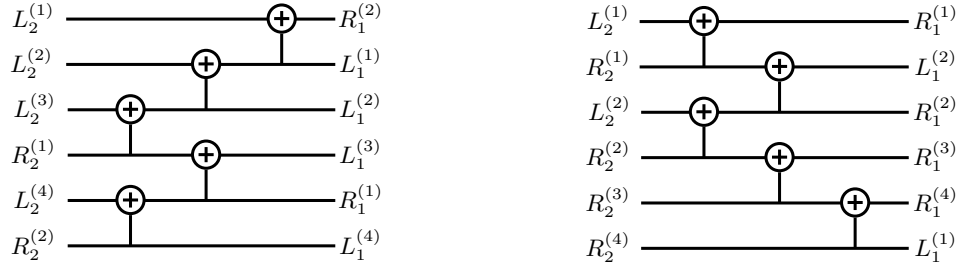


Figure 2.7. A rate-2/6 transform (left) and a rate-4/6 transform (right). Channels enter both transforms on the right-hand-side and produce the channels on the left-hand-side.

where the inputs and outputs are numbered in increasing order from top to bottom. We label the channels as follows (see Figure 2.6): If a wire's connection to the bottom is on the left side of its connection to the top, then the corresponding channel is called a type- L channel. The i th such channel from the top is called $L_1^{(i)}$. Similarly, a channel whose connection to the bottom is on the right side of its connection to the top is called a type- R channel. In addition, the top channel is a type- L channel and the bottom channel is a type- R channel. Observe that the fraction of type- L and type- R channels can be adjusted to arbitrary non-zero values by an appropriate choice of transform.

We will restrict our attention to two classes of transforms for which the claims will be easy to verify. For $g \leq b$ (that is, when the target rate is less than $1/2$), we will use the transform that produces the two channel types in the order

$$\underbrace{LL \dots LL}_{b-g} \underbrace{LRRLR \dots LR}_{g \text{ pairs}} \quad (2.3)$$

That is, the top $b - g$ channels will be type- L , followed by an alternating sequence of type- L and type- R channels. In order to define a recursion, we need to specify the order in which the transform combines these $b + g$ in each level. In the present case, the input order is obtained by cyclically down-shifting (2.3) by one:

$$R \underbrace{LL \dots LL}_{b-g+1} \underbrace{RL \dots RL}_{g-1 \text{ pairs}} \quad (2.4)$$

When $g \geq b$, the top channels produced by the transform will be of alternating

types, followed by a sequence of type- R channels:

$$\underbrace{LRLR \dots LRLR}_{b \text{ pairs}} \underbrace{RR \dots R}_{g-b} \quad (2.5)$$

These channels will be combined in the each recursion after up-shifting the order (2.5) by one:

$$\underbrace{RL \dots RL}_{b-1 \text{ pairs}} \underbrace{RR \dots RL}_{g-b+1} \quad (2.6)$$

Examples of both recursions are shown in Figure 2.7. We will label the channels produced by these recursions as in the previous section: If $g \leq b$, then the channels $L_n^{(1)}, \dots, L_n^{(b)}$ and $R_n^{(1)} \dots R_n^{(g)}$ after the n th recursion are transformed through (2.3) and (2.4) to produce $L_{n+1}^{(1)}, \dots, L_{n+1}^{(b)}$ and $R_{n+1}^{(1)} \dots R_{n+1}^{(g)}$. The first recursion takes $b + g$ copies of W as input. For the case $g \geq b$, the recursions are defined through (2.5) and (2.6).

The reason for the labeling above is the analogy between type- L (respectively, type- R) channels and the channel L_1 (respectively, R_1) of Section 2.2. Indeed, suppose that we combine $b + g$ copies of W through a transform that produces the channels $L_1^{(1)}, \dots, L_1^{(b)}$ and $R_1^{(1)}, \dots, R_1^{(g)}$. We clearly have

$$\sum_{i=1}^b I(L_n^{(i)}) + \sum_{i=1}^g I(R_n^{(i)}) = (b + g)I(W).$$

Moreover, type- L channels are worse than W and type- R channels are better:

Proposition 2.3.1. *For all $i = 1, \dots, b$ and $j = 1, \dots, g$ we have*

$$I(L_1^{(i)}) \leq I(W) \leq I(R_1^{(j)}).$$

Both inequalities are strict unless $I(W) \in \{0, 1\}$.

Proof. We prove the statement for the case $g \leq b$. The case $g > b$ can be proved similarly.

By construction, we have from top to bottom the following sequence of channels

$$\underbrace{L_1^{(1)}, \dots, L_1^{(b-g)}}_{b-g}, \underbrace{L_1^{(b-g+1)}, R_1^{(1)}, \dots, L_1^{(b)}, R_1^{(g)}}_{g \text{ pairs}}.$$

Define $Q_1 = W$ and $Q_{i+1} = (Q_i, W)^+$ for $i = 1, 2, \dots$. If $g = 1$, we have $L_1^{(i)} = (Q_i, W)^-$ for $1 \leq i \leq b$ and $R_1^{(1)} = (Q_b, W)^+$. If $g > 1$, we have

$$L_1^{(i)} = \begin{cases} (Q_i, W)^- & 1 \leq i \leq b-g \\ (Q_{b+g-1}, W^-)^- & i = b-g+1 \\ (W^+, W^-)^- & b-g+1 < i < b \\ (W^+, W)^- & i = b \end{cases}$$

and

$$R_1^{(j)} = \begin{cases} (Q_{b-g+1}, W^-)^+ & j = 1 \\ (W^+, W^-)^+ & 1 < j < g \\ (W^+, W)^+ & j = g \end{cases}.$$

The claim then follows by noting that

$$\begin{aligned} I((W, V)^-) &\leq \min\{I(W), I(V)\} \\ &\leq \max\{I(W), I(V)\} \\ &\leq I((W, V)^+) \end{aligned}$$

for any two channels W and V . Strict inequalities follow again from [65, Lemma 2.1]. \square

Having created b bad and g good channels out of W , we wish to enhance polarization by making the bad channels worse and the good channels better. The main result of this section is that these recursions indeed polarize channels universally:

Proposition 2.3.2.

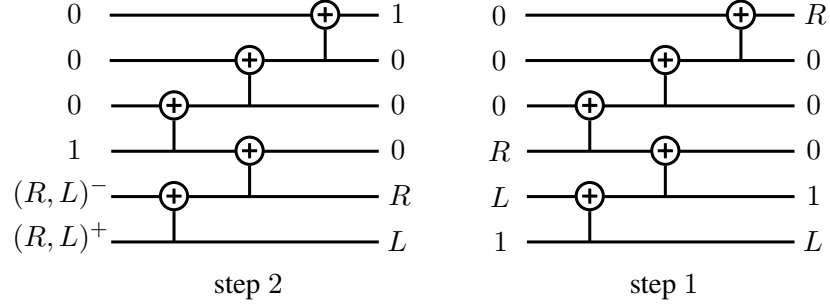


Figure 2.8. An example of polarization.

(i) If $I(W) \geq g/(b+g)$, then for all $1 \leq i \leq g$

$$\lim_{n \rightarrow \infty} I(R_n^{(i)}) = 1.$$

(ii) If $I(W) \leq g/(b+g)$, then for all $1 \leq i \leq b$

$$\lim_{n \rightarrow \infty} I(L_n^{(i)}) = 0.$$

Proof. We prove (i) for the case $g \leq b$. The arguments for the remaining three cases are similar. Recall the recursions (2.3) and (2.4). Define $Q_1 = R_n^{(g)}$ and $Q_{i+1} = (Q_i, L_n^{(i)})^+$ for $1 \leq i \leq b-g$. If $g = 1$, we have

$$L_{n+1}^{(i)} = (Q_i, L_n^{(i)})^- \text{ for } 1 \leq i \leq b$$

and

$$R_{n+1}^{(1)} = (Q_b, L_n^{(b)})^+.$$

Therefore, $I(L_n^{(i)})$ is decreasing while $I(R_n^{(1)})$ is increasing. If $g > 1$, define

$$P_i^+ = (L_n^{(b-g+i)}, R_n^{(i)})^+$$

$$P_i^- = (L_n^{(b-g+i)}, R_n^{(i)})^-$$

for $1 \leq i < g$. One can check that the order of inputs to the recursion implies (see Figure 2.7 for reference)

$$L_{n+1}^{(i)} = \begin{cases} (Q_i, L_n^{(i)})^- & 1 \leq i \leq b-g \\ (Q_{b-g+1}, P_1^-)^- & i = b-g+1 \\ (P_{g-b+i-1}^+, P_{g-b+i}^-)^- & b-g+1 < i < b \\ (P_{g-1}^+, L_n^{(b)})^- & i = b \end{cases}.$$

Note that $I(P_{g-b+i}^-) = I((L_n^{(i)}, R_n^{(g-b+i)})^-) \leq I(L_n^{(i)})$ for $b-g+1 \leq i < b$. It follows that $I(L_{n+1}^{(i)}) \leq I(L_n^{(i)})$ for all $1 \leq i \leq b$. Similarly, one can check that

$$R_{n+1}^{(j)} = \begin{cases} (Q_{b-g+1}, P_1^-)^+ & j = 1 \\ (P_{j-1}^+, P_j^-)^+ & 1 < j < g \\ (P_{g-1}^+, L_n^{(b)})^+ & j = g \end{cases}.$$

Note that $I(Q_{b-g+1}) \geq I(R_n^{(g)})$ and $I(P_{j-1}^+) = I((L_n^{(b-g+j-1)}, R_n^{(j-1)})^+) \geq I(R_n^{(j-1)})$ for $1 < j \leq g$. It then follows that $I(R_{n+1}^{(1)}) \geq I(R_n^{(g)})$ and $I(R_{n+1}^{(j)}) \geq I(R_n^{(j-1)})$ for $1 < j \leq g$. That is, type- R channels at level $n+1$ are better than the ones at level n , with a shift in indices.

To show that the improvement in $I(R_n^{(i)})$ is strict unless all the type- R channels are perfect, one needs to rule out the following possibility: If at some point in the polarization process some, but not all, type- R channels become perfect, then the perfect channels entering subsequent recursions may stall the polarization of the non-perfect ones. We now argue that the structure of the channel combinations does not allow this. Suppose that all but one type- R channels polarize to perfect ones. Then, there must be at least one unpolarized type- L channel, since otherwise the inequality $I(W) \geq g/(b+g)$ would be violated. Suppose that there is only one such type- L channel $L^{(k)}$ and all others are polarized to useless ones. One can then check that either the unpolarized type- L

and type- R channels will be combined in the next recursion (which will further polarize the type- R channel), or their positions will change. In particular, the type- L channel index k will remain unchanged after each recursion, while the type- R channel index will be cyclically shifted by one. If $1 \leq k \leq b - g + 1$, then $L^{(k)}$ will be combined with the unpolarized type- R channel when the type- R channel index is shifted to g . On the other hand, if $b - g + 1 < k \leq b$, then $L^{(k)}$ will be combined with the unpolarized type- R channel when the type- R channel index is shifted to neighboring positions $k - b + g - 1$ or $k - b + g$. Therefore, regardless of the unpolarized type- R channel's position, strict polarization will take place in at most g recursions. Figure 2.8 is an example of strict polarization of period two over the rate-2/6 recursion. Suppose that at some point in time the channels are polarized as in the graph on the right. Here, 0 denotes a completely noisy channel, 1 is a perfect channel, and R and L are the mediocre R - and type- L channels. Channels enter the recursion from the right. After the first step, the channels change positions, but no polarization takes place. Nevertheless, the new positions of R and L ensure that they are combined in the next step (left). One can check that mediocre channels are always combined eventually, regardless of their initial positions. Therefore, the type- R channel will polarize further, eventually becoming perfect. The same reasoning can be used when there is more than one unpolarized type- R channel and type- L channel. \square

2.3.1 Polar coding

Fix a transform of rate $g/(b+g)$. The code construction is identical to the one in Section 2.2: In the first level, channels are combined in the usual fashion. This is followed by a single step of chaining K transforms that combines channels of different types. Then, each subsequent step combines $b+g$ transform blocks in the same fashion. Once sufficient universal polarization is attained, the good channels $R_n^{(i)}$ are fast-polarized further using the Arıkan transform.

2.3.2 Rate

As in the rate-1/2 case, the slow polarization transform involves leaving some channels in lower levels of polarization. Similar arguments to those in Section 2.2 show that the fraction of such channels is upper bounded by

$$\frac{(g+b)^2}{K},$$

which can be made as small as desired by picking a large K .

2.3.3 Error probability

Since the reliability of the good channels are determined essentially by the fast polarization stage, the error probability of the SC decoder can again be upper bounded for all $W \in \mathcal{W}(g/b + g)$ by

$$No(2^{-M^\beta}),$$

where N and M respectively are the lengths of the slow and the fast polarization stages.

2.3.4 Complexity

The present construction differs from the one in Section 2.2 only in the size $b+g$ of the one-level transform, and it is easily seen that the transforms discussed in this section can be encoded and decoded in linear time. Hence, $b+g$ does not affect the encoding and decoding complexities, which are both $O(MN \log MN)$ for a blocklength- MN code.

2.4 Discussion

In independent work [36], Hassani and Urbanke propose two polarization-based methods to construct universal codes. On close inspection, one of these methods and the one presented here are seen to be complementary. In particular, the method here guarantees universality in the first stage and reliability in the second, whereas the con-

struction in [36] reverses this order by combining identical channels in the first stage (i.e., fast polarization) and distinct channels in the second (i.e., slow polarization). It is evident from both works that many other variations are possible for constructing universal polar codes, such as interleaving the fast and slow polarization stages. Such alternatives may help reduce the impractically large blocklengths that the present chapter's methods require (see Table 2.1) to simultaneously achieve universality and reliability. For this purpose one may also consider using larger $(b + g)$ -type constructions for simple fractional rates such as $1/2$, or mixing the unconnected channels into the process to increase the speed of slow polarization. The investigation of these are left for future study.

In addition to providing robustness to point-to-point channel coding, universal polarization is also of interest from a theoretical perspective. Recall that one of the many appeals of polarization methods is the ease with which they have been extended to other communication settings. Polar codes' optimality have already been established for multiple-access channels [5], degraded wiretap channels [52], lossless [4], lossy [44], distributed source coding [5], and some special cases of broadcast channels [31]. However, standard polarization methods are difficult to extend to settings with two or more receivers, and the main bottleneck appears to be the incompatibility of polar code designs for different receivers. Since the appearance of [36] and [64] on arXiv.org, universal polar coding techniques have been shown to achieve the best known inner bounds in various network communication settings, such as general broadcast channels [56], interference channels [74, 76] (see Chapter 3), relay channels [75] (see Chapter 6), and general wiretap channels [19, 33, 78].

It is worth mentioning that the methods discussed here also yield universal source codes, and can be extended to non-binary alphabets using standard arguments [65, Chapter 3].

Acknowledgment

This chapter is, in part, a reprint of the material in the paper: Eren Şaşođlu and Lele Wang, “Universal polarization,” accepted for publication in *IEEE Transactions on Information Theory*, 2015.

Chapter 3

Interference Channels: Polar Codes

In the following two chapters, we will investigate an important random coding scheme, *simultaneous decoding*, in interference channels. Two low-complexity channel coding schemes will be proposed, both of which achieve desired theoretical performance. In this chapter, we present the first solution using polar codes. It achieves the Han–Kobayashi inner bound for two-user interference channels and generalizes to interference networks.

3.1 Introduction

Consider an interference channel $p(y_1, y_2|x_1, x_2)$ as depicted in Figure 3.1, in which sender $i \in \{1, 2\}$ wishes to communicate an independent message reliably to its respective receiver i .

A $(2^{nR_1}, 2^{nR_2}, n)$ code for the interference channel consists of

- two message sets $[1 : 2^{nR_1}]$ and $[1 : 2^{nR_2}]$,
- two encoders, where encoder $i \in \{1, 2\}$ assigns a codeword $x_i^n(m_i)$ to each message

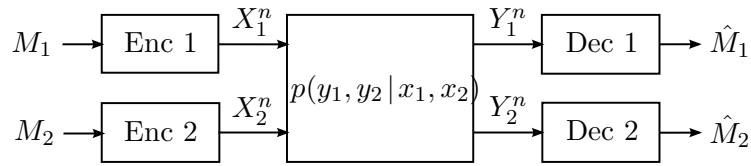


Figure 3.1. Two-user interference channel.

$m_i \in [1 : 2^{nR_i}]$, and

- two decoders, where decoder $i \in \{1, 2\}$ assigns an estimate \hat{m}_i or an error message e to each received sequence y_i^n .

We assume that the message pair (M_1, M_2) is uniformly distributed over $[1 : 2^{nR_1}] \times [1 : 2^{nR_2}]$. The average probability of error is defined as $P_e^{(n)} = \mathbb{P}\{(\hat{M}_1, \hat{M}_2) \neq (M_1, M_2)\}$. A rate pair (R_1, R_2) is said to be *achievable* if there exists a sequence of $(2^{nR_1}, 2^{nR_2}, n)$ codes such that $\lim_{n \rightarrow \infty} P_e^{(n)} = 0$. The capacity region is the closure of the set of achievable rate pairs (R_1, R_2) .

One important decoding scheme for interference channels is *simultaneous decoding*. It is a key component in the Han–Kobayashi coding scheme [34], whereby each receiver, instead of treating interference as noise, decodes for the intended message as well as part of the interfering message. Recently, Bandemer, El Gamal, and Kim showed that simultaneous nonunique decoding is rate-optimal for random code ensembles with superposition coding and time sharing [8].

Unfortunately, simultaneous decoding uses multiuser sequence detection at the core of its operation and it is not known how this can be implemented in low complexity. Consequently, several heuristic approaches have been developed that attempt to achieve “similar” performance; see, for example, [48, 82].

In Chapters 3 and 4, we address this problem from a different angle and ask the following questions. Is simultaneous decoding really needed? Is there an alternative coding scheme that achieves the same performance at low complexity?

Treating interference as noise and successive cancellation decoding (with no rate-splitting) are the two main decoding schemes used in practice, both of which achieve

strictly smaller rate regions than simultaneous decoding. Recently, Zhao, Tan, Avestimehr, Diggavi, and Pottie [83] studied successive cancellation decoding for more than two layers of Gaussian superposition codes, as an application of the rate-splitting scheme by Rimoldi and Urbanke [60] and Grant, Rimoldi, Urbanke, and Whiting [32] to interference channels. In Section 3.2, we investigate this application in full generality by considering arbitrary code distributions for superposition coding, which is sometimes necessary as pointed out in [80]. We show that regardless of the number of layers and the code distribution of each layer, the standard single-block rate-splitting scheme fails to achieve the simultaneous decoding inner bound in interference channels.

In the following sections, we present a polar coding solution that achieves the simultaneous decoding performance. The method is built on two techniques developed recently by Hassani and Urbanke [36], and Arikan [5]. We explain these two building blocks in Sections 3.4, and show how they can be combined and applied to achieve the simultaneous decoding performance in Section 3.5. In Section 3.6, we extend the result to general interference networks, which includes the Han–Kobayashi inner bound as an important special case.

3.2 Insufficiency of Single-Block Rate-Splitting

In this section, we consider the symmetric Gaussian interference channels. We show a corner point of the simultaneous decoding inner bound is not achievable using rate-splitting with successive cancellation decoding. We assume average power constraint P . The channel outputs at the receivers for inputs X_1 and X_2 are

$$Y_1 = X_1 + gX_2 + Z_1,$$

$$Y_2 = gX_1 + X_2 + Z_2,$$

where g is a fixed constant and $Z_1, Z_2 \sim \mathcal{N}(0, 1)$ are additive Gaussian noise components, independent of (X_1, X_2) . We define the received *signal-to-noise ratio* as $S = P$ and the

received *interference-to-noise ratio* as $I = g^2P$.

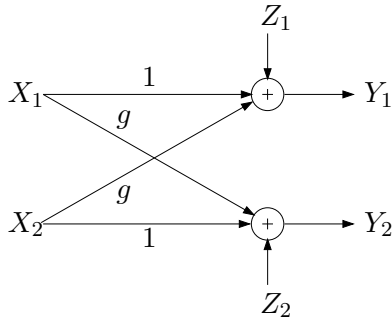


Figure 3.2. Symmetric Gaussian interference channel.

The (s, t, d_1, d_2, F) rate-splitting scheme and its achievable rate region are defined as follows.

Rate splitting. We represent the message M_1 by s independent parts $M_{11}, M_{12}, \dots, M_{1s}$ at rates $R_{11}, R_{12}, \dots, R_{1s}$, respectively, and the message M_2 by t independent parts $M_{21}, M_{22}, \dots, M_{2t}$ at rates $R_{21}, R_{22}, \dots, R_{2t}$, respectively.

Codebook generation. We use superposition coding. Fix a cdf $F = F(q)F(u^s|q)F(v^t|q)$ such that Q is finite, $E[(U_s)^2] \leq P$, and $E[(V_t)^2] \leq P$. Randomly and independently generate q^n according to $\prod_{k=1}^n F(q_k)$. Randomly and conditionally independently generate $2^{nR_{11}}$ sequences $u_1^n(m_{11})$, $m_{11} \in [1 : 2^{nR_{11}}]$, each according to a product cdf of $F(u_1|q)$. For $j \in [2 : s]$, for each m_1^{j-1} , randomly and conditionally independently generate $2^{nR_{1j}}$ sequences $u_j^n(m_{1j}|m_1^{j-1})$, $m_{1j} \in [1 : 2^{nR_{1j}}]$, each according to a product cdf of $F(u_j|u^{j-1}, q)$. Randomly and conditionally independently generate $2^{nR_{21}}$ sequences $v_1^n(m_{21})$, $m_{21} \in [1 : 2^{nR_{21}}]$, each according to a product cdf of $F(v_1|q)$. For $j \in [2 : t]$, for each m_2^{j-1} , randomly and conditionally independently generate $2^{nR_{2j}}$ sequences $v_j^n(m_{2j}|m_2^{j-1})$, $m_{2j} \in [1 : 2^{nR_{2j}}]$, each according to a product cdf of $F(v_j|v^{j-1}, q)$.

Encoding. To send message pair $(m_1, m_2) = (m_1^s, m_2^t)$, encoder 1 transmits $x_1^n(m_1^s) = u_s^n(m_{1s}|m_1^{s-1})$ and encoder 2 transmits $x_2^n(m_2^t) = v_t^n(m_{2t}|m_2^{t-1})$.

Decoding. We use successive cancellation decoding. Define the *decoding order* d_1 at decoder 1 as an ordering of elements in $\{U^s, V_{\mathcal{A}}\}$ and d_2 at decoder 2 as an ordering of elements in $\{U_{\mathcal{B}}, V^t\}$, where $\mathcal{A} \subseteq [1 : t]$ and $\mathcal{B} \subseteq [1 : s]$.

As an example, suppose that message M_1 is split into two parts while message M_2 is not split. The decoding orders are

$$d_1: U \rightarrow X_2 \rightarrow X_1,$$

$$d_2: U \rightarrow X_1 \rightarrow X_2,$$

where in case of a single split, we write $(U, X_1) = (U_1, U_2)$ and $X_2 = V_1$. This means that decoder 1 recovers M_{11}, M_2 , and M_{12} successively and decoder 2 recovers M_{11}, M_{12} and M_2 successively. More precisely, upon receiving y_1^n at decoder 1, decoding proceeds in three steps:

1. Decoder 1 finds the unique message \hat{m}_{11} such that

$$(u^n(\hat{m}_{11}), y_1^n, q^n) \in \mathcal{T}_\epsilon^{(n)}.$$

2. If \hat{m}_{11} is found, decoder 1 finds the unique \hat{m}_2 such that

$$(u^n(\hat{m}_{11}), x_2^n(\hat{m}_2), y_1^n, q^n) \in \mathcal{T}_\epsilon^{(n)}.$$

3. If $(\hat{m}_{11}, \hat{m}_2)$ is found, find the unique \hat{m}_{12} such that

$$(u^n(\hat{m}_{11}), x_2^n(\hat{m}_2), x_1^n(\hat{m}_{11}, \hat{m}_{12}), y_1^n, q^n) \in \mathcal{T}_\epsilon^{(n)}.$$

Similarly, upon receiving y_2^n , decoding proceeds in three steps:

1. Decoder 2 finds the unique message \hat{m}_{11} such that

$$(u^n(\hat{m}_{11}), y_2^n, q^n) \in \mathcal{T}_\epsilon^{(n)}.$$

2. If \hat{m}_{11} is found, decoder 2 finds the unique \hat{m}_{12} such that

$$(u^n(\hat{m}_{11}), x_1^n(\hat{m}_{11}, \hat{m}_{12}), y_2^n, q^n) \in \mathcal{T}_\epsilon^{(n)}.$$

3. If $(\hat{m}_{11}, \hat{m}_{12})$ is found, find the unique \hat{m}_2 such that

$$(u^n(\hat{m}_{11}), x_1^n(\hat{m}_{11}, \hat{m}_{12}), x_2^n(\hat{m}_2), y_2^n, q^n) \in \mathcal{T}_\epsilon^{(n)}.$$

Following the standard analysis of the error probability [27, Sec. 4.5.1], $P_e^{(n)}$ tends to zero as $n \rightarrow \infty$ if

$$R_{11} < I(U; Y_1 | Q) - \delta(\epsilon), \quad (3.1a)$$

$$R_2 < I(X_2; Y_1 | U, Q) - \delta(\epsilon), \quad (3.1b)$$

$$R_{12} < I(X_1; Y_1 | U, X_2, Q) - \delta(\epsilon), \quad (3.1c)$$

$$R_{11} < I(U; Y_2 | Q) - \delta(\epsilon), \quad (3.1d)$$

$$R_{12} < I(X_1; Y_2 | U, Q) - \delta(\epsilon), \quad (3.1e)$$

$$R_2 < I(X_2; Y_2 | X_1, Q) - \delta(\epsilon). \quad (3.1f)$$

By Fourier–Motzkin elimination, (R_1, R_2) is achievable if

$$R_1 < \min\{I(U; Y_1 | Q), I(U; Y_2 | Q)\} + \min\{I(X_1; Y_1 | U, X_2, Q), I(X_1; Y_2 | U, Q)\}, \quad (3.2)$$

$$R_2 < \min\{I(X_2; Y_1 | U, Q), I(X_2; Y_2 | X_1, Q)\}.$$

We note some common misconception in the literature (see [28] and the references therein) that the bounds on R_{11} and R_{12} in (3.1) simplify to

$$R_1 < \min\{I(U; Y_1 | Q) + I(X_1; Y_1 | U, X_2, Q), I(X_1; Y_2 | Q)\},$$

which leads to an incorrect conclusion that the Han–Kobayashi inner bound can be

achieved by rate-splitting and successive cancellation. As pointed out in [28], successive decoding requires individual rate constraints (3.1d) and (3.1e) instead of the sum-rate constraint $R_1 < I(X_1; Y_2|Q)$. Moreover, a proper application of the Fourier–Motzkin elimination procedure requires taking the minimum for four cases of sum-rates, which leads to (3.2).

For more layers of splitting and general decoding orders, decoding can be performed in a similar fashion. Thus, an (s, t, d_1, d_2, F) rate-splitting scheme is specified by

- the numbers s and t of independent parts in messages $M_1 = (M_{11}, \dots, M_{1s})$ and $M_2 = (M_{21}, \dots, M_{2t})$,
- the cdf $F = F(q)F(u^s|q)F(v^t|q)$, and
- the decoding orders d_1 and d_2 .

Let $\mathcal{R}(s, t, d_1, d_2, F)$ denote the achievable rate region of the (s, t, d_1, d_2, F) rate-splitting scheme. Let $\mathcal{R}^*(s, t, d_1, d_2)$ be the closure of $\cup_F \mathcal{R}(s, t, d_1, d_2, F)$. Let $\mathsf{C}(x) := \frac{1}{2} \log(1+x)$ and define

$$R_1^*(s, t, d_1, d_2) = \max\{R_1 : (R_1, \mathsf{C}(S)) \in \mathcal{R}^*(s, t, d_1, d_2)\}$$

as the maximal achievable rate R_1 such that R_2 is at individual capacity.

Now we are ready to state the main result of this section. Assume that the symmetric Gaussian interference channel has *strong but not very strong* interference, i.e., $S < I < S(S+1)$. The capacity region is the set of rate pairs (R_1, R_2) such that

$$R_1 \leq \mathsf{C}(S),$$

$$R_2 \leq \mathsf{C}(S),$$

$$R_1 + R_2 \leq \mathsf{C}(I + S),$$

which is achieved by simultaneous decoding with $X_1, X_2 \sim N(0, P)$ and $Q = \emptyset$ [34, 67].

Theorem 3.2.1 states that the corner point of this region is not achievable using any (s, t, d_1, d_2, F) rate-splitting scheme.

Theorem 3.2.1. *For the symmetric Gaussian interference channel with $S < I < S(S + 1)$,*

$$R_1^*(s, t, d_1, d_2) < C\left(\frac{I}{1+S}\right)$$

for any finite s, t and decoding orders d_1, d_2 .

Remark 3.2.1. *The idea of the standard rate-splitting scheme for the multiple access channel is to represent each message by multiple parts and encode them into superimposed layers. Combined with successive cancellation decoding, this superposition coding scheme transforms the multiple access channel into a sequence of point-to-point channels. For the interference channel, which consists of two underlying multiple access channels $p(y_i|x_1, x_2), i = 1, 2$, however, this idea no longer works. Here rate-splitting induces two sequences of point-to-point channels that have different qualities in general. To ensure reliable communication, the messages have to be loaded at the rate of the worse channel on each layer, which in general incurs a total rate loss. Theorem 3.2.1 essentially states that there is no split of the messages that “equalizes” the qualities of the two point-to-point channels on each layer, even when the decoding orders of the layers are optimized. Rate-splitting is alternatively viewed as mapping a boundary point of one multiple access rate region to a corner point of another multiple access rate region in a higher dimensional space [32, 60]. Theorem 3.2.1 shows that there is no such mapping in general under which the corresponding corner points for the two multiple access channels coincide.*

3.3 Proof of Theorem 3.2.1

For the simplicity of notation, we prove the claim for $Q = \emptyset$. The case for general Q follows the same logic. First, we show the special property of the (s, t, d_1, d_2, F) rate-splitting scheme that achieves $R_1^*(s, t, d_1, d_2)$.

Lemma 3.3.1. *For any (s, t, d_1, d_2, F) rate-splitting scheme that achieves $R_1^*(s, t, d_1, d_2)$, we can assume without loss of generality that $s = t$ and the decoding orders are*

$$\begin{aligned} d_1^* &: U_1 \rightarrow V_1 \rightarrow \cdots \rightarrow U_{s-1} \rightarrow V_{s-1} \rightarrow U_s \rightarrow V_s, \\ d_2^* &: U_1 \rightarrow U_2 \rightarrow \cdots \rightarrow U_s \rightarrow V_1 \rightarrow V_2 \rightarrow \cdots \rightarrow V_s. \end{aligned}$$

Proof. First, we specify the optimal decoding order at decoder 2. Fix any (s, t, d_1, d_2, F) rate-splitting scheme that guarantees $R_2 = C(S)$. Suppose that some part V_j of the message M_2 is decoded earlier than U_k of the message M_1 at decoder 2, that is,

$$d_2: d_{21} \rightarrow V_j \rightarrow U_k \rightarrow d_{22}.$$

Now flip the decoding order of V_j and U_k in \tilde{d}_2 as

$$\tilde{d}_2: d_{21} \rightarrow U_k \rightarrow V_j \rightarrow d_{22}$$

and construct $(s, t, d_1, \tilde{d}_2, F)$ rate-splitting scheme, where the message splitting, the underlying distribution, and decoding order d_1 remain the same. Let \tilde{R}_{ij} be the rate of message m_{ij} in the $(s, t, d_1, \tilde{d}_2, F)$ rate-splitting scheme. Then we have all the rates remain the same except

$$\begin{aligned} R_{2j} &= I(V_j; Y_2 | V^{j-1}, U^{k-1}), \\ \tilde{R}_{2j} &= I(V_j; Y_2 | V^{j-1}, U^k), \\ R_{1k} &= I(U_k; Y_2 | V^j, U^{k-1}), \\ \tilde{R}_{1k} &= I(U_k; Y_2 | V^{j-1}, U^{k-1}). \end{aligned}$$

Note that $R_{2j} \leq \tilde{R}_{2j}$ since U^k is independent of V^j . On the other hand, since R_{2j} already

results in full rate at R_2 , we must have $\tilde{R}_{2j} = R_{2j}$. It follows that

$$I(U_k; V_j | Y_2, V^{j-1}, U^{k-1}) = 0$$

and therefore $R_{1j} = \tilde{R}_{1j}$. Thus, without loss of generality, d_2 can be of the form $U_1 \rightarrow U_2 \rightarrow \dots \rightarrow U_s \rightarrow V_1 \rightarrow V_2 \rightarrow \dots \rightarrow V_t$.

Next, we show merging two consecutive parts of the same message in both decoding orders improves the rates in general. Consider an (s, t, d_1, d_2, F) rate-splitting scheme with decoding order

$$\begin{aligned} d_1: d_{11} &\rightarrow U_k \rightarrow U_{k+1} \rightarrow d_{12}, \\ d_2: d_{21} &\rightarrow U_k \rightarrow U_{k+1} \rightarrow d_{21}, \end{aligned}$$

where $d_{ij}, i, j \in \{1, 2\}$, are parts of the decoding order pair (d_1, d_2) that are not explicitly specified. Construct a $(s-1, t, d'_1, d'_2, F')$ rate-splitting scheme, where

$$U'_j = \begin{cases} U_j & \text{if } j \in [1 : k-1] \\ (U_k, U_{k+1}) & \text{if } j = k \\ U_{j+1} & \text{if } j \in [k+1 : s-1], \end{cases}$$

$V'_j = V_j$ for $j \in [1 : t]$, and decoding orders are

$$d'_1: d_{11} \rightarrow U'_k \rightarrow d_{12} \quad \text{and} \quad d'_2: d_{21} \rightarrow U'_k \rightarrow d_{22}.$$

Let R_{ij} be the rate of m_{ij} in the (s, t, d_1, d_2, F) superposition coding scheme and R'_{ij} be the rate of m_{ij} in the $(s-1, t, d'_1, d'_2, F')$ rate-splitting scheme. Following standard

analysis of error probability, we have

$$R'_{1j} = \begin{cases} R_{1j} & \text{if } j \in [1 : k - 1] \\ R_{1,j+1} & \text{if } j \in [k + 1 : s - 1] \end{cases}$$

and $R'_{1k} \geq R_{1k} + R_{1,k+1}$, which follows since $\min\{a_1 + a_2, b_1 + b_2\} \geq \min\{a_1, b_1\} + \min\{a_2, b_2\}$. Moreover, we have $R'_{2j} = R_{2j}$ for $j \in [1 : t]$. By Fourier–Motzkin elimination, we have

$$\mathcal{R}(s, t, d_1, d_2, F) \subseteq \mathcal{R}(s - 1, t, d'_1, d'_2, F').$$

A similar conclusion holds for two consecutive parts (V_k, V_{k+1}) of message M_2 .

Finally, combining the merge operation and the optimal decoding order d_2^* , we can assume without loss of generality that the message parts from M_1 and M_2 are decoded alternately at decoder 1. Therefore, the class of rate-splitting schemes can be narrowed down to (s, s, d_1^*, d_2^*, F) rate-splitting scheme with

$$d_1^* : U_1 \rightarrow V_1 \rightarrow \cdots \rightarrow U_{s-1} \rightarrow V_{s-1} \rightarrow U_s \rightarrow V_s.$$

Note that by setting $U_1 = \emptyset$, this coding scheme recovers the case when V_1 of message M_2 is decoded first at the decoder 1. By setting $V_s = \emptyset$, this coding scheme recovers the case when the whole message M_2 is decoded at decoder 1. Therefore, there is no loss of generality in assuming the (s, s, d_1^*, d_2^*, F) rate-splitting scheme. \square

The following Lemma 3.3.2 that provides a necessary condition for achieving the corner point of the capacity region.

Lemma 3.3.2. *A necessary condition for $(2, 2, d_1^*, d_2^*, F)$ rate-splitting scheme to attain*

the corner point is that the distribution F is such that

$$X_1 \sim \mathcal{N}(0, P) \quad \text{and} \quad X_2 \sim \mathcal{N}(0, P).$$

Proof of Lemma 3.3.2. Suppose the $(2, 2, d_1^*, d_2^*, F)$ rate-splitting scheme attains the corner point $(\mathsf{C}(I/(1+S)), \mathsf{C}(S))$. Let $U = U_1$ and $V = V_1$. Then, for R_2 we have

$$\begin{aligned} \mathsf{C}(S) &= \min\{I(V; Y_1|U), I(V; Y_2|X_1)\} \\ &\quad + I(X_2; Y_2|X_1, V) \\ &\leq I(X_2; Y_2|X_1) \\ &\leq \mathsf{C}(S). \end{aligned} \tag{3.3}$$

Given X_1 , the channel from X_2 to Y_2 is a Gaussian channel with SNR S . Therefore the condition $X_2 \sim \mathcal{N}(0, P)$ is necessary for (3.3) to hold with equality. At the corner point, R_1 must satisfy the following

$$\begin{aligned} \mathsf{C}(I/(1+S)) &= \min\{I(U; Y_1), I(U; Y_2)\} \\ &\quad + \min\{I(X_1; Y_1|U, V), I(X_1; Y_2|U)\} \\ &\leq I(U; Y_2) + I(X_1; Y_2|U) \\ &= I(X_1; Y_2) \\ &\leq \mathsf{C}(I/(1+S)). \end{aligned} \tag{3.4}$$

Given $X_2 \sim \mathcal{N}(0, P)$, the channel from X_1 to Y_2 is a Gaussian channel with SNR $I/(1+S)$. Therefore, the condition $X_1 \sim \mathcal{N}(0, P)$ is necessary for (3.4) to hold with equality. \square

We also need the following technical lemma.

Lemma 3.3.3. *Let $F(u, x)$ be any distribution such that $X \sim \mathcal{N}(0, P)$ and $I(U; Y) = 0$, where $Y = X + Z$ with $Z \sim \mathcal{N}(0, 1)$ independent of X . Then, $I(U; X) = 0$.*

Proof. For every $u \in \mathcal{U}$, we have

$$\begin{aligned} I(X; Y | U = u) &= h(Y | U = u) - h(Y | X, U = u) \\ &\stackrel{(a)}{=} h(Y) - h(Y | X) \\ &= \mathsf{C}(P), \end{aligned}$$

where (a) follows since Y is independent of U and $U \rightarrow X \rightarrow Y$ form a Markov chain. Suppose for some u , $E(X^2 | U = u) < P$, i.e., the effective channel SNR is strictly less than P . Then $I(X; Y | U = u) < P$. As a result, we must have $E(X^2 | U = u) \geq P$ for all $u \in \mathcal{U}$. On the other hand,

$$\begin{aligned} P &\leq \int E(X^2 | U = u) dF(u) \\ &= E(X^2) \\ &= P, \end{aligned}$$

which forces $E(X^2 | U = u) = P$ for almost all u . Since Gaussian input $\mathsf{N}(0, P)$ is the unique distribution that attains the rate $\mathsf{C}(P)$ in the Gaussian channel with SNR P , the distribution $F(x|u)$ must be $\mathsf{N}(0, P)$ for almost all u . Therefore $I(U; X) = 0$. \square

Now we are ready to establish the suboptimality of the rate-splitting scheme.

It is straightforward to check for the case $s = 1$. For $s = 2$, we prove by contradiction. Let $U = U_1$ and $V = V_1$. The achievable rate region of the $(2, 2, d_1^*, d_2^*, F)$ rate-splitting scheme is the set of rate pairs (R_1, R_2) such that

$$\begin{aligned} R_1 &< \min\{I(U; Y_1), I(U; Y_2)\} \\ &\quad + \min\{I(X_1; Y_1 | U, V), I(X_1; Y_2 | U)\} := I_1, \\ R_2 &< \min\{I(V; Y_1 | U), I(V; Y_2 | X_1)\} \\ &\quad + I(X_2; Y_2 | X_1, V) := I_2. \end{aligned}$$

Assume that the corner point of the capacity region is achieved by the $(2, 2, d_1^*, d_2^*, F)$ rate-splitting scheme, that is,

$$I_1 = C(I/(1+S)), \quad (3.5)$$

$$I_2 = C(S). \quad (3.6)$$

Then, by Lemma 3.3.2, we must have $X_1 \sim N(0, P)$ and $X_2 \sim N(0, P)$. Consider

$$\begin{aligned} I_1 &= \min\{I(U; Y_1), I(U; Y_2)\} \\ &\quad + \min\{I(X_1; Y_1|U, V), I(X_1; Y_2|U)\}, \\ &\leq I(U; Y_1) + I(X_1; Y_2|U) \end{aligned} \quad (3.7)$$

$$= h(Y_1) - h(Y_1|U) + h(Y_2'|U) - h(Y_2'|X_1), \quad (3.8)$$

where $Y_2' = Y_2/g = X_1 + (X_2 + Z_2)/g$. Since

$$\begin{aligned} \frac{1}{2} \log(2\pi e(S+1)/g^2) &= h(Y_2'|X_1) \\ &\leq h(Y_2'|U) \\ &\leq h(Y_2') \\ &= \frac{1}{2} \log(2\pi e(I+S+1)/g^2), \end{aligned}$$

there exists an $\alpha \in [0, 1]$ such that $h(Y_2'|U) = (1/2) \log(2\pi e(\alpha I + S + 1)/g^2)$. Moreover, since $X_2 \sim N(0, P)$ and $I < S(1+S)$, the channel $X_1 \rightarrow Y_1$ is a degraded version of the channel $X_1 \rightarrow Y_2'$, i.e., $Y_1 = Y_2' + Z'$, where $Z' \sim N(0, I + 1 - (S+1)/g^2)$ is independent of X_1 and X_2 . By the entropy power inequality,

$$\begin{aligned} 2^{2h(Y_1|U)} &\geq 2^{2h(Y_2'|U)} + 2^{2h(Z'|U)} \\ &= 2\pi e(\alpha S + I + 1). \end{aligned}$$

Therefore, it follows from (3.8) that

$$\begin{aligned}
I_1 &\leq h(Y_1) - h(Y_1|U) + h(Y_2'|U) - h(Y_2'|X_1) \\
&\leq \frac{1}{2} \log \left(\frac{(I+S+1)(\alpha I+S+1)}{(\alpha S+I+1)(1+S)} \right) \\
&\stackrel{(a)}{\leq} C(I/(1+S)),
\end{aligned}$$

where (a) follows since $S < I$. To match the standing assumption in (3.5), we must have equality in (a), which forces $\alpha = 1$ and $h(Y_2'|U) = (1/2) \log(2\pi e(I+S+1)/g^2) = h(Y_2')$, i.e., $I(U; Y_2') = 0$. Note that $X_1, X_2 \sim N(0, P)$ and the channel from X_1 to Y_2' is a Gaussian channel. Applying Lemma 3.3.3 yields

$$I(U; X_1) = 0. \quad (3.9)$$

Now, I_2 can be simplified to

$$I_2 = \min\{I(V; Y_1|U), I(V; Y_2|X_1)\} + I(X_2; Y_2|X_1, V) \quad (3.10)$$

$$\begin{aligned}
&\stackrel{(b)}{=} \min\{I(V; Y_1), I(V; Y_2|X_1)\} + I(X_2; Y_2|X_1, V) \\
&\leq I(V; Y_1) + I(X_2; Y_2|X_1, V) \\
&= h(\tilde{Y}_1) - h(\tilde{Y}_1|V) + h(\tilde{Y}_2|V) - h(Y_2|X_1, X_2),
\end{aligned} \quad (3.11)$$

where (b) follows since $I(U; Y_1|V) \leq I(U; Y_1|X_2) = I(U; X_1 + Z_1) \leq I(U; X_1) = 0$, which implies $I(V; Y_1|U) = I(V; Y_1)$. In (3.11), we denote $\tilde{Y}_1 = Y_1/g = X_2 + (X_1 + Z_1)/g$ and $\tilde{Y}_2 = X_2 + Z_2$. Since

$$\begin{aligned}
\frac{1}{2} \log(2\pi e) &= h(\tilde{Y}_2|X_2) \\
&\leq h(\tilde{Y}_2|V) \\
&\leq h(\tilde{Y}_2)
\end{aligned}$$

$$= \frac{1}{2} \log(2\pi e(1+S)),$$

there exists a $\beta \in [0, 1]$ such that $h(\tilde{Y}_2|V) = (1/2) \log(2\pi e(1+\beta S))$. Moreover, since $X_1 \sim \mathcal{N}(0, P)$ and $I < S(1+S)$, \tilde{Y}_1 is a degraded version of \tilde{Y}_2 , i.e., $\tilde{Y}_1 = \tilde{Y}_2 + \tilde{Z}$, where $\tilde{Z} \sim \mathcal{N}(0, (1+s)/g^2 - 1)$ is independent of X_1 and X_2 . Applying the entropy power inequality, we have

$$\begin{aligned} 2^{2h(\tilde{Y}_1|V)} &\geq 2^{2h(\tilde{Y}_2|V)} + 2^{2h(\tilde{Z}|V)} \\ &= 2\pi e(\beta S + (1+S)/g^2). \end{aligned}$$

Therefore, it follows from (3.11) that

$$\begin{aligned} I_2 &\leq h(\tilde{Y}_1) - h(\tilde{Y}_1|V) + h(\tilde{Y}_2|V) - h(Y_2|X_1, X_2) \\ &\leq \frac{1}{2} \log \left(\frac{(I+S+1)(1+\beta S)}{g^2(\beta S + (1+S)/g^2)} \right) \\ &\stackrel{(c)}{\leq} \mathbf{C}(S), \end{aligned}$$

where (c) follows from the channel condition $I < (1+S)S$. To match the standing assumption in (3.6), we must have equality in (c), which forces $\beta = 1$ and $h(\tilde{Y}_2|V) = (1/2) \log(2\pi e(1+S)) = h(\tilde{Y}_2)$, i.e., $I(V; \tilde{Y}_2) = 0$. Note that $X_2 \sim \mathcal{N}(0, P)$ and the channel from X_2 to \tilde{Y}_2 is a Gaussian channel. Applying Lemma 3.3.3 yields

$$I(V; X_2) = 0. \tag{3.12}$$

However, conditions (3.9) and (3.12) implies

$$\begin{aligned} I(X_1; Y_1|U, V) &= I(U, V, X_1; Y_1) - I(U, V; Y_1) \\ &= I(X_1; Y_1) + I(V; Y_1|X_1) \\ &\quad - I(U; Y_1) - I(V; Y_1|U) \end{aligned}$$

$$\stackrel{(d)}{=} I(X_1; Y_1),$$

where (d) follows since $I(U; Y_1) \leq I(U; X_1) = 0$ and $I(V; Y_1|U) \leq I(V; Y_1|X_1) \leq I(V; X_2) = 0$. Therefore,

$$\begin{aligned} I_1 &= \min\{I(X_1; Y_1), I(X_1; Y_2)\} \\ &= I(X_1; Y_1) \\ &= \mathsf{C}(S/(1+I)) \\ &< \mathsf{C}(I/(1+S)), \end{aligned}$$

which contradicts (3.5) and completes the proof for $s = 2$.

Finally, to show the suboptimality of the (s, s, d_1^*, d_2^*, F) superposition coding scheme for $s > 2$, denote the achievable rate region as the set of rate pairs (R_1, R_2) such that $R_1 < I_1$ and $R_2 < I_2$. Then

$$\begin{aligned} I_1 &= \min\{I(U_1; Y_1), I(U_1; Y_2)\} \\ &\quad + \sum_{j=2}^s \min\{I(U_j; Y_1|U^{j-1}, V^{j-1}), I(U_j; Y_2|U^{j-1})\} \\ &\leq I(U_1; Y_1) + \sum_{j=2}^s I(U_j; Y_2|U^{j-1}) \\ &= I(U_1; Y_1) + I(U_2^s; Y_2|U_1) \\ &\stackrel{(a)}{=} I(U_1; Y_1) + I(X_1; Y_2|U_1), \end{aligned} \tag{3.13}$$

where (a) follows since $U^s \rightarrow X_1 \rightarrow Y_1$ form a Markov chain. Similarly, we have

$$\begin{aligned} I_2 &= \sum_{j=1}^{s-1} \min\{I(V_j, Y_1|U^{j-1}, V^{j-1}), I(V_j; Y_2|X_1, V^{j-1})\} \\ &\quad + I(V_s; Y_2|X_1, V^{s-1}) \\ &\leq \min\{I(V_1, Y_1|U_1), I(V_1; Y_2|X_1)\} \end{aligned}$$

$$\begin{aligned}
& + \sum_{j=2}^s I(V_j; Y_2 | X_1, V^{j-1}) \\
& = \min\{I(V_1, Y_1 | U_1), I(V_1; Y_2 | X_1)\} \\
& \quad + I(X_2; Y_2 | X_1, V_1). \tag{3.14}
\end{aligned}$$

Note that (3.13) and (3.14) are of the same form as (3.7) and (3.10), respectively. Therefore, the suboptimality follows from the same arguments as when $s = 2$. This completes the proof.

3.4 Polar Coding Preliminaries

In this section, we present a coding scheme that achieves the simultaneous decoding inner bound using the recently invented polar codes [7]. We start by reviewing two techniques by Hassani and Urbanke [38] and by Arıkan [5] that serve as building blocks for our design.

3.4.1 Aligning Polarized Indices

Consider two binary-input memoryless channels $P : X \rightarrow Y$ and $Q : X \rightarrow Z$ with equal symmetric capacities $I(P) = I(Q)$. Recall that the symmetric capacity is the mutual information between the input and the output of the channel when the input is distributed uniformly. Suppose we wish to design a polar code that performs well over both of these channels. For $n = 2^k$, define $U^n = X^n G_n$, where $G_n = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}^{\otimes k} B_n$ is the standard polar transformation. Here, $\otimes k$ denotes the k th Kronecker power and B_n is the ‘bit-reversal’ permutation. Assume that X^n is uniform. Define the channels

$P_i : U_i \rightarrow Y^n U^{i-1}$ and $Q_i : U_i \rightarrow Z^n U^{i-1}$ and sets

$$\begin{aligned}
 \mathcal{G}_Y &= \{i \in [1:n] : I(P_i) > 1 - 2^{-n^\beta}\}, \\
 \mathcal{G}_Z &= \{i \in [1:n] : I(Q_i) > 1 - 2^{-n^\beta}\}, \\
 \mathcal{B}_Y &= \{i \in [1:n] : I(P_i) < 2^{-n^\beta}\}, \\
 \mathcal{B}_Z &= \{i \in [1:n] : I(Q_i) < 2^{-n^\beta}\}.
 \end{aligned} \tag{3.15}$$

for some $\beta < 1/2$. Standard polarization results imply that $|\mathcal{G}_Y|/n \approx I(P) = I(Q) \approx |\mathcal{G}_Z|/n$ for large n , and thus almost all bit indices belong to one of the following four sets:

$$\begin{aligned}
 \mathcal{A}_I &= \mathcal{G}_Y \cap \mathcal{G}_Z, \\
 \mathcal{A}_{II} &= \mathcal{G}_Y \cap \mathcal{B}_Z, \\
 \mathcal{A}_{III} &= \mathcal{B}_Y \cap \mathcal{G}_Z, \\
 \mathcal{A}_{IV} &= \mathcal{B}_Y \cap \mathcal{B}_Z.
 \end{aligned}$$

To understand the performance of standard polar codes on channels P and Q , it suffices to consider the bit indices of the above four types, and assume that the remaining bit values are fixed and revealed to all receivers. Note that type-I indices, i.e., those in \mathcal{A}_I , see clean channels under both P and Q and thus can carry information. Similarly, type-IV indices are bad under both P and Q and must be fixed. Type-II and III indices are *incompatible*, i.e., they are good under one channel and under the other. Moreover, the fraction $(|\mathcal{A}_{II}| + |\mathcal{A}_{III}|)/n$ of incompatible indices is non-negligible in general [37], and therefore standard polar coding does not achieve the compound capacity of arbitrary channels P and Q .

Hassani and Urbanke [36] propose a simple solution to the incompatibility problem, which *aligns* the good indices of the two channels. Given two binary-input memo-

ryless channels $P : X_1 \rightarrow Y_1$ and $Q : X_2 \rightarrow Y_2$, define the binary-input channels

$$(P, Q)^-(y_1, y_2 | u_1) = \sum_{u_2} \frac{1}{2} P(y_1 | u_1 \oplus u_2) Q(y_2 | u_2),$$

$$(P, Q)^+(y_1, y_2, u_1 | u_2) = \frac{1}{2} P(y_1 | u_1 \oplus u_2) Q(y_2 | u_2),$$

and note that

$$\begin{aligned} I((P, Q)^-) &\leq \min\{I(P), I(Q)\} \\ I((P, Q)^+) &\geq \max\{I(P), I(Q)\}. \end{aligned} \tag{3.16}$$

Now let i and j be a type-II and a type-III index, respectively. That is,

$$\begin{aligned} I(P_i) &\approx 1 \quad \text{and} \quad I(P_j) \approx 0, \\ I(Q_i) &\approx 0 \quad \text{and} \quad I(Q_j) \approx 1. \end{aligned}$$

It then follows from (3.16) that

$$\begin{aligned} I((P_i, P_j)^-) &\approx 0 \quad \text{and} \quad I((P_i, P_j)^+) \approx 1, \\ I((Q_i, Q_j)^-) &\approx 0 \quad \text{and} \quad I((Q_i, Q_j)^+) \approx 1, \end{aligned}$$

In words, combining two incompatible indices results in an almost perfect ‘plus’ channel and almost useless ‘minus’ channel, regardless of the underlying channel. This ‘aligns’ the mutual informations for such indices. In particular, taking two blocks of U^n , one can combine almost all type-II indices from one block with type-III indices from the other block, since $|\mathcal{A}_{\text{II}}|/n \approx |\mathcal{A}_{\text{III}}|/n$. More precisely, suppose $\mathcal{A}_{\text{II}} = \{c_1, c_2, \dots, c_{q_1}\}$ and $\mathcal{A}_{\text{III}} = \{d_1, d_2, \dots, d_{q_2}\}$, where the elements are written in increasing order. Define $U^n = X^n G_n$ and $E^n = X_{n+1}^{2n} G_n$. Then, combining U_{c_j} with E_{d_j} , $j = 1, \dots, q = \min\{q_1, q_2\}$,

and leaving the remaining symbols uncombined yields the length- $2n$ sequence

$$\begin{aligned} \tilde{U}^{2n} = & (U^{c_1-1}, E^{d_1-1}, U_{c_1} \oplus E_{d_1}, E_{d_1}, \\ & \dots \\ & U_{c_{q-1}+1}^{c_{q-1}}, E_{d_{q-1}+1}^{d_{q-1}}, U_{c_q} \oplus E_{d_q}, E_{d_q}, \\ & U_{c_{q+1}}^n, E_{d_{q+1}}^n). \end{aligned}$$

Then, the mutual informations of channels $\tilde{U}_i \rightarrow Y^{2n}\tilde{U}^{i-1}$ and $\tilde{U}_i \rightarrow Z^{2n}\tilde{U}^{i-1}$ are aligned for the combined indices $\tilde{U}_i = U_{c_j} \oplus E_{d_j}$ and $\tilde{U}_i = E_{d_j}$, and unchanged for the remaining ones. Note again that the indices in \mathcal{A}_{III} of the first block and \mathcal{A}_{II} of the second block are not combined with each other and remain incompatible. This is to ensure that the combined indices are polarized as desired. The fraction of incompatible indices is thus halved by this alignment, to $(|\mathcal{A}_{\text{II}}| + |\mathcal{A}_{\text{III}}|)/2n$. Recursively aligning the indices m times in this fashion then reduces this fraction to $(|\mathcal{A}_{\text{II}}| + |\mathcal{A}_{\text{III}}|)/2^r n$, and thus the rate $I(P) = I(Q)$ can be achieved on both channels by picking a large m .

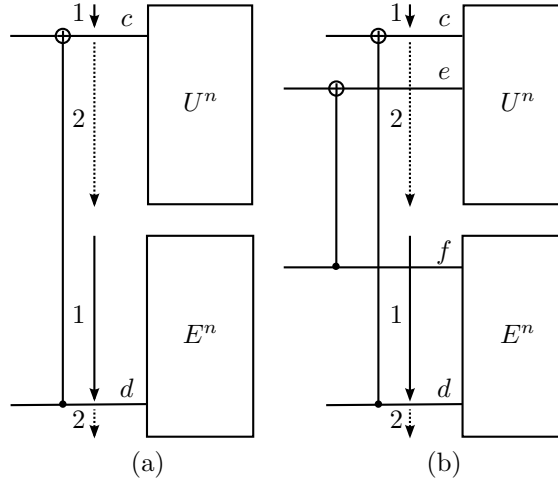


Figure 3.3. Alignment of the incompatible indices

In the following, we illustrate how to align incompatible indices through examples.

Example 3.4.1. Suppose that $\mathcal{A}_{\text{II}} = \{c\}$ and $\mathcal{A}_{\text{III}} = \{d\}$. We combine U_c from block 1

with E_d from block 2 as in Figure 3.3 (a). The numbers on the arrow indicate the decoding order. That is, variables along the ‘1’s are decoded before those along the ‘2’s.

Example 3.4.2. Suppose now that $\mathcal{A}_{II} = \{c, e\}$ and $\mathcal{A}_{III} = \{f, d\}$, with $f < c$, and that one tries to align all indices in one recursion, as in Figure 3.3 (b). Here again, the order of successive decoding is indicated by the numbered arrows. Observe, however, that in order to decode $U_c \oplus E_d$, one needs to know E^{d-1} , and in particular E_f . On the other hand, the decoding of E_f involves U_e , which is not available before knowing $U_c \oplus E_d$. Successive decoding is therefore infeasible. This shows the necessity of sorting type II and type III indices in increasing order and combining the j -th type II index from one block with the j -th type III index from the other block.

3.4.2 ‘Polar Splitting’ for MACs

Consider a two-user MAC $(\mathcal{X} \times \mathcal{W}, P(y|x, w), \mathcal{Y})$, where sender 1 and sender 2 wish to communicate M_1 and M_2 to the receiver by respectively sending codewords $X^n(M_1)$ and $W^n(M_2)$ over n uses of the channel. The capacity region of this channel is given by

$$\bigcup_p \mathcal{R}(p), \quad (3.17)$$

where the union is over all distributions of the form $p = p(q)p(x|q)p(w|q)P(y|x, w)$, and $\mathcal{R}(p)$ is the set of non-negative rate pairs (R_1, R_2) satisfying

$$\begin{aligned} R_1 &\leq I(X; Y, W|Q), \\ R_2 &\leq I(W; Y, X|Q), \\ R_1 + R_2 &\leq I(X, W; Y|Q). \end{aligned} \quad (3.18)$$

The subset of $\mathcal{R}(p)$ satisfying $R_1 + R_2 = I(X, W; Y|Q)$ is called its *dominant face*, and the two points $(I(X; Y|Q), I(W; Y, X|Q))$ and $(I(X; Y, W|Q), I(W; Y|Q))$ are called its

corner points. We will first consider uniform X and W and constant Q ; generalizations to arbitrary distributions are discussed in Section 3.5.2.

In [5], Arikan develops a polar coding method that achieves the entire dominant face based on the following observations: Let $U^n = X^n G_n$ and $V^n = W^n G_n$. Consider the chain rules of the form

$$\sum_{i=1}^{2n} I(S_i; Y^n | S^{i-1}),$$

where $S^{2n} = (S_1, \dots, S_{2n})$ is a *monotone* permutation of $U^n V^n$, i.e., elements of both U^n and V^n appear in increasing order in S^{2n} . Let \mathcal{S}_U and \mathcal{S}_V respectively denote the set of indices of S^{2n} with $S_i = U_k$ and $S_i = V_l$ for some $k, l \in [1 : n]$, and define the rates

$$\begin{aligned} R_1 &= \frac{1}{n} \sum_{i \in \mathcal{S}_U} I(S_i; Y^n | S^{i-1}), \\ R_2 &= \frac{1}{n} \sum_{i \in \mathcal{S}_V} I(S_i; Y^n | S^{i-1}). \end{aligned} \tag{3.19}$$

The entire region $\mathcal{R}(p)$ can be achieved by polar coding if (R_1, R_2) can be set to arbitrary values on the dominant face and if the mutual informations $I(S_i; Y^n | S^{i-1})$ are polarized. It turns out that these requirements are satisfied by permutations of the form $S^{2n} = (U^i, V^n, U_{i+1}^n)$.

Proposition 3.4.1 ([5]). *For every $\epsilon > 0$, $\beta < 1/2$, and rate pair (I_1, I_2) on the dominant face of $\mathcal{R}(p)$, there exist an n and a permutation $S^{2n} = (U^i, V^n, U_{i+1}^n)$ such that*

$$(i) \quad |R_1 - I_1| < \epsilon \text{ and } |R_2 - I_2| < \epsilon,$$

(ii)

$$\frac{|\mathcal{G}^{(1)}|}{n} > R_1 - \epsilon \quad \text{and} \quad \frac{|\mathcal{G}^{(2)}|}{n} > R_2 - \epsilon,$$

where

$$\mathcal{G}^{(1)} = \{i \in \mathcal{S}_U : I(S_i; Y^n | S^{i-1}) > 1 - 2^{-n^\beta}\},$$

$$\mathcal{G}^{(2)} = \{i \in \mathcal{S}_V : I(S_i; Y^n | S^{i-1}) > 1 - 2^{-n^\beta}\}.$$

3.5 Two-user Compound MAC

We are now ready to describe a polar coding scheme for the two-user compound MAC consisting of two channels $P_Y(y|x, w)$ and $P_Z(z|x, w)$. The channel is assumed to be known at the receiver but not at the transmitter. A rate pair (R_1, R_2) is achievable if there exists a sequence of codes with rates approaching (R_1, R_2) and vanishing error probability on both MACs. The capacity region is described by

$$\bigcup_p (\mathcal{R}_Y(p) \cap \mathcal{R}_Z(p)), \quad (3.20)$$

where $\mathcal{R}_Y(p)$ is as in (3.18) and $\mathcal{R}_Z(p)$ is the rate region obtained by replacing Y by Z in (3.18). Recall that for the simple MAC, the time-sharing random variable Q in (3.18) can be replaced by a convex hull operation on the union in (3.17). However, in the compound case, Q is necessary since the rate region (3.20) is in general strictly larger than the convex hull of $\bigcup_p (\mathcal{R}_Y(p) \cap \mathcal{R}_Z(p))$ when p is of the form $p(x)p(w)P_Y(y|x, w)P_Z(z|x, w)$.

3.5.1 Uniform Independent Inputs

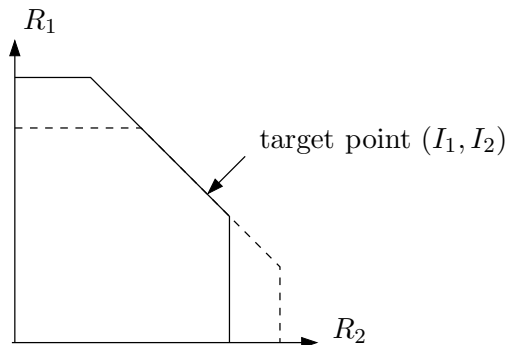


Figure 3.4. Two MAC regions with equal sum-rate.

Assume that X and W are uniform and independent, $Q = \emptyset$. The simplest nontrivial case is when the two pentagons in (3.20) intersect as in Figure 3.4, where

the sum-rates are equal and the two dominant faces have non-empty intersection. Let (I_1, I_2) be a rate point on the dominant face of this intersection. Let $U^n = X^n G_n$ and $V^n = W^n G_n$. By Proposition 3.4.1, there exists an n and two monotone permutations S^{2n} and T^{2n} for which the mutual informations $I(S_i; Y^n | S^{i-1})$ and $I(T_i; Z^n | T^{i-1})$ are polarized, and the corresponding rate pairs in (3.19) are close to (I_1, I_2) . However, as in the point-to-point case, the two sets of mutual informations $\{I(S_i; Y^n | S^{i-1}): i \in \mathcal{S}_U\}$ and $\{I(T_i; Z^n | T^{i-1}): i \in \mathcal{T}_U\}$ may be incompatible. Similarly to the point-to-point case, one can define the type of index U_i by comparing the mutual informations $I(S_j; Y^n | S^{j-1})$ and $I(T_k; Z^n | T^{k-1})$ where $S_j = T_k = U_i$. The indices for V can be defined similarly.

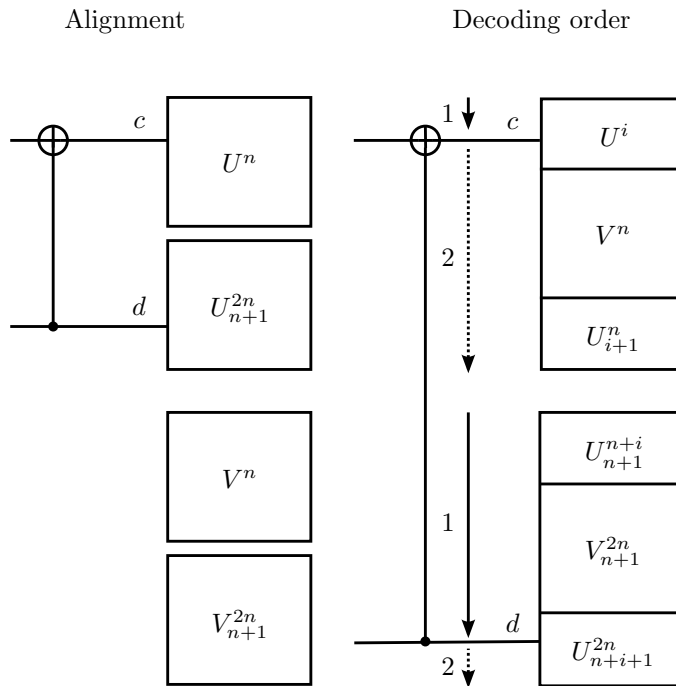


Figure 3.5. First recursion.

We can now apply the technique in Section 3.4.1 to align the incompatible indices of both U 's and V 's. Here, as in the point-to-point case, care must be taken to combine the random variables in a way that guarantees successive decodability. This can be done by aligning either the U 's or V 's, but not both, in any given recursion. As before, only half of the incompatible indices of U 's (or V 's) will be aligned in a single recursion.

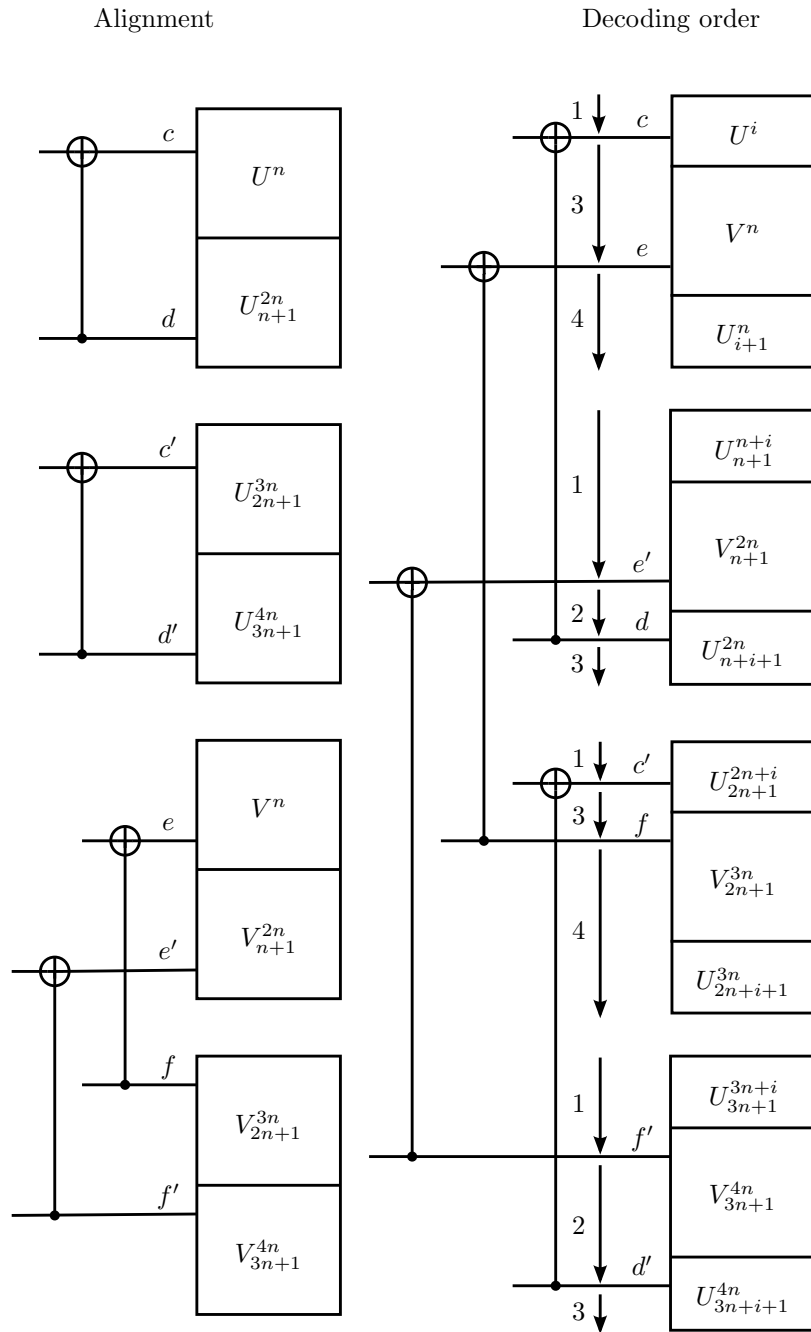


Figure 3.6. Second recursion. Here, $c' = 2n + c$, $d' = 2n + d$, $e' = n + e$, and $f' = n + f$.

Aligning the two index sets alternately over $2m$ recursions, both fractions of incompatible indices can be reduced to $1/2^r$ times their original values.

As an example, consider two recursions of alignment, where incompatible U 's are aligned in the first recursion and incompatible V 's are aligned in the second. Suppose that $\{c\}$ and $\{d\}$ are the type II and type III incompatible indices for U respectively, and $\{e\}$ and $\{f\}$ are the type II and type III incompatible indices for V respectively (Figures 3.5 and 3.6).

In the first recursion, blocks U^n and U_{n+1}^{2n} are aligned, while blocks V^n and V_{n+1}^{2n} are left uncombined (Figure 3.5, left). The decoding order at the each receiver can be identified according to this alignment and the corresponding monotone permutation (U^i, V^n, U_{i+1}^n) (Figure 3.5, right), because the receiver knows the channel and thus which i to pick for polar splitting. The overall decoding order over $4n$ variables is identified in a similar fashion as in the point-to-point alignment (recall Figure 3.3): That is, variables along the solid arrows are decoded before the variables along the dashed arrows.

In the second recursion, The two length- $2n$ blocks (V^n, V_{n+1}^{2n}) and $(V_{2n+1}^{3n}, V_{3n+1}^{4n})$ are aligned while the two length- $2n$ blocks (U^n, U_{n+1}^{2n}) and $(U_{2n+1}^{3n}, U_{3n+1}^{4n})$ are left uncombined (Figure 3.6, left). The decoding order at each receiver can be identified as follows (Figure 3.6, right). Uncombined indices in each block are decoded until reaching a pair of combined indices. Then the pair of combined indices are decoded. This pattern is then repeated until all bits are decoded. More specifically, as depicted in Figure 3.6, variables along arrows with smaller indices are decoded before those with bigger indices.

Note that since in each recursion, only incompatible indices for U 's (or V 's) are combined appropriately, successive decodability is guaranteed as in the point-to-point case. Moreover, the generic decoding order on the right of Figures 3.5 and 3.6 illustrates the general principle that works for both receivers. To identify the specific orders, receiver 1 and 2 start from the monotone permutation S^{2n} and T^{2n} , respectively, and form the permutation S^{8n} and T^{8n} , respectively, based on the above procedure. The

corresponding rate pair (R_1^s, R_2^s) are defined as before

$$R_1^s = \frac{1}{n} \sum_{i \in S_U} I(S_i; Y^{4n} | S^{i-1}),$$

$$R_2^s = \frac{1}{n} \sum_{i \in S_V} I(S_i; Y^{4n} | S^{i-1}).$$

The rate pair (R_1^t, R_2^t) are defined similarly. Clearly, the fraction of incompatible indices for U (and V) is halved in the first (second) recursion.

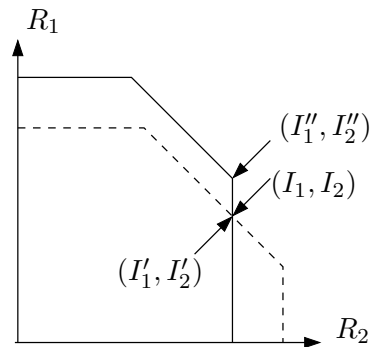


Figure 3.7. Two MAC regions with unequal sum-rates.

To achieve a rate point (I_1, I_2) in the general case as in Figure 4.2, one can find two monotone permutations, which respectively approximate rate pairs (I'_1, I'_2) on the dominant face of one pentagon and (I''_1, I''_2) on the dominant face of the other pentagon such that

$$I_1 \leq \min\{I'_1, I''_1\},$$

$$I_2 \leq \min\{I'_2, I''_2\}.$$

Then, applying the approach above guarantees that I_j fraction of the bit channels are good for both receivers to communicate M_j , $j = 1, 2$. Those bit channels that are good for only one or none of the receivers are fed with frozen valued inputs.

3.5.2 Arbitrary Inputs

Based on the polar coding scheme developed for uniform and independent X and W , one can adapt the method in [66, Section III-D] to design a polar coding scheme for independent *nonuniform* X and W . For *correlated* input distributions $p(q)p(x|q)p(w|q)$, there exist (X', W', Q) jointly independent such that X and W can be represented by functions $x(x', q)$ and $w(w', q)$ (see, for example, [27, Appendix B]). Now consider a new MAC with inputs X' and W' , vector output (Y, Q) , and conditional distribution $P'(y, q|x', w') = p(q)P(y|x(x', q), w(w', q))$, where Q is the common randomness shared at the senders and the receiver. Then the achievable rate region for the new MAC is the set of rate pairs (R_1, R_2) such that

$$\begin{aligned} R_1 &\leq I(X'; Y, Q, W') = I(X; Y, W|Q), \\ R_2 &\leq I(W'; Y, Q, X') = I(W; Y, X|Q), \\ R_1 + R_2 &\leq I(X', W'; Y, Q) = I(X, W; Y|Q) \end{aligned}$$

for distribution $p' = p(q)p(w')p(x')p(x|x', q)p(w|w', q)P(y|x(x', q), w(w', q))$, where $p(x|x', q)$ and $p(w|w', q)$ are $\{0, 1\}$ -valued according to $x(x', q)$ and $w(w', q)$. This rate region is identical to $\mathcal{R}_Y(p)$ as the mutual informations involved are only a function of the joint distribution on (X, W, Y, Q) and $\sum_{x', w'} p'(q, x', w', x, w, y) = p(q, x, w, y)$.

Similarly the rate region $\mathcal{R}_Y(p) \cap \mathcal{R}_Z(p)$ can be described by considering the compound MAC with inputs X' and W' , vector output (Y, Z, Q) , and conditional distribution $P'(y, z, q|x', w') = p(q)P_Y(y|x(x', q), w(w', q))P_Z(z|x(x', q), w(w', q))$. One can apply the method designed for independent inputs to achieve arbitrary points in the rate region of the new compound MAC. To complete the argument, one only needs to show the existence of a good common random sequence q^n , which is shared at the senders and the receiver before the transmission. But this is guaranteed since the average probability of error over all possible choices of q^n is small.

3.5.3 Main Result

Sections 3.5.2 and 3.5.1 together established the following main result of this chapter.

Theorem 3.5.1. *For every $\epsilon > 0$, $\beta < 1/2$, and rate pair (I_1, I_2) in the rate region $\mathcal{R}_Y(p) \cap \mathcal{R}_Z(p)$, there exist $n, m = 2^n n$, and two monotone permutations S^{2m} and T^{2m} with associated rate pairs (R_1^s, R_2^s) and (R_1^t, R_2^t) such that for $j = 1, 2$,*

(i)

$$|\min\{R_j^s, R_j^t\} - I_j| < \epsilon,$$

(ii)

$$\frac{|\mathcal{G}_Y^{(j)} \cap \mathcal{G}_Z^{(j)}|}{m} > \min\{R_j^s, R_j^t\} - \epsilon,$$

where

$$\mathcal{G}_Y^{(1)} = \{i \in \mathcal{S}_U : I(S_i; Y^m | S^{i-1}) > 1 - 2^{-n^\beta}\},$$

$$\mathcal{G}_Z^{(1)} = \{i \in \mathcal{T}_U : I(T_i; Z^m | T^{i-1}) > 1 - 2^{-n^\beta}\},$$

$\mathcal{G}_Y^{(2)}, \mathcal{G}_Z^{(2)}$ are defined similarly by replacing U by V .

This polarization result readily implies the associated polar coding scheme. To design a polar codes for rate pair (I_1, I_2) in the rate region $\mathcal{R}_Y(p) \cap \mathcal{R}_Z(p)$, find two permutations S^{2m} and T^{2m} satisfying (i) and (ii) in the above theorem.

Encoding. Message M_1 is carried by $\{U_i : i \in \mathcal{G}_Y^{(1)} \cap \mathcal{G}_Z^{(1)}\}$ and message M_2 is carried by $\{V_i : i \in \mathcal{G}_Y^{(2)} \cap \mathcal{G}_Z^{(2)}\}$. The remaining indices of U and V are fixed and revealed to the senders and the receivers. Sender 1 transmits $X^m = U^m G_m$ and sender 2 transmits $W^m = V^m G_m$.

Decoding. Upon receiving y^m , receiver 1 applies the standard successive cancellation decoding following the decoding orders of S^{2m} to recover both $\{\hat{U}_i : i \in \mathcal{G}_Y^{(1)} \cap \mathcal{G}_Z^{(1)}\}$

and $\{\hat{V}_i: i \in \mathcal{G}_Y^{(2)} \cap \mathcal{G}_Z^{(2)}\}$. Similarly receiver 2 decodes the received sequence z^m by following the decoding order of T^{2m} .

Rates. The rate pair achieved by this code is

$$\left(\frac{|\mathcal{G}_Y^{(1)} \cap \mathcal{G}_Z^{(1)}|}{m}, \frac{|\mathcal{G}_Y^{(2)} \cap \mathcal{G}_Z^{(2)}|}{m} \right),$$

which, by Theorem 3.5.1, approaches (I_1, I_2) arbitrary closely for large m .

Average probability of error. Since the probability of error in the first ‘polar splitting’ phase is bounded by $O(2^{-n^\beta})$ [5], the probability of error in the compound setting is bounded by $2^r O(2^{-n^\beta})$, which, for fixed r , goes to zero as n tends to infinity.

Complexity. Combining the encoding and decoding complexities of the ‘polar splitting’ phase [5] and the ‘alignment’ phase [36], the complexity in the compound setting is bounded by $O(m \log m)$.

3.6 Interference Networks

Theorem 3.5.1 implies that arbitrary points in the capacity region of the two-user compound MAC are achievable with the proposed polar coding scheme. In the two-user *strong* interference channel, that is, when $I(X; Y, W) \leq I(X; Z, W)$ and $I(W; Z, X) \leq I(W; Y, X)$ for all $p(x)p(w)$, decoding both messages at each receiver is optimal and the two-user compound MAC region coincides with the capacity region of the interference channel. Therefore, the same technique applies to the two-user strong interference channels.

Now we generalize the result to K -sender L -receiver interference networks with input alphabets $\mathcal{X}_1, \dots, \mathcal{X}_K$, and output alphabets $\mathcal{Y}_1, \dots, \mathcal{Y}_L$, and conditional distribution $P(y^L|x^K)$ as depicted in Figure 3.8. Each sender $j \in [1 : K]$ communicates an independent message M_j at rate R_j and each receiver $l \in [1 : L]$ wishes to recover a subset $\mathcal{D}_l \subseteq [1 : K]$ of the messages. A $(2^{nR_1}, \dots, 2^{nR_K}, n)$ code consists of

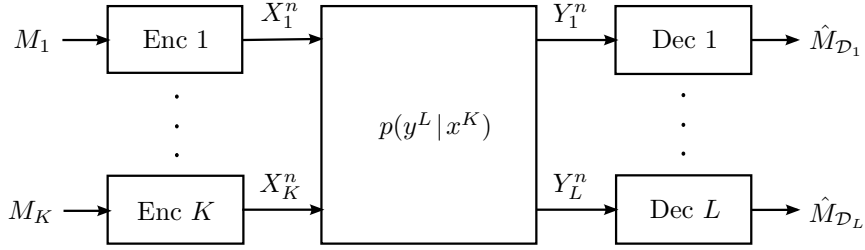


Figure 3.8. K -sender L -receiver interference networks.

- K message sets $[1 : 2^{nR_1}], \dots, [1 : 2^{nR_K}]$,
- K encoders, where encoder $k \in [1 : K]$ assigns a codeword $x_k^n(m_k)$ for each $m_k \in [1 : 2^{nR_k}]$, and
- L decoders, where decoder $l \in [1 : L]$ assigns estimates $\hat{m}_{kl}, k \in \mathcal{D}_l$, or an error e to each received sequence y_l^n .

We assume that the message tuple (M_1, \dots, M_K) is uniformly distributed over $[1 : 2^{nR_1}] \times \dots \times [1 : 2^{nR_K}]$. The average probability of error is defined as

$$P_e^{(n)} = \{\hat{M}_{kl} \neq M_k \text{ for some } l \in [1 : L], k \in \mathcal{D}_l\}.$$

A rate tuple (R_1, \dots, R_K) is *achievable* if there exists a sequence of $(2^{nR_1}, \dots, 2^{nR_K}, n)$ codes with $\lim_{n \rightarrow \infty} P_e^{(n)} = 0$.

The optimal achievable rate region when the encoding is restricted to random coding ensembles [9] is the union over the distribution p and the decoding sets $\{(\mathcal{A}_1, \dots, \mathcal{A}_L) : \mathcal{A}_l \supseteq \mathcal{D}_l, l \in [1 : L]\}$ of the region

$$\bigcap_{l \in [1 : L]} \mathcal{R}_{\mathcal{A}_l}(p), \quad (3.21)$$

where the input distribution is of the form $p = (\prod_{j=1}^K p(x_j)) P(y^L | x^K)$ and $\mathcal{R}_{\mathcal{A}_l}(p)$ is

the set of rate tuples $(R_j: j \in \mathcal{A}_l)$ such that

$$R(\mathcal{J}) \leq I(X_{\mathcal{J}}; Y_l, X_{\mathcal{A}_l \setminus \mathcal{J}}) \quad (3.22)$$

for all $\mathcal{J} \subseteq \mathcal{A}_l$. Here we introduce notation

$$R(\mathcal{J}) := \sum_{j \in \mathcal{J}} R_j \quad (3.23)$$

and

$$X_{\mathcal{J}} := (X_j: j \in \mathcal{J}) \quad (3.24)$$

for an index set \mathcal{J} . It is clear from (3.21) that this rate region is also a compound MAC region.

To apply the proposed polar coding scheme to the interference networks, one needs to (i) generalize Arikan's polar splitting result to K -user MAC and (ii) align more than two incompatible polarization processes, each of which involves codes from K users. We prove (i) in Section 3.6.1 and discuss (ii) in Section 3.6.2. We show the achievability of the Han–Kobayashi inner bound in Section 4.4.

3.6.1 ‘Polar Splitting’ for K -User MAC

Consider a K -user MAC, where transmitter $j, j \in [1:K]$, wishes to communicate a message M_j reliably to the receiver by sending a codeword $X_j^n(M_j) = (X_{j1}, X_{j2}, \dots, X_{jn})$ over the memoryless channel $P(y|x^K)$. The receiver wishes to recover all the messages $M_{[1:K]}$. The capacity region of the K -user MAC is described by

$$\bigcup_p \mathcal{R}_{[1:K]}(p),$$

where the union is over all distributions of the form $p = p(q)(\prod_{i=1}^K p(x_i|q))P(y|x^K)$, and $\mathcal{R}_{[1:K]}(p)$ is defined as in (3.22).

Let $U_j^n = X_j^n G_n$ for $j \in [1 : K]$. Similar to the two-user MAC case, we have the chain rule of the form

$$\sum_{i=1}^{Kn} I(S_i; Y^n | S^{i-1}),$$

where S^{Kn} is a *monotone* permutation of (U_1^n, \dots, U_K^n) , i.e., elements of U_j^n appear in increasing order in S^{Kn} for all $j \in [1 : K]$. Let \mathcal{S}_j denote the index set $\{i : S_i = U_{jk} \text{ for some } k\}$. Define the associated rate tuple (R_1, \dots, R_K) of the monotone permutation as

$$R_j = \frac{1}{n} \sum_{i \in \mathcal{S}_j} I(S_i; Y^n | S^{i-1})$$

for $j \in [1 : K]$. We now generalize Arıkan's polar-splitting result to K users.

Proposition 3.6.1. *For every $\epsilon > 0, \beta < 1/2$, and rate tuple (I_1, \dots, I_K) on the dominant face of $\mathcal{R}_{[1:K]}(p)$, there exists an n and a monotone permutation S^{Kn} such that for all $j \in [1 : K]$,*

(i)

$$|R_j - I_j| \leq \epsilon,$$

(ii)

$$\frac{|\mathcal{G}^{(j)}|}{n} > R_j - \epsilon,$$

where

$$\mathcal{G}^{(j)} = \{i \in \mathcal{S}_j : I(S_i; Y^n | S^{i-1}) > 1 - 2^{-n^\beta}\}.$$

Proof. We prove statement (i) by induction. The case $K = 2$ holds by Proposition 3.4.1. Suppose the statement holds for up to $K - 1$. We prove the statement for K .

Assume without loss of generality that we start by decoding $U_1^{i_0}$ for some $i_0 \in [1 : n]$. We specify i_0 by the following procedure. Let i increase from 0 to n and consider the quantities

$$\frac{1}{n} I(U_{\mathcal{J}}^n; Y^n, U_1^i) \tag{3.25}$$

for each $\mathcal{J} \subseteq [2 : n]$, where $U_{\mathcal{J}}^n := (U_j^n : j \in \mathcal{J})$. Some observations follow:

1. As i increases, each mutual information term increases by at most $1/n$ in each step, since the increment is $I(U_{1i}; U_{\mathcal{J}}^n | Y^n, U_1^{i-1})/n \leq 1/n$.
2. There exists an i such that for at least one $\mathcal{J} \subseteq [2 : K]$, the following is violated

$$\frac{1}{n}I(U_{\mathcal{J}}^n; Y^n, U_1^i) < I(\mathcal{J}), \quad (3.26)$$

where $I(\mathcal{J}) := \sum_{j \in \mathcal{J}} I_j$.

To see 2), set $U_1^i = \emptyset$ and $U_1^i = U_1^n$ respectively. We have

$$\begin{aligned} \frac{1}{n}I(U_{\mathcal{J}}^n; Y^n) &\leq I(\mathcal{J}) \quad \text{for } \mathcal{J} \subset [2 : K], \\ \frac{1}{n}I(U_{[2:K]}^n; Y^n) &\leq I([2 : K]) \leq I(U_{[2:K]}^n; Y^n, U_1^n). \end{aligned} \quad (3.27)$$

As i increases, the mutual information terms in (3.25) increase steadily. In particular, the second inequality in (3.27) guarantees that there exists an i such that (3.26) is violated for some $\mathcal{J} \subseteq [2 : K]$ and $\mathcal{J} \neq \emptyset$. Take the smallest such i as i_0 .

Suppose at $i = i_0$, for index set $\mathcal{J}_0 \subseteq [2 : K]$, the inequality in (3.26) is violated. As the increment on the left-hand-side of (3.26) is bounded by $1/n$, we have

$$\frac{1}{n}I(U_{\mathcal{J}_0}^n; Y^n, U_1^{i_0}) = I(\mathcal{J}_0) + \delta, \quad (3.28)$$

where $0 \leq \delta < 1/n$. This divides the K -dimensional rate-approximation into two sub-problems of smaller dimensions.

Problem 1: For $\mathcal{J} \subseteq \mathcal{J}_0$, we have

$$\frac{1}{n}I(U_{\mathcal{J}}^n; Y^n, U_1^{i_0}) < I(\mathcal{J}) \quad \text{for all } \mathcal{J} \subset \mathcal{J}_0$$

and

$$\frac{1}{n}I(U_{\mathcal{J}_0}^n; Y^n, U_1^{i_0}) = I(\mathcal{J}_0) + \delta.$$

Treating $(Y^n, U_1^{i_0})$ as the output of the MAC, This is a rate-approximation problem for the rate tuple $(I_j: j \in \mathcal{J}_0)$ on the dominant face of a $|\mathcal{J}_0|$ -dimension polyhedron. Note that $\emptyset \neq \mathcal{J}_0 \subseteq [2:K]$, and thus $1 \leq |\mathcal{J}_0| \leq K-1$. We can therefore approximate the rate tuple $(I_j: j \in \mathcal{J}_0)$ within ϵ distance due to the induction hypothesis.

Problem 2: To handle the rate indices outside of $\mathcal{J}_0 \cup \{1\}$, consider for all $\mathcal{J} \supseteq \mathcal{J}_0$, and subtract (3.28) from (3.26). Let $\mathcal{T} = \mathcal{J} \setminus \mathcal{J}_0$, $\mathcal{T}_0 = [2:K] \setminus \mathcal{J}_0$, and $I'_1 = I_1 - \frac{1}{n}I(U_1^{i_0}; Y^n)$. This yields

$$\begin{aligned} \frac{1}{n}I(U_{\mathcal{T}}^n; Y^n, U_{\mathcal{J}_0}^n, U_1^{i_0}) &\leq I(\mathcal{T}) - \delta, \\ \frac{1}{n}I(U_{\mathcal{T}}^n, U_{1,i_0+1}^n; Y^n, U_{\mathcal{J}_0}^n, U_1^{i_0}) &\leq I(\mathcal{T}) + I'_1 - \delta, \\ \frac{1}{n}I(U_{\mathcal{T}_0}^n, U_{1,i_0+1}^n; Y^n, U_{\mathcal{J}_0}^n, U_1^{i_0}) &= I(\mathcal{T}_0) + I'_1 - \delta. \end{aligned}$$

Treating $(Y^n, U_1^{i_0}, U_{\mathcal{J}_0}^n)$ as the output of the MAC, this is a rate-approximation problem for the rate tuple $(I'_1, (I_j: j \in \mathcal{T}_0))$ on the dominant face of a $(K - |\mathcal{J}_0|)$ -dimensional polyhedron. Note that $1 \leq K - |\mathcal{J}_0| \leq K-1$ and therefore we can again approximate the rate tuple $(I'_1, (I_j: j \in \mathcal{T}_0))$ within ϵ distance by the induction hypothesis.

The cumulative approximation error is bounded by $K(\epsilon + \delta) \leq K(\epsilon + 1/n)$, which can be made arbitrarily small by choosing a small ϵ and a large n . The final permutation is obtained by cascading $U_1^{i_0}$, $S^{|\mathcal{J}_0|n}$ (the solution from problem 1), and $T^{K-|\mathcal{J}_0|n-i_0}$ (the solution from problem 2).

The polarization result (ii) is obtained by the standard arguments as in [5]. This concludes the proof. \square

3.6.2 Aligning Polarized Indices for K Users

Suppose we have two monotone permutations for two K -user MACs. To align the incompatible indices for all users, one can continue the method in Section 3.5.1 and sequentially align the incompatible indices for each $U_j^n, j \in [1 : K]$. After alternately aligning K index sets over Km recursions, the fraction of the incompatible indices for each user is reduced to $1/2^r$ times the original fraction. The method for aligning L monotone permutations can be done by recursively aligning two permutations as in [36].

3.6.3 Han–Kobayashi Inner Bound

As an important special case, we show how the scheme above can be used to achieve the Han–Kobayashi inner bound, the best known inner bound for general two-user interference channels $P(y_1, y_2 | x_1, x_2)$.

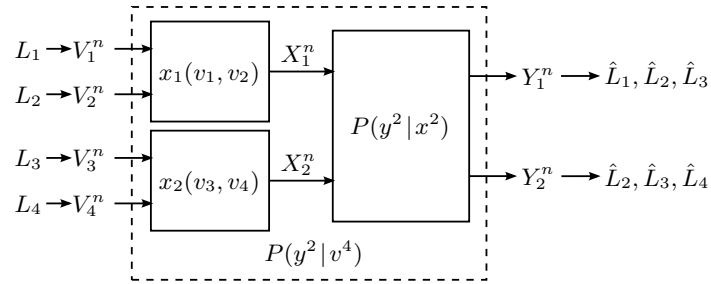


Figure 3.9. Han–Kobayashi coding scheme.

The Han–Kobayashi coding scheme is illustrated in Figure 4.9. Message M_1 is split into two independent parts (L_1, L_2) with rates (R'_1, R'_2) and message M_2 is split into two independent parts (L_3, L_4) with rates (R'_3, R'_4) . Message $L_j, j \in [1 : 4]$, is carried by codeword $V_j^n(L_j)$. Then the channel inputs X_1^n and X_2^n are formed using two symbol-by-symbol mappings $x_1(v_1, v_2)$ and $x_2(v_3, v_4)$. Receiver 1 uniquely decodes $(\hat{L}_1, \hat{L}_2, \hat{L}_3)$ upon receiving Y_1^n , while receiver 2 uniquely decodes $(\hat{L}_2, \hat{L}_3, \hat{L}_4)$ upon receiving Y_2^n . For a fixed input distribution p , the Han–Kobayashi inner bound can be expressed with

auxiliary rate tuple (R'_1, R'_2, R'_3, R'_4) as

$$\mathcal{R}_1(p) \cap \mathcal{R}_2(p). \quad (3.29)$$

Here the input distribution is of the form

$$p = p(q) \left(\prod_{j=1}^4 p(v_j | q) \right) p(x_1 | v_1, v_2, q) p(x_2 | v_3, v_4, q) P(y_1, y_2 | x_1, x_2),$$

where $p(x_1 | v_1, v_2, q)$ and $p(x_2 | v_3, v_4, q)$ are $\{0, 1\}$ -valued according to functions $x_1(v_1, v_2, q)$ and $x_2(v_3, v_4, q)$. The rate region $\mathcal{R}_1(p)$ is the set of rate triples (R'_1, R'_2, R'_3) such that

$$R'_{\mathcal{J}} \leq I(V_{\mathcal{J}}; Y_1^n, V_{[1:3] \setminus \mathcal{J}} | Q)$$

for all $\mathcal{J} \subseteq [1:3]$. The rate region $\mathcal{R}_2(p)$ is the set of rate triples (R'_2, R'_3, R'_4) such that

$$R'_{\mathcal{J}} \leq I(V_{\mathcal{J}}; Y_2^n, V_{[2:4] \setminus \mathcal{J}} | Q)$$

for all $\mathcal{J} \subseteq [2:4]$.

It is clear from the Han–Kobayashi coding scheme that for each pair of functions $x_1(v_1, v_2)$ and $x_2(v_3, v_4)$, the message splitting transforms the original two-user interference channel into a four-sender two-receiver interference network

$$P(y^2 | v^4) = P(y_1, y_2 | x_1(v_1, v_2), x_2(v_3, v_4)),$$

where sender $j \in \{1, 2, 3, 4\}$ communicates an independent message L_j at rate R'_j , receiver 1 recovers the subset $\mathcal{D}_1 = \{1, 2, 3\}$ of the four messages, and receiver 2 recovers the subset $\mathcal{D}_2 = \{2, 3, 4\}$ of the four messages.

Note from expression (4.17) that the auxiliary rate region (R'_1, R'_2, R'_3, R'_4) is the intersection of two 3-dimensional MAC regions, two dimensions of which are in common.

Therefore, one just needs to find two monotone permutations that achieves any target point in the two MACs respectively and sequentially align the two codes shared in common using the method in Section 3.5.1.

3.7 Discussion

We have shown a polar coding method for general interference networks that achieves the optimal rate region when the encoding is restricted to random coding ensembles with superposition coding and time sharing [9]. As special cases, the method achieves the capacity region of the compound MAC and the Han–Kobayashi inner bound for two-user interference channels.

One drawback of the current method is the long blocklength needed for large scale networks. When there are L receivers in the network, one needs to do $L - 1$ alignments to resolve the incompatible indices in L permutations, which makes the blocklength scale with the network size.

One crucial component in the current method is Arikan’s ‘polar splitting’ for MAC. It would be interesting to compare it to regular rate splitting for MAC as in [32]. Both schemes achieve optimal performance in MAC. However, for interference channels, the former, together with the alignment method, achieves the best known rate region while the latter is strictly suboptimal information theoretically [77].

In the next chapter, a successive decoding based random coding scheme is presented, which also achieves Han–Kobayashi inner bound. Some similarities and connections can be found in the way the two schemes resolve the incompatibility of the two MACs.

Acknowledgment

This chapter is, in part, a reprint of the material in the papers: Lele Wang and Eren Şaşoğlu, “Polar coding for interference networks,” submitted to *IEEE Trans-*

actions on Information Theory, 2014; and Lele Wang, Eren Şaşoğlu, and Young-Han Kim, “Sliding-window superposition coding for interference networks,” to be submitted to *IEEE Transactions on Information Theory*.

Chapter 4

Interference Channels: COTS Codes

In Chapter 3, we showed that single block rate-splitting is strictly suboptimal in achieving the simultaneous decoding inner bound in interference channels. This fact makes us wonder whether point-to-point coding techniques, which can achieve capacity for multiple access and single-antenna Gaussian broadcast channels, are fundamentally deficient for the interference channel. Interestingly, using polar codes, a channel coding technique that was initially invented for the point-to-point coding, the simultaneous decoding inner bound was shown to be achievable, even with a low-complexity successive cancellation decoding algorithm. How does the polar coding solution circumvent the incompatibility that is shown to be unresolvable in rate-splitting? Does the polar coding solution introduce a fundamentally new technique for the network communication? Can there be a “random coding equivalence” of the polar coding solution? These questions motivate the sliding-window superposition coding (SWSC) scheme that will be presented in this Chapter.

4.1 Sliding-Window Superposition Coding

In this section, we present the SWSC scheme for the two-user general discrete memoryless interference channels. We first illustrate the simplest SWSC scheme in Section 4.1.1, which is sufficient to achieve the corner point on the simultaneous decoding inner bound with only two superposition layers at one user. In Section 4.1.2, we show how to achieve an arbitrary rate point by having more superposition layers.

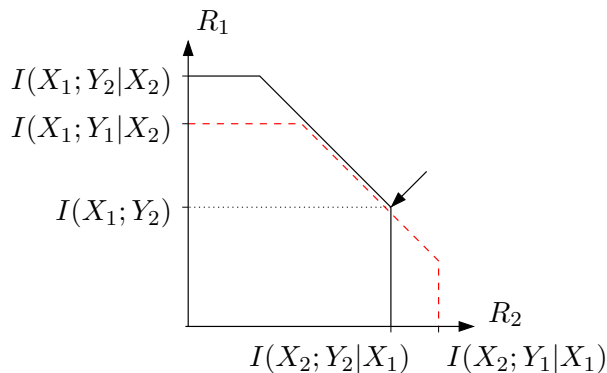


Figure 4.1. The simultaneous decoding inner bound and the corner point (4.1) (the one with the arrow) that will be illustrated to be achievable by SWSC.

4.1.1 Corner Point: The First Illustration

As the simplest illustration, we first describe how to achieve the corner point

$$(R_1, R_2) = (I(X_1; Y_2), I(X_2; Y_2|X_1)) \quad (4.1)$$

of the simultaneous decoding inner bound when $Q = \emptyset$ and the sum-rates are equal, i.e.,

$$I(X_1, X_2; Y_1) = I(X_1, X_2; Y_2) \quad (4.2)$$

as shown in Figure 4.1. For the symmetric Gaussian interference channel with $S < I < S(S+1)$, this is the exact same point that demonstrated the insufficiency of single-block rate-splitting in Section 3.2.

Theorem 4.1.1. *A rate pair (R_1, R_2) is achievable with the SWSC scheme if*

$$R_1 < \min\{I(U; Y_1) + I(X_1; Y_1|U, X_2), I(X_1; Y_2)\} := I_1,$$

$$R_2 < \min\{I(X_2; Y_1|U), I(X_2; Y_2|X_1)\} := I_2$$

for some pmf $p(u, x_1)p(x_2)$. In addition, there exists a pmf $p(u, x_1)p(x_2)$ such that $(I_1, I_2) = (I(X_1; Y_2), I(X_2; Y_2|X_1))$; in other words, the corner point (4.1) is achievable.

block j	1	2	3	...	$b-1$	b
U	1	$m_1(1)$	$m_1(2)$	$m_1(b-1)$
X_1	$m_1(1)$	$m_1(2)$	$m_1(b-1)$	1
X_2	$m_2(1)$	$m_2(2)$	$m_2(b)$
Receiver 1	\emptyset	$\hat{m}_1(1)$	$\hat{m}_1(2)$	$\hat{m}_1(b-1)$
	$\hat{m}_2(1)$	$\hat{m}_2(2)$	$\hat{m}_2(3)$	$\hat{m}_2(b)$
Receiver 2	\emptyset	$\hat{m}_1(1)$	$\hat{m}_1(2)$	$\hat{m}_1(b-1)$
	\emptyset	$\hat{m}_2(1)$	$\hat{m}_2(2)$	$\hat{m}_2(b-1), \hat{m}_2(b)$

Table 4.1. Sliding-window superposition coding scheme.

We now describe our coding scheme. Roughly speaking, instead of splitting the message M_1 into two parts and recovering the two parts separately, we send M_1 without split over two consecutive blocks and recover it using sliding-window decoding. A sequence of $(b-1)$ messages $M_1(j), j \in [1 : b-1]$ and b messages $M_2(j), j \in [1 : b]$ are transmitted over b blocks. The average achievable rate pair over b blocks is $(\frac{b-1}{b}R_1, R_2)$, which can be made as close to (R_1, R_2) as desired. Details are as follows.

Codebook generation. We use superposition coding. Fix the pmf $p(u, x_1)p(x_2)$ that attains the target rate pair. Randomly and independently generate a codebook for each block. For notational convention, we assume $m_1(0) = m_1(b) = 1$. For $j \in [1 : b]$, randomly and independently generate 2^{nR_1} sequences $u^n(m_1(j-1)), m_1(j-1) \in [1 : 2^{nR_1}]$, each according to a product of $p(u)$. For each $m_1(j-1)$, randomly and conditionally

independently generate 2^{nR_1} sequences $x_1^n(m_1(j)|m_1(j-1)), m_1(j) \in [1 : 2^{nR_1}]$, each according to a product of $p(x_1|u)$. Randomly and independently generate 2^{nR_2} sequences $x_2^n(m_2(j)), m_2(j) \in [1 : 2^{nR_2}]$, each according to a product of $p(x_2)$. This defines the codebook

$$\mathcal{C}_j = \{u^n(m_1(j-1)), x_1^n(m_1(j)|m_1(j-1)), x_2^n(m_2(j)) : \\ m_1(j-1), m_1(j) \in [1 : 2^{nR_1}], m_2(j) \in [1 : 2^{nR_2}]\}, \quad j \in [1 : b].$$

Encoding. Sender 1 transmits $x_1^n(m_1(j)|m_1(j-1))$ and sender 2 transmits $x_2^n(m_2(j))$ in block $j \in [1 : b]$.

Decoding. Decoder 1 successively recovers $\hat{m}_2(1) \rightarrow \hat{m}_1(1) \rightarrow \dots \rightarrow \hat{m}_2(b-1) \rightarrow \hat{m}_1(b-1) \rightarrow \hat{m}_2(b)$ and decoder 2 successively recovers $\hat{m}_1(1) \rightarrow \hat{m}_2(1) \rightarrow \dots \rightarrow \hat{m}_1(b-1) \rightarrow \hat{m}_2(b-1) \rightarrow \hat{m}_2(b)$, where the decoding of $\hat{m}_1(j)$ is done by a sliding-window decoding over blocks j and $j+1$. Table 4.1 reveals the scheduling of the messages.

Let the received sequences in block j be $y_1^n(j)$ and $y_2^n(j)$, $j \in [1 : b]$. For receiver 1, at the end of block 1, it finds the unique message $\hat{m}_2(1)$ such that

$$(x_2^n(\hat{m}_2(1)), y_1^n(1), u^n(m_1(0))) \in \mathcal{T}_\epsilon^{(n)}.$$

At the end of block $j+1$, $j \in [1 : b-1]$, it finds the unique message $\hat{m}_1(j)$ such that

$$(u^n(\hat{m}_1(j-1)), x_1^n(\hat{m}_1(j)|\hat{m}_1(j-1)), x_2^n(\hat{m}_2(j)), y_1^n(j)) \in \mathcal{T}_\epsilon^{(n)}$$

and

$$(u^n(\hat{m}_1(j)), y_1^n(j+1)) \in \mathcal{T}_\epsilon^{(n)}$$

simultaneously. Then it finds the unique $\hat{m}_2(j+1)$ such that

$$(x_2^n(\hat{m}_2(j+1)), y_1^n(j+1), u^n(\hat{m}_1(j))) \in \mathcal{T}_\epsilon^{(n)}.$$

If any of the typicality checks fails, it declares an error.

For receiver 2, at the end of block $j + 1$, $j \in [1 : b - 1]$, it finds the unique $\hat{m}_1(j)$ such that

$$(u^n(\hat{m}_1(j-1)), x_1^n(\hat{m}_1(j)|\hat{m}_1(j-1)), y_2^n(j)) \in \mathcal{T}_\epsilon^{(n)}$$

and

$$(u^n(\hat{m}_1(j)), y_2^n(j+1)) \in \mathcal{T}_\epsilon^{(n)}$$

simultaneously. Then it finds the unique $\hat{m}_2(j)$ such that

$$(x_2^n(\hat{m}_2(j)), y_2^n(j), u^n(\hat{m}_1(j-1)), x_1^n(\hat{m}_1(j)|\hat{m}_1(j-1))) \in \mathcal{T}_\epsilon^{(n)}.$$

In the end, receiver 2 finds the unique $\hat{m}_2(b)$ such that

$$(x_2^n(\hat{m}_2(b)), y_2^n(b), u^n(\hat{m}_1(b-1)), x_1^n(1|\hat{m}_1(b-1))) \in \mathcal{T}_\epsilon^{(n)}.$$

If any of the typicality checks fails, it declares an error.

Analysis of the probability of error. We analyze the probability of decoding error averaged over codebooks. Assume without loss of generality that $M_1(j) = M_2(j) = 1$. First consider receiver 1. We divide the error events as follows

$$\mathcal{E}_{11}(j-1) = \{\hat{M}_1(j-1) \neq 1\},$$

$$\mathcal{E}_{12}(j) = \{\hat{M}_2(j) \neq 1\},$$

$$\begin{aligned} \mathcal{E}_{13}(j) = \{ & (U^n(\hat{M}_1(j-1)), X_1^n(1|\hat{M}_1(j-1)), X_2^n(\hat{M}_2(j)), Y_1^n(j)) \notin \mathcal{T}_\epsilon^{(n)} \\ & \text{or } (U^n(1), Y_1^n(j+1)) \notin \mathcal{T}_\epsilon^{(n)}\}, \end{aligned}$$

$$\begin{aligned} \mathcal{E}_{14}(j) = \{ & (U^n(\hat{M}_1(j-1)), X_1^n(m_1(j)|\hat{M}_1(j-1)), X_2^n(\hat{M}_2(j)), Y_1^n(j)) \in \mathcal{T}_\epsilon^{(n)} \\ & \text{and } (U^n(m_1(j)), Y_1^n(j+1)) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_1(j) \neq 1\}, \end{aligned}$$

$$\mathcal{E}_{15}(j+1) = \{(X_2^n(1), Y_1^n(j+1), U^n(\hat{M}_1(j))) \notin \mathcal{T}_\epsilon^{(n)}\},$$

$$\mathcal{E}_{16}(j+1) = \{(X_2^n(m_2(j+1)), Y_1^n(j+1), U^n(\hat{M}_1(j))) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_2(j+1) \neq 1\}.$$

We analyze by induction. By assumption $\mathcal{E}_{11}(0) = \emptyset$. Thus in block 1, the probability of error is upper bounded as

$$\begin{aligned} \mathbb{P}\{\hat{M}_2(1) \neq 1\} &= \mathbb{P}(\mathcal{E}_{12}(1)) \\ &\leq \mathbb{P}(\mathcal{E}_{15}(1)) + \mathbb{P}(\mathcal{E}_{16}(1)). \end{aligned}$$

Now by the law of large numbers, $\mathbb{P}(\mathcal{E}_{15}(1)) \rightarrow 0$ as $n \rightarrow \infty$. By the packing lemma, $\mathbb{P}(\mathcal{E}_{16}(1)) \rightarrow 0$ as $n \rightarrow \infty$ if $R_2 < I(X_2; Y_1|U) - \delta(\epsilon)$. Now assume that the probability of error $\mathbb{P}(\mathcal{E}_{11}(j-1) \cup \mathcal{E}_{12}(j))$ in block j tends to zero as $n \rightarrow \infty$. In block $j+1$, the probability of error is upper bounded as

$$\begin{aligned} &\mathbb{P}\{(\hat{M}_1(j), \hat{M}_2(j+1)) \neq (1, 1)\} \\ &\leq \mathbb{P}(\mathcal{E}_{11}(j-1) \cup \mathcal{E}_{12}(j) \cup \mathcal{E}_{11}(j) \cup \mathcal{E}_{12}(j+1)) \\ &\leq \mathbb{P}(\mathcal{E}_{11}(j-1) \cup \mathcal{E}_{12}(j)) + \mathbb{P}(\mathcal{E}_{11}(j) \cap \mathcal{E}_{11}^c(j-1) \cap \mathcal{E}_{12}^c(j)) + \mathbb{P}(\mathcal{E}_{12}(j+1) \cap \mathcal{E}_{11}^c(j)) \\ &\leq \mathbb{P}(\mathcal{E}_{11}(j-1) \cup \mathcal{E}_{12}(j)) + \mathbb{P}(\mathcal{E}_{13}(j) \cap \mathcal{E}_{11}^c(j-1) \cap \mathcal{E}_{12}^c(j)) \\ &\quad + \mathbb{P}(\mathcal{E}_{14}(j) \cap \mathcal{E}_{11}^c(j-1) \cap \mathcal{E}_{12}^c(j)) + \mathbb{P}(\mathcal{E}_{15}(j+1) \cap \mathcal{E}_{11}^c(j)) + \mathbb{P}(\mathcal{E}_{16}(j+1) \cap \mathcal{E}_{11}^c(j)). \end{aligned}$$

By the induction assumption, the first term tends to zero as $n \rightarrow \infty$. By the independence of the codebooks, the law of large numbers, and the packing lemma, the second, fourth, and fifth terms tend to zero as $n \rightarrow \infty$ if $R_2 < I(X_2; Y_1|U) - \delta(\epsilon)$. The third term $\mathbb{P}(\mathcal{E}_{14}(j) \cap \mathcal{E}_{11}^c(j-1) \cap \mathcal{E}_{12}^c(j))$ requires a special care. We have

$$\begin{aligned} &\mathbb{P}(\mathcal{E}_{14}(j) \cap \mathcal{E}_{11}^c(j-1) \cap \mathcal{E}_{12}^c(j)) \\ &= \mathbb{P}\{(U^n(1), X_1^n(m_1(j)|1), X_2^n(1), Y_1^n(j)) \in \mathcal{T}_\epsilon^{(n)} \text{ and } (U^n(m_1(j)), Y^n(j+1)) \in \mathcal{T}_\epsilon^{(n)} \\ &\quad \text{for some } m_1(j) \neq 1\} \\ &= \sum_{m_1(j) \neq 1} \mathbb{P}\{(U^n(1), X_1^n(m_1(j)|1), X_2^n(1), Y_1^n(j)) \in \mathcal{T}_\epsilon^{(n)} \text{ and} \\ &\quad (U^n(m_1(j)), Y_1^n(j+1)) \in \mathcal{T}_\epsilon^{(n)}\} \end{aligned}$$

$$\begin{aligned}
&\stackrel{(a)}{=} \sum_{m_1(j) \neq 1} \mathbb{P}\{(U^n(1), X_1^n(m_1(j)|1), X_2^n(1), Y_1^n(j)) \in \mathcal{T}_\epsilon^{(n)}\} \\
&\quad \cdot \mathbb{P}\{(U^n(m_1(j)), Y_1^n(j+1)) \in \mathcal{T}_\epsilon^{(n)}\} \\
&\stackrel{(b)}{\leq} 2^{nR_1} 2^{-n(I(X_1; Y_1|X_2, U) - \delta(\epsilon))} 2^{-n(I(U; Y_1) - \delta(\epsilon))},
\end{aligned}$$

which tends to zero if $R_1 < I(U; Y_1) + I(X_1; Y_1|X_2, U) - 2\delta(\epsilon)$. Here (a) follows since, by the independence of the codebooks, the events

$$\{(U^n(1), X_1^n(m_1(j)|1), X_2^n(1), Y_1^n(j)) \in \mathcal{T}_\epsilon^{(n)}\}$$

and

$$\{(U^n(m_1(j)), Y_1^n(j+1)) \in \mathcal{T}_\epsilon^{(n)}\}$$

are independent for each $m_1(j) \neq 1$, and (b) follows by the independence of the codebooks and the joint typicality lemma.

For receiver 2, we divide the error events as follows

$$\begin{aligned}
\mathcal{E}_{21}(j) &= \{\hat{M}_1(j) \neq 1\}, \\
\mathcal{E}_{22}(j) &= \{\hat{M}_2(j) \neq 1\}, \\
\mathcal{E}_{23}(j) &= \{(U^n(\hat{M}_1(j-1)), X_1^n(1|\hat{M}_1(j-1)), Y_2^n(j)) \notin \mathcal{T}_\epsilon^{(n)} \\
&\quad \text{or } (U^n(1), Y_2^n(j+1)) \notin \mathcal{T}_\epsilon^{(n)}\}, \\
\mathcal{E}_{24}(j) &= \{(U^n(\hat{M}_1(j-1)), X_1^n(m_1(j)|\hat{M}_1(j-1)), Y_2^n(j)) \in \mathcal{T}_\epsilon^{(n)} \text{ and} \\
&\quad (U^n(m_1(j)), Y_2^n(j+1)) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_1(j) \neq 1\}, \\
\mathcal{E}_{25}(j) &= \{(X_2^n(1), Y_2^n(j), X_1^n(\hat{M}_1(j)|\hat{M}_1(j-1)), U^n(\hat{M}_1(j-1))) \notin \mathcal{T}_\epsilon^{(n)}\}, \\
\mathcal{E}_{26}(j) &= \{(X_2^n(m_2(j)), Y_2^n(j), X_1^n(\hat{M}_1(j)|\hat{M}_1(j-1)), U^n(\hat{M}_1(j-1))) \in \mathcal{T}_\epsilon^{(n)} \\
&\quad \text{for some } m_2(j) \neq 1\}.
\end{aligned}$$

We analyze by induction. In block 1, since $\mathcal{E}_{21}(0) = \emptyset$ by assumption, $\mathbb{P}\{\hat{M}_1(0) \neq$

$1\} = 0$. Now assume that in block j , the probability of error $\mathbb{P}\{(\hat{M}_1(j-1), \hat{M}_2(j-1)) \neq (1, 1)\} = \mathbb{P}(\mathcal{E}_{21}(j-1) \cup \mathcal{E}_{22}(j-1))$ tends to zero as $n \rightarrow \infty$, then in block $j+1$, we have

$$\begin{aligned} & \mathbb{P}\{(\hat{M}_1(j), \hat{M}_2(j)) \neq (1, 1)\} \\ & \leq \mathbb{P}(\mathcal{E}_{21}(j-1) \cup \mathcal{E}_{21}(j) \cup \mathcal{E}_{22}(j)) \\ & \leq \mathbb{P}(\mathcal{E}_{21}(j-1)) + \mathbb{P}(\mathcal{E}_{21}^c(j-1) \cap \mathcal{E}_{21}(j)) + \mathbb{P}(\mathcal{E}_{21}^c(j-1) \cap \mathcal{E}_{21}^c(j) \cap \mathcal{E}_{22}(j)) \\ & \leq \mathbb{P}(\mathcal{E}_{21}(j-1)) + \mathbb{P}(\mathcal{E}_{21}^c(j-1) \cap \mathcal{E}_{23}(j)) + \mathbb{P}(\mathcal{E}_{21}^c(j-1) \cap \mathcal{E}_{24}(j)) \\ & \quad + \mathbb{P}(\mathcal{E}_{21}^c(j-1) \cap \mathcal{E}_{21}^c(j) \cap \mathcal{E}_{25}(j)) + \mathbb{P}(\mathcal{E}_{21}^c(j-1) \cap \mathcal{E}_{21}^c(j) \cap \mathcal{E}_{26}(j)). \end{aligned}$$

By the induction assumption, the first term $\mathbb{P}(\mathcal{E}_{21}(j-1))$ tends to zero as $n \rightarrow \infty$. By the law of large numbers, the second and fourth terms $\mathbb{P}(\mathcal{E}_{21}^c(j-1) \cap \mathcal{E}_{23}(j))$ and $\mathbb{P}(\mathcal{E}_{21}^c(j-1) \cap \mathcal{E}_{21}^c(j) \cap \mathcal{E}_{25}(j))$ tend to zero as $n \rightarrow \infty$. By the packing lemma, the last term $\mathbb{P}(\mathcal{E}_{21}^c(j-1) \cap \mathcal{E}_{21}^c(j) \cap \mathcal{E}_{26}(j))$ tends to zero as $n \rightarrow \infty$ if $R_2 < I(X_2; Y_2 | X_1) - \delta(\epsilon)$. The third term requires a special care. We have

$$\begin{aligned} & \mathbb{P}(\mathcal{E}_{21}^c(j-1) \cap \mathcal{E}_{24}(j)) \\ & = \mathbb{P}\{(U^n(1), X_1^n(m_1(j)|1), Y_2^n(j)) \in \mathcal{T}_\epsilon^{(n)} \text{ and } (U^n(m_1(j)), Y_2^n(j+1)) \in \mathcal{T}_\epsilon^{(n)} \\ & \quad \text{for some } m_1(j) \neq 1\} \\ & = \sum_{m_1(j) \neq 1} \mathbb{P}\{(U^n(1), X_1^n(m_1(j)|1), Y_2^n(j)) \in \mathcal{T}_\epsilon^{(n)} \text{ and } (U^n(m_1(j)), Y_2^n(j+1)) \in \mathcal{T}_\epsilon^{(n)}\} \\ & \stackrel{(a)}{=} \sum_{m_1(j) \neq 1} \mathbb{P}\{(U^n(1), X_1^n(m_1(j)|1), Y_2^n(j)) \in \mathcal{T}_\epsilon^{(n)}\} \cdot \mathbb{P}\{(U^n(m_1(j)), Y_2^n(j+1)) \in \mathcal{T}_\epsilon^{(n)}\} \\ & \stackrel{(b)}{\leq} 2^{nR_1} 2^{-n(I(X_1; Y_2|U) - \delta(\epsilon))} 2^{-n(I(U; Y_2) - \delta(\epsilon))}, \end{aligned}$$

where (a) follows since, by the independence of codebooks, the events

$$(U^n(1), X_1^n(m_1(j)|1), Y_2^n(j)) \in \mathcal{T}_\epsilon^{(n)}$$

and

$$(U^n(m_1(j)), Y_2^n(j+1)) \in \mathcal{T}_\epsilon^{(n)}$$

are independent for each $m_1(j) \neq 1$, and (b) follows from the independence of the codebooks and the joint typicality lemma. Therefore, the probability of error $\mathbb{P}(\mathcal{E}_{21}^c(j-1) \cap \mathcal{E}_{24}(j))$ tends to zero as $n \rightarrow \infty$ if $R_1 < I(X_1; Y_2|U) + I(U; Y_2) - 2\delta(\epsilon) = I(X_1; Y_2) - 2\delta(\epsilon)$. In the last block b , $\mathbb{P}\{\hat{M}_2(b) \neq 1\}$ tends to zero as $n \rightarrow \infty$ if $R_2 < I(X_2; Y_2|X_1) - \delta(\epsilon)$.

Finally, note that $I(X_2; Y_1) \leq I(X_2; Y_2|X_1) \leq I(X_2; Y_1|X_1)$, which guarantees the existence of $p(u|x_1)$ such that $I(X_2; Y_1|U) = I(X_2; Y_2|X_1)$. Combined with (4.2), this implies that $I(X_1; Y_2) = I(U; Y_1) + I(V; Y_1|X_2, U)$ and the corner point is achievable. For the symmetric Gaussian interference channels with $S < I < S(S+1)$, the corner point is achieved when $U \sim \mathcal{N}(0, \alpha P)$, $W \sim \mathcal{N}(0, (1-\alpha)P)$ and $X_1 = U + W$, where U and W are independent and $\alpha = (S^2 + S - I)/S^2$. This completes the proof of Theorem 4.1.1 and establishes the achievability of the corner point $(I(X_1; Y_2), I(X_2; Y_1|X_1))$.

4.1.2 General Rate Point: Single Dimensional SWSC

Now we relax the assumption that the two multiple access rate regions have the same sum-rate and show how to achieve an arbitrary point (I_1, I_2) on the dominant face of the simultaneous decoding inner bound. We split X_1 into three parts and keep X_2 unsplit. The following theory states the achievable rate region.

Theorem 4.1.2. *A rate pair (R_1, R_2) is achievable with the SWSC scheme if*

$$\begin{aligned} R_1 &< \min\{I(U_1; Y_1) + I(X_1; Y_1|U_1, X_2), I(U_1, U_2; Y_2) + I(X_1; Y_2|U_1, U_2, X_2)\}, \\ R_2 &< \min\{I(X_2; Y_1|U_1), I(X_2; Y_2|U_1, U_2)\} \end{aligned} \quad (4.3)$$

or

$$\begin{aligned}
 R_1 &< \min\{I(U_1; Y_2) + I(X_1; Y_2|U_1, X_2), I(U_1, U_2; Y_1) + I(X_1; Y_1|U_1, U_2, X_2)\}, \\
 R_2 &< \min\{I(X_2; Y_2|U_1), I(X_2; Y_1|U_1, U_2)\}
 \end{aligned} \tag{4.4}$$

for some pmf $p(u_1, u_2, x_1)p(x_2)$. In particular, for any rate point (I_1, I_2) in the simultaneous decoding inner bound, there exists a pmf $p(u_1, u_2, x_1)p(x_2)$ such that (I_1, I_2) is achievable.

In the following, we show how to achieve the above rate pair via a SWSC scheme, in which each message $M_1(j), j \in [1 : b - 2]$ is spread over three blocks and carried by $u_1^n(M_1(j))$ in block $j + 2$, $u_2^n(M_1(j)|M_1(j - 1))$ in block $j + 1$, and $x_1^n(M_1(j)|M_1(j - 1), M_1(j - 2))$ in block j . Two decoding rules will be considered, resulting in two achievable rate regions. The receivers can choose which rule to follow before the communication commences. Note that a sequence of $(b - 2)$ messages $M_1(j), j \in [1 : b - 2]$ and b messages $M_2(j), j \in [1 : b]$ are transmitted over b blocks. The average achievable rate pair over b blocks is $(\frac{b-2}{b}R_1, R_2)$, which can be made as close to (R_1, R_2) as desired. Details are as follows.

Codebook Generation. We use superposition coding. Fix the pmf $p(u_1, u_2, x_1)p(x_2)$ that attains the target rate pair. For notational convention, we assume $m_1(-1) = m_1(0) = m_1(b - 1) = m_1(b) = 1$. For $j \in [1 : b]$, randomly and independently generate 2^{nR_1} sequences $u_1^n(m_1(j - 2)), m_1(j - 2) \in [1 : 2^{nR_1}]$, each according to a product of $p(u_1)$. For each $m_1(j - 2)$, randomly and conditionally independently generate 2^{nR_1} sequences $u_2^n(m_1(j - 1)|m_1(j - 2)), m_1(j - 1) \in [1 : 2^{nR_1}]$, each according to a product of $p(u_2|u_1)$. For each pair $(m_1(j - 1), m_1(j - 2))$, randomly and conditionally independently generate 2^{nR_1} sequences $x_1^n(m_1(j)|m_1(j - 2), m_1(j - 1)), m_1(j) \in [1 : 2^{nR_1}]$, each according to a product of $p(x_1|u_1, u_2)$. Randomly and independently generate 2^{nR_2} sequences $x_2^n(m_2(j)), m_2(j) \in [1 : 2^{nR_2}]$, each according to a product of $p(x_2)$. This defines the

codebook

$$\mathcal{C}_j = \{u_1^n(m_1(j-2)), u_2^n(m_1(j-1)|m_1(j-2)), x_1^n(m_1(j)|m_1(j-2), m_1(j-1)), \\ x_2^n(m_2(j)): m_1(j-2), m_1(j-1), m_1(j) \in [1:2^{nR_1}], m_2(j) \in [1:2^{nR_2}]\}, j \in [1:b].$$

Encoding. Sender 1 transmits $x_1^n(m_1(j)|m_1(j-2), m_1(j-1))$ and sender 2 transmits $x_2^n(m_2(j))$ in block $j \in [1:b]$.

block j	1	2	3	4	...	$b-1$	b
U_1	1	1	$m_1(1)$	$m_1(2)$	$m_1(b-2)$
U_2	1	$m_1(1)$	$m_1(2)$	$m_1(b-2)$	1
X_1	$m_1(1)$	$m_1(2)$	$m_1(b-2)$	1	1
X_2	$m_2(1)$	$m_2(2)$	$m_2(b)$
Receiver 1	\emptyset	\emptyset	$\hat{m}_1(1)$	$\hat{m}_1(2)$	$\hat{m}_1(b-2)$
	$\hat{m}_2(1)$	$\hat{m}_2(2)$	$\hat{m}_2(3)$	$\hat{m}_2(4)$	$\hat{m}_2(b)$
Receiver 2	\emptyset	\emptyset	$\hat{m}_1(1)$	$\hat{m}_1(2)$	$\hat{m}_1(b-2)$
	\emptyset	$\hat{m}_2(1)$	$\hat{m}_2(2)$	$\hat{m}_2(3)$	$\hat{m}_2(b-1), \hat{m}_2(b)$

Table 4.2. SWSC scheme with decoding rule 1.

Decoding rule 1. Let the received sequences at block $j \in [1:b]$ be $y_1^n(j)$ and $y_2^n(j)$. We recover message $\hat{M}_1(j)$ via sliding-window decoding over three blocks. Table 4.2 reveals the scheduling of the messages and decoding orders.

For receiver 1, at the end of block $j = 1, 2$, it finds the unique $\hat{m}_2(j)$ such that

$$(x_2^n(\hat{m}_2(j)), u_1^n(m_1(j-2)), y_1^n(j)) \in \mathcal{T}_\epsilon^{(n)}.$$

At the end of block $j+2, j \in [1:b-2]$, receiver 1 first finds the unique $\hat{m}_1(j)$ such that

$$(x_1^n(\hat{m}_1(j)|\hat{m}_1(j-2), \hat{m}_1(j-1)), y_1^n(j), x_2^n(\hat{m}_2(j)), u_1^n(\hat{m}_1(j-2)), \\ u_2^n(\hat{m}_1(j-1)|\hat{m}_1(j-2))) \in \mathcal{T}_\epsilon^{(n)},$$

$$(u_2^n(\hat{m}_1(j)|\hat{m}_1(j-1)), y_1^n(j+1), x_2^n(\hat{m}_2(j+1)), u_1^n(\hat{m}_1(j-1))) \in \mathcal{T}_\epsilon^{(n)},$$

$$(u_1^n(\hat{m}_1(j)), y_1^n(j+2)) \in \mathcal{T}_\epsilon^{(n)}$$

simultaneously. Then it finds the unique $\hat{m}_2(j+2)$ such that

$$(x_2^n(\hat{m}_2(j+2)), u_1^n(\hat{m}_1(j)), y_1^n(j+2)) \in \mathcal{T}_\epsilon^{(n)}.$$

Following standard analysis, the rate constraints for successful decoding are

$$R_1 < I(U_1; Y_1) + I(X_1; Y_1 | U_1, X_2) - 3\delta(\epsilon)$$

$$R_2 < I(X_2; Y_1 | U_1) - \delta(\epsilon).$$
(4.5)

For receiver 2, at the end of block 2, it finds the unique message $\hat{m}_2(1)$ such that

$$(x_2^n(\hat{m}_2(1)), u_1^n(m_1(-1)), u_2^n(m_1(0)|m_1(-1)), y_2^n(1)) \in \mathcal{T}_\epsilon^{(n)}.$$

At the end of block $j+2$, $j \in [1 : b-2]$, it first finds the unique message $\hat{m}_1(j)$ such that

$$(x_1^n(\hat{m}_1(j)|\hat{m}_1(j-2), \hat{m}_1(j-1)), y_2^n(j), x_2^n(\hat{m}_2(j)), u_1^n(\hat{m}_1(j-2))),$$

$$u_2^n(\hat{m}_1(j-1)|\hat{m}_1(j-2))) \in \mathcal{T}_\epsilon^{(n)},$$

$$(u_2^n(\hat{m}_1(j)|\hat{m}_1(j-1)), y_2^n(j+1), u_1^n(\hat{m}_1(j-1))) \in \mathcal{T}_\epsilon^{(n)},$$

$$(u_1^n(\hat{m}_1(j)), y_2^n(j+2)) \in \mathcal{T}_\epsilon^{(n)}$$

simultaneously. It then finds the unique message $\hat{m}_2(j+1)$ such that

$$(x_2^n(\hat{m}_2(j+1)), u_1^n(\hat{m}_1(j-1)), u_2^n(\hat{m}_1(j)|\hat{m}_1(j-1)), y_2^n(j+1)) \in \mathcal{T}_\epsilon^{(n)}.$$

In the end, receiver 2 finds the unique message $\hat{m}_2(b)$, such that

$$(x_2^n(\hat{m}_2(b)), u_1^n(\hat{m}_1(b-2)), u_2^n(m_1(b-1)|\hat{m}_1(b-2)), y_2^n(b)) \in \mathcal{T}_\epsilon^{(n)}.$$

Following standard analysis, the rate constraints for successful decoding are

$$\begin{aligned} R_1 &< I(U_1, U_2; Y_2) + I(X_1; Y_2 | U_1, U_2, X_2) - 3\delta(\epsilon), \\ R_2 &< I(X_2; Y_2 | U_1, U_2) - \delta(\epsilon). \end{aligned} \quad (4.6)$$

Combining (4.5) and (4.6) results in the rate region in (4.3).

Decoding rule 2. We swap the role of receiver 1 and 2 in decoding rule 1. In particular, we swap receiver 1 with receiver 2, $y_1^n(j)$ with $y_2^n(j)$, and Y_1 with Y_2 in the description of the decoding rule 1. This results in the rate region in (4.4).

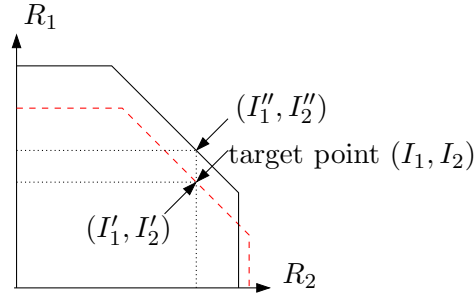


Figure 4.2. Simultaneous decoding inner bound formed by two multiple access rate regions with unequal sum-rates.

Finally, we show that the union of rate regions (4.3) and (4.4) covers arbitrary rate point (I_1, I_2) in the simultaneous decoding inner bound. For a fixed $(X_1, X_2) \sim p(x_1)p(x_2)$, suppose the corresponding multiple access rate regions and the target rate point (I_1, I_2) are as depicted in Figure 4.2. Since (I_1, I_2) lies in the intersection of the two multiple access channel rate regions, we can find two points (I'_1, I'_2) and (I''_1, I''_2) , one from the dominant face of each rate region, such that

$$\begin{aligned} I_1 &\leq \min\{I'_1, I''_1\}, \\ I_2 &\leq \min\{I'_2, I''_2\}. \end{aligned}$$

Figure 4.2 gives an example of such a choice. In certain cases, there can be multiple valid choices. For the example in Figure 4.2, in the red-dashed rate region, any rate point on the northeast of the target point can also be a valid choice for (I''_1, I''_2) . Then we keep

the marginal pmf on X_1 as $p(x_1)$ and choose a joint pmf $(U_1, U_2, X_1) \sim p(u_1, u_2, x_1)$ such that at least one of the following is true

$$\begin{aligned} I_2' &= I(X_2; Y_1 | U_1), \\ I_2'' &= I(X_2; Y_2 | U_1, U_2) \end{aligned} \tag{4.7}$$

or

$$\begin{aligned} I_2' &= I(X_2; Y_2 | U_1), \\ I_2'' &= I(X_2; Y_1 | U_1, U_2). \end{aligned} \tag{4.8}$$

Suppose that (4.7) is true. Since (I_1', I_2') and (I_1'', I_2'') are chosen from the dominant face of the corresponding multiple access rate regions, we have

$$\begin{aligned} I_1' &= I(U_1; Y_1) + I(X_1; Y_1 | U_1, X_2), \\ I_1'' &= I(U_1, U_2; Y_2) + I(X_1; Y_2 | U_1, U_2, X_2). \end{aligned}$$

In this case, the above SWSC scheme with decoding rule 1 achieves the target rate point (I_1, I_2) . Suppose otherwise that (4.8) is true. Then the target rate point will be achieved by following decoding rule 2. This completes the proof of Theorem 4.1.2.

4.2 Variations for Practical Purpose

4.2.1 Rate Loss

In the SWSC above, there is a $\frac{K-1}{b}R_1$ rate loss when X_1 is split into K layers. For example for $K = 2$, the loss comes since no message is scheduled for U^n in block 1 and for X_1^n in block b . We can instead send some messages at a lower rate. For example, one can always send a message at the treating-interference-as-noise rate $\min\{I(U; Y_1), I(U; Y_2)\}$ for U^n in block 1. For the coding scheme for Theorem 4.1.1, one can send a message at rate $\min\{I(X_1; Y_1 | U, X_2), I(X_1; Y_2 | U)\}$ for X_1^n in block b . In general, one can the

rate corresponding to the rate-splitting scheme for those unscheduled codewords at the beginning and end of transmission. This reduces the rate loss from $\frac{1}{b}R_1$ to $\frac{1}{b}\Delta R_1$, where ΔR_1 is the rate difference between the SWSC and the rate-splitting for the same decoding order. In the running example,

$$\begin{aligned} \Delta R_1 &= \min\{I(U; Y_1) + I(X_1; Y_1|U, X_2), I(X_1; Y_2)\} \\ &\quad - [\min\{I(U; Y_1), I(U; Y_2) + \min\{I(X_1; Y_1|U, X_2), I(X_1; Y_2|U)\}]. \end{aligned}$$

4.2.2 Homogeneous Superposition Coding

In the description of the sliding-window superposition coding scheme above, we applied the *heterogeneous* superposition coding [10], in which the signal X_1 is superimposed on top of U_1 according to the conditional pmf $p(x_1|u_1)$. For practical implementation purpose, it is sometimes easier to apply homogeneous superposition coding [22], in which the two signal layers U_1 and U_2 are independent according to $p(u_1)p(u_2)$ and are superimposed to the physical channel input X_1 through a symbol-by-symbol mapping $x_1(u_1, u_2)$. (For a comparison of the two superposition coding schemes under optimal decoding, see Chapter 5.)

For example, for Gaussian channels, a typical choice using homogeneous superposition coding is $U_1 \sim N(0, \alpha P)$, $U_2 \sim N(0, (1 - \alpha)P)$ and $X_1 = U_1 + U_2$. In practical systems, when U_1 and U_2 are BPSK signals and distributed according to $\text{Unif}\{-1, +1\}$, one can form a 4-PAM channel input by choosing $X_1 = \sqrt{\alpha P}U_1 + \sqrt{(1 - \alpha)P}U_2$. In particular, choosing $\alpha = 0.8$ makes X_1 a uniformly-spaced 4-PAM signal

$$X_1 \in \{-3\sqrt{P}/\sqrt{5}, -\sqrt{P}/\sqrt{5}, +\sqrt{P}/\sqrt{5}, +3\sqrt{P}/\sqrt{5}\}.$$

4.2.3 Decoding Orders

To show the achievability of an arbitrary rate point in the simultaneous decoding inner bound, it is sufficient to consider the union of rate regions (4.3) and (4.4) in

Theorem 4.1.2. However, there can be other valid decoding rules for the given codebook generation and encoding. Simply put, receiver $i = 1, 2$, can either choose from one of the decoding orders $\{d_{ik} : k \in [1 : 4]\}$, or treat interference as noise.

For decoding order $d_{i1} : \hat{m}_1(1) \rightarrow \hat{m}_2(1) \rightarrow \hat{m}_1(2) \rightarrow \hat{m}_2(2) \rightarrow \dots \rightarrow \hat{m}_1(b-2) \rightarrow \hat{m}_2(b-2) \rightarrow \hat{m}_2(b-1) \rightarrow \hat{m}_2(b)$, the achievable rate region $\mathcal{R}_{i1}(p)$ is the set of rate pairs (R_1, R_2) such that

$$\begin{aligned} R_1 &< I(X_1; Y_i), \\ R_2 &< I(X_2; Y_i | X_1). \end{aligned}$$

For decoding order $d_{i2} : \hat{m}_2(1) \rightarrow \hat{m}_1(1) \rightarrow \hat{m}_2(2) \rightarrow \hat{m}_1(2) \rightarrow \dots \rightarrow \hat{m}_2(b-2) \rightarrow \hat{m}_1(b-2) \rightarrow \hat{m}_2(b-1) \rightarrow \hat{m}_2(b)$, the achievable rate region $\mathcal{R}_{i2}(p)$ is the set of rate pairs (R_1, R_2) such that

$$\begin{aligned} R_1 &< I(U_1, U_2; Y_i) + I(X_1; Y_i | U_1, U_2, X_2), \\ R_2 &< I(X_2; Y_i | U_1, U_2). \end{aligned}$$

For decoding order $d_{i3} : \hat{m}_2(1) \rightarrow \hat{m}_2(2) \rightarrow \hat{m}_1(1) \rightarrow \hat{m}_2(3) \rightarrow \hat{m}_1(2) \rightarrow \hat{m}_2(4) \rightarrow \dots \rightarrow \hat{m}_1(b-2) \rightarrow \hat{m}_2(b)$, the achievable rate region $\mathcal{R}_{i3}(p)$ is the set of rate pairs (R_1, R_2) such that

$$\begin{aligned} R_1 &< I(U_1; Y_i) + I(X_1; Y_i | U_1, X_2), \\ R_2 &< I(X_2; Y_i | U_1). \end{aligned}$$

For decoding order $d_{i4} : \hat{m}_2(1) \rightarrow \hat{m}_2(2) \rightarrow \hat{m}_2(3) \rightarrow \hat{m}_1(1) \rightarrow \hat{m}_2(4) \rightarrow \hat{m}_1(2) \rightarrow \dots \rightarrow \hat{m}_2(b) \rightarrow \hat{m}_1(b-2)$, the achievable rate region $\mathcal{R}_{i4}(p)$ is the set of rate pairs (R_1, R_2) such that

$$R_1 < I(X_1; Y_i | X_2),$$

$$R_2 < I(X_2; Y_i).$$

We denote the achievable rate region for treating interference as noise at receiver i as $\mathcal{R}_{i0}(p)$, which is the set of rate pairs (R_1, R_2) such that

$$R_i < I(X_i; Y_i).$$

The combined achievable rate region given current encoding is

$$\bigcup_{k=0}^4 \bigcup_{l=0}^4 (\mathcal{R}_{1k}(p) \cap \mathcal{R}_{2l}(p)).$$

As subsets, decoding rule 1 in the proof of Theorem 4.1.2 corresponds to $\mathcal{R}_{13}(p) \cap \mathcal{R}_{22}(p)$, and decoding rule 2 corresponds to $\mathcal{R}_{12}(p) \cap \mathcal{R}_{23}(p)$.

4.2.4 Superposition Layers

For some applications, one might want to have a more symmetric encoding scheme that splits each user's channel input into two layers, i.e., (U, X_1) for user 1 and (V, X_2) for user 2. Encoding is illustrated in Table 4.3.

block j	1	2	3	...	$b-1$	b
U	1	$m_1(1)$	$m_1(2)$	$m_1(b-1)$
X_1	$m_1(1)$	$m_1(2)$	$m_1(b-1)$	1
V	1	$m_2(1)$	$m_2(2)$	$m_2(b-1)$
X_2	$m_2(1)$	$m_2(2)$	$m_2(b-1)$	1

Table 4.3. Sliding-window superposition encoding for (U, X_1) and (V, X_2) .

As for decoding, one can similarly work out all valid decoding orders for the given encoding. Here we introduce a more compact way of writing the corresponding achievable rate regions. We define a *layering order* $s: S_1 \rightarrow S_2 \rightarrow S_3 \rightarrow S_4$ as the an ordering of the variables $\{U, X_1, V, X_2\}$ such that the relative orders $U \rightarrow X_1$ and

$V \rightarrow X_2$ are preserved. There are six different layering orders for the current split:

$$\begin{aligned}
s_1: U &\rightarrow X_1 \rightarrow V \rightarrow X_2 \\
s_2: U &\rightarrow V \rightarrow X_1 \rightarrow X_2 \\
s_3: V &\rightarrow U \rightarrow X_2 \rightarrow X_1 \\
s_4: V &\rightarrow X_2 \rightarrow U \rightarrow X_1 \\
s_5: U &\rightarrow V \rightarrow X_2 \rightarrow X_1 \\
s_6: V &\rightarrow U \rightarrow X_1 \rightarrow X_2.
\end{aligned} \tag{4.9}$$

For each layering order, we define the index sets $\mathcal{I}_1 := \{i: S_i = U \text{ or } X_1\}$ and $\mathcal{I}_2 := \{i: S_i = V \text{ or } X_2\}$. For example, for layering order s_2 , $\mathcal{I}_1 = \{1, 3\}$ and $\mathcal{I}_2 = \{2, 4\}$. Now, for each layering order s_i , we associate a rate region $\mathcal{R}_{ji}(p)$ defined as the set of rate pairs (R_1, R_2) such that

$$\begin{aligned}
R_1 &< \sum_{i \in \mathcal{I}_1} I(S_i; Y_j | S^{i-1}), \\
R_2 &< \sum_{i \in \mathcal{I}_2} I(S_i; Y_j | S^{i-1}).
\end{aligned} \tag{4.10}$$

We say a layering order is *achievable* if the associated rate region is achievable. We further denote the treating interference as noise region $\mathcal{R}_{j0}(p)$, $j = 1, 2$, as the set of rate pairs (R_1, R_2) such that

$$R_j < I(X_j; Y_j). \tag{4.11}$$

It can be checked that the achievable rate region for the given encoding is

$$\bigcup_{j=0}^4 \bigcup_{k=0}^4 [\mathcal{R}_{1j}(p) \cap \mathcal{R}_{2k}(p)]. \tag{4.12}$$

Remark 4.2.1. *One might notice that the rate regions corresponding to layering orders s_5 and s_6 are included in the above expression (4.12). Indeed, this is a limitation of the single dimensional SWSC. We will introduce a two dimensional SWSC that achieves all*

possible layering orders in Section 4.4.

More generally, one can split X_1 into K layers (U_1, \dots, U_K) and X_2 into L layers (V_1, \dots, V_L) . Message $M_1(i)$, $i \in [1 : b - K + 1]$, is sent through signal U_k in block $i + K - k$ for $k \in [1 : K]$. Message $M_2(i)$, $i \in [1 : b - L + 1]$, is sent through signal V_l in block $i + L - l$ for $l \in [1 : L]$. Let us call such splitting and message scheduling as a *single dimensional K-L split*. Similar to the case of the single dimensional 2-2 split above, not every layering order is achievable. The layering orders achievable are of the following form. It starts with a sequence of consecutive U 's (or V 's respectively). Then, it alternates between one V (U) and one U (V) until one of them is exhausted. It ends with the rest unexhausted variables. For example, assuming $K > L$, the $K+L$ achievable *alternating* decoding orders are

$$\begin{aligned}
s_1: & U_1 \rightarrow \dots \rightarrow U_K \rightarrow V_1 \rightarrow \dots \rightarrow V_L \\
s_2: & U_1 \rightarrow \dots \rightarrow U_{K-1} \rightarrow V_1 \rightarrow U_K \rightarrow V_2 \rightarrow \dots \rightarrow V_L \\
s_3: & U_1 \rightarrow \dots \rightarrow U_{K-2} \rightarrow V_1 \rightarrow U_{K-1} \rightarrow V_2 \rightarrow U_K \rightarrow V_3 \rightarrow \dots \rightarrow V_L \\
& \vdots \\
s_K: & U_1 \rightarrow V_1 \rightarrow U_2 \rightarrow V_2 \rightarrow \dots \rightarrow U_L \rightarrow V_L \rightarrow U_{L+1} \rightarrow \dots \rightarrow U_K \\
s_{K+1}: & V_1 \rightarrow U_1 \rightarrow V_2 \rightarrow U_2 \rightarrow \dots \rightarrow V_L \rightarrow U_L \rightarrow \dots \rightarrow U_K \\
s_{K+2}: & V_1 \rightarrow V_2 \rightarrow U_1 \rightarrow V_3 \rightarrow U_2 \rightarrow \dots \rightarrow V_L \rightarrow U_{L-1} \rightarrow \dots \rightarrow U_K \\
& \vdots \\
s_{K+L}: & V_1 \rightarrow \dots \rightarrow V_L \rightarrow U_1 \rightarrow \dots \rightarrow U_K
\end{aligned}$$

The index sets \mathcal{I}_1 and \mathcal{I}_2 are similarly defined as follows

$$\begin{aligned}
\mathcal{I}_1 &= \{i: S_i = U_k \text{ for some } k \in [1 : K]\}, \\
\mathcal{I}_2 &= \{i: S_i = V_l \text{ for some } l \in [1 : L]\}.
\end{aligned}$$

The rate region $\mathcal{R}_{jk}(p)$ corresponding to decoding order $s_k: S_1 \rightarrow \dots \rightarrow S^{K+L}$ is the set of rate pairs (R_1, R_2) such that

$$R_1 < \sum_{i \in \mathcal{I}_1} I(S_i; Y_j^n | S^{i-1}),$$

$$R_2 < \sum_{i \in \mathcal{I}_2} I(S_i; Y_j^n | S^{i-1}).$$

Theorem 4.2.1. *The achievable rate region for the above single dimensional K - L split encoding for a two-user interference channel $p(y_1, y_2 | x_1, x_2)$ is given by*

$$\bigcup_{k=0}^{K+L} \bigcup_{l=0}^{K+L} [\mathcal{R}_{1k}(p) \cap \mathcal{R}_{2l}(p)].$$

4.3 Sliding-Window Coded Modulation

Now we combine the SWSC scheme with coded modulation and simulate its performance using 4G LTE turbo codes in the two-user Gaussian interference channel.

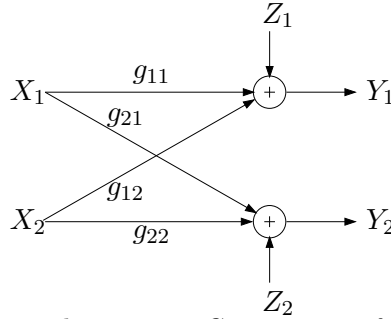


Figure 4.3. The two-user Gaussian interference channel.

Consider the two-user Gaussian interference channel (Figure 4.3)

$$Y_1 = g_{11}X_1 + g_{12}X_2 + Z_1,$$

$$Y_2 = g_{21}X_1 + g_{22}X_2 + Z_2,$$
(4.13)

where $X_i \in \mathcal{X}$, $i = 1, 2$, is the transmitted signal from sender i with average power constraint P_i , $Y_i \in \mathbb{R}$ is the received signal at receiver i , and $Z_i \in \mathbb{R} \sim \mathcal{N}(0, 1)$ is the

Gaussian noise, independent of X_i . We assume that each receiver i knows local channel gain coefficients $g_{ij} \in \mathbb{R}$, $j = 1, 2$, from both senders, which are constant during the communication.

The original SWSC scheme allows for full flexibility in the number of superimposed layers, the number and structure of auxiliary random variables for superposition coding, and the decoding order. Here, we limit our attention to two layers of BPSK signals that form a 4-PAM signal by superposition and a fixed decoding order. In particular,

$$\begin{aligned} X_1 &= \sqrt{P_1}\sqrt{\alpha}U + \sqrt{P_1}\sqrt{1-\alpha}V, \\ X_2 &= \sqrt{P_2}W, \end{aligned} \tag{4.14}$$

where U , V , and $W \in \{-1, +1\}$ are independent BPSK signals. The parameter α determines the ratio of powers split into U and V . We choose $\alpha = 0.8$, which makes $X_1 \in \{-3\sqrt{P_1}/\sqrt{5}, -\sqrt{P_1}/\sqrt{5}, +\sqrt{P_1}/\sqrt{5}, +3\sqrt{P_1}/\sqrt{5}\}$ a *uniformly-spaced* 4-PAM signal. The corresponding channel outputs are

$$\begin{aligned} Y_1 &= g_{11}\sqrt{P_1}\sqrt{\alpha}U + g_{11}\sqrt{P_1}\sqrt{1-\alpha}V + g_{12}\sqrt{P_2}W + Z_1, \\ Y_2 &= g_{21}\sqrt{P_1}\sqrt{\alpha}U + g_{21}\sqrt{P_1}\sqrt{1-\alpha}V + g_{22}\sqrt{P_2}W + Z_2. \end{aligned}$$

4.3.1 Theoretical Performance Comparison

Given the current encoding and modulation scheme, we compare the theoretical performance of treating interference as noise, simultaneous nonunique decoding, and sliding-window coded modulation.

- *Treating interference as noise.*

The achievable rate region is the set of rate pairs (R_1, R_2) such that

$$R_1 < I(X_1; Y_1),$$

$$R_2 < I(X_2; Y_2),$$

where $X_1 \sim \text{Unif}\{-3\sqrt{P_1}/\sqrt{5}, -\sqrt{P_1}/\sqrt{5}, +\sqrt{P_1}/\sqrt{5}, +3\sqrt{P_1}/\sqrt{5}\}$ and $X_2 \sim \text{Unif}\{-\sqrt{P_2}, +\sqrt{P_2}\}$.

- *Simultaneous nonunique decoding.*

The achievable rate region is the set of rate pairs (R_1, R_2) such that

$$R_1 < I(X_1; Y_1 | X_2),$$

$$R_2 < I(X_2; Y_2 | X_1),$$

$$R_1 + R_2 < \min\{I(X_1, X_2; Y_1), I(X_1, X_2; Y_2)\},$$

where $X_1 \sim \text{Unif}\{-3\sqrt{P_1}/\sqrt{5}, -\sqrt{P_1}/\sqrt{5}, +\sqrt{P_1}/\sqrt{5}, +3\sqrt{P_1}/\sqrt{5}\}$ and $X_2 \sim \text{Unif}\{-\sqrt{P_2}, +\sqrt{P_2}\}$.

- *Sliding-window coded modulation.*

The achievable rate region by sliding-window coded modulation is the set of rate pairs (R_1, R_2) such that

$$R_1 < \min\{I(U; Y_1) + I(V; Y_1 | W) I(U, V; Y_2)\},$$

$$R_2 < \min\{I(W; Y_1 | U) I(W; Y_2 | U, V)\}.$$

where U , V , and W are independent $\text{Unif}\{-1, +1\}$ random variables,

$X_1 \sim \text{Unif}\{-3\sqrt{P_1}/\sqrt{5}, -\sqrt{P_1}/\sqrt{5}, +\sqrt{P_1}/\sqrt{5}, +3\sqrt{P_1}/\sqrt{5}\}$, and

$X_2 \sim \text{Unif}\{-\sqrt{P_2}, +\sqrt{P_2}\}$.

4.3.2 Implementation With LTE Turbo Codes

For $j = 1, \dots, b-1$, we use a binary linear code of length $2n$ and rate $r_1/2$ to encode message $m_1(j)$ through $v^n(m_1(j))$ in the j -th block and $u^n(m_1(j))$ in the $(j+1)$ -st block. For $j = 1, \dots, b$, we use a binary linear code of length n and rate r_2

to encode message $m_2(j)$ through $w^n(m_2(j))$ in the j -th block. We adopt the turbo codes used in the LTE standard [1], which allow flexibility in code rate and block length. In particular, we start with the rate $1/3$ mother code and adjust the rates and lengths according to the rate matching algorithm in the standard. Note that for $r_1 < 2/3$, some code bits are repeated and for $r_1 > 2/3$, some code bits are punctured. We set the block length $n = 2048$ and the number of blocks $b = 20^1$ respectively. We use the LOG-MAP algorithm for the turbo decoding with the maximum number of iterations set to 8 for each stage of decoding. We assume that a rate pair (R_1, R_2) is achieved for given P_i and g_{ij} if the resulting bit-error rate (BER) is below 10^{-3} over 1000 independent sets of simulations.

4.3.3 Simulation Results

We assume symmetric rate, power, and channel gains, i.e., $R_1 = R_2 = R, P_1 = P_2 = P, g_{11} = g_{22} = 1$, and $g_{12} = g_{21} = g$. We fix the signal-to-noise ratio SNR to be 10dB and vary the interference-to-noise ration INR from 8-12dB.

As shown in Figure 4.4, the sliding-window coded modulation scheme outperforms treating interference as noise in all strong and weak interference regimes and approaches the theoretically best known performance of simultaneous decoding as the interference becomes strong.

4.4 Han–Kobayashi Inner Bound

In this section, we describe an alternative SWSC scheme for the two-user interference channel that achieves an arbitrary rate point on the simultaneous decoding inner bound by a *two-dimensional 2-2* split (U, X_1) and (V, X_2) . We first show the achievability of an arbitrary rate pair in the two-user simultaneous decoding region in Theorem 4.4.1

¹It should be stressed that b is the total number of blocks, not the size of the decoding window (which is 2). Every message is recovered with one-block delay. While a larger b reduces the rate penalty of $1/b$, it also incurs error propagation over multiple blocks, both of which were properly taken into account in our rate and BER calculation.

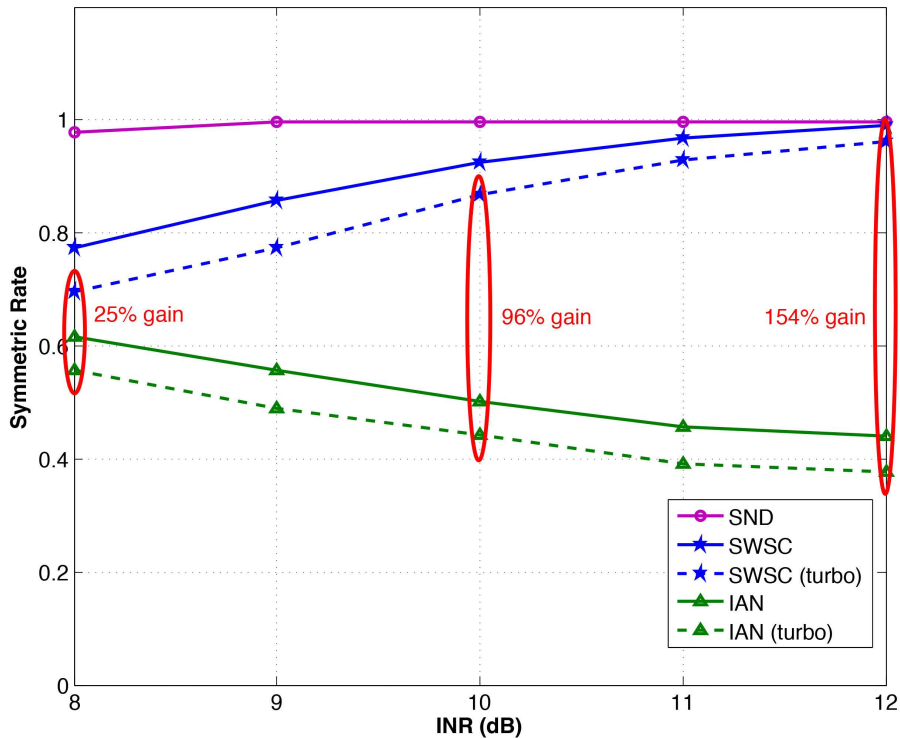


Figure 4.4. Performance comparison in the symmetric Gaussian interference channel.

in Section 4.4.1. Then, we show that for the given two-dimensional 2-2 split, all possible layering orders s_1, s_2, \dots, s_6 in (4.9) are achievable by varying over different decoding orders in Theorem 4.4.2. As an important application of this theorem, we establish in Section 4.4.2 the achievability of the Han–Kobayashi inner bound for the two-user interference channel [34] using a SWSC scheme in Theorem 4.4.3.

4.4.1 Two-dimensional SWSC

Theorem 4.4.1. *A rate pair (R_1, R_2) is achievable with the two-dimensional SWSC scheme if*

$$R_1 < \min\{I(U; Y_1) + I(X_1; Y_1|U, X_2), I(X_1; Y_2|V)\},$$

$$R_2 < \min\{I(X_2; Y_1|U), I(V; Y_2) + I(X_2; Y_2|V, X_1)\}$$

for some pmf $p(u, x_1)p(v, x_2)$. In particular, for any rate point (I_1, I_2) in the simultaneous decoding inner bound, there exists a pmf $p(u, x_1)p(v, x_2)$ such that (I_1, I_2) is achievable.

Remark 4.4.1. *The achievable rate region above is equivalent as*

$$\bigcup_{p=p(u, x_1)p(v, x_2)} (\mathcal{R}_{15}(p) \cap \mathcal{R}_{26}(p)),$$

where $\mathcal{R}_{15}(p)$ and $\mathcal{R}_{26}(p)$ are the same as in (4.10). In other words, receiver 1 achieves the rate region corresponding to layering order $s_5: U \rightarrow V \rightarrow X_2 \rightarrow X_1$ and receiver 2 achieves the rate region corresponding to layering order $s_6: V \rightarrow U \rightarrow X_1 \rightarrow X_2$.

Remark 4.4.2. *One advantage of the two-dimensional SWSC scheme is that the split of (U, X_1) and (V, X_2) can be chosen separately. In other words, one only needs to separately make sure that $p(u, x_1)$ is chosen such that*

$$\begin{aligned} I_1 &\leq I(U; Y_1) + I(X_1; Y_1 | U, X_2), \\ I_2 &\leq I(X_2; Y_1 | U) \end{aligned}$$

and that $p(v, x_2)$ is chosen such that

$$\begin{aligned} I_1 &\leq I(X_1; Y_2 | V), \\ I_2 &\leq I(V; Y_2) + I(X_2; Y_2 | V, X_1). \end{aligned}$$

However, as we will see, the decoding delay and finite-length rate loss incurred by two dimensional SWSC scheme is larger.

Now we briefly sketch the coding scheme. We transmit $b_1(b_2 - 1)$ messages $M_1(jk), j \in [1 : b_1], k \in [1 : b_2 - 1]$, and $(b_1 - 1)b_2$ messages $M_2(jk), j \in [1 : b_1 - 1], k \in [1 : b_2]$, over $b_1 b_2$ blocks. The average achievable rate pair over $b_1 b_2$ blocks is $(\frac{b_2-1}{b_2} R_1, \frac{b_1-1}{b_1} R_2)$, which can be made arbitrarily close to (R_1, R_2) . In Figures 4.5 and

4.6, we illustrate the encoding and decoding for $b_1 = b_2 = 4$.

U, V X_1, X_2	$i_2 = 1$	2	3	$i_2 = 4$
$i_1 = 1$	1, 1 $m_1(11), m_2(11)$	1, $m_2(11)$ $m_1(12), m_2(12)$	1, $m_2(12)$ $m_1(13), m_2(13)$	1, $m_2(13)$ $m_1(14), 1$
2	$m_1(11), 1$ $m_1(21), m_2(21)$	$m_1(12), m_2(21)$ $m_1(22), m_2(22)$	$m_1(13), m_2(22)$ $m_1(23), m_2(23)$	$m_1(1, 4), m_2(23)$ $m_2(1, b_2), 1$
3	$m_1(21), 1$ $m_1(31), m_2(31)$	$m_1(22), m_2(31)$ $m_1(32), m_2(32)$	$m_1(23), m_2(32)$ $m_1(33), m_2(33)$	$m_1(24), m_2(33)$ $m_1(34), m_2(34)$
$i_1 = 4$	$m_1(31), 1$ 1, $m_2(41)$	$m_1(32), m_2(41)$ 1, $m_2(42)$	$m_1(33), m_2(42)$ 1, $m_2(43)$	$m_1(34), m_2(43)$ 1, 1

Figure 4.5. Message scheduling for the two-dimensional SWSC.

Encoding. We use superposition coding to generate independent codebooks, one for each block. On each column $i_2 \in [1 : b_2]$, message $M_1(i_1, i_2), i_1 \in [1 : b_1 - 1]$, is spread over two blocks and carried by $u^n(M_1(i_1, i_2))$ in block $(i_1 + 1, i_2)$ and $x_1^n(M_1(i_1, i_2)|M_1(i_1 - 1, i_2))$ in block (i_1, i_2) . On each row $i_1 \in [1 : b_1]$, message $M_2(i_1, i_2), i_2 \in [1 : b_2 - 1]$, is spread over two blocks and carried by $v^n(M_2(i_1, i_2))$ in block $(i_1, i_2 + 1)$ and $x_2^n(M_2(i_1, i_2)|M_2(i_1, i_2 - 1))$ in block (i_1, i_2) .

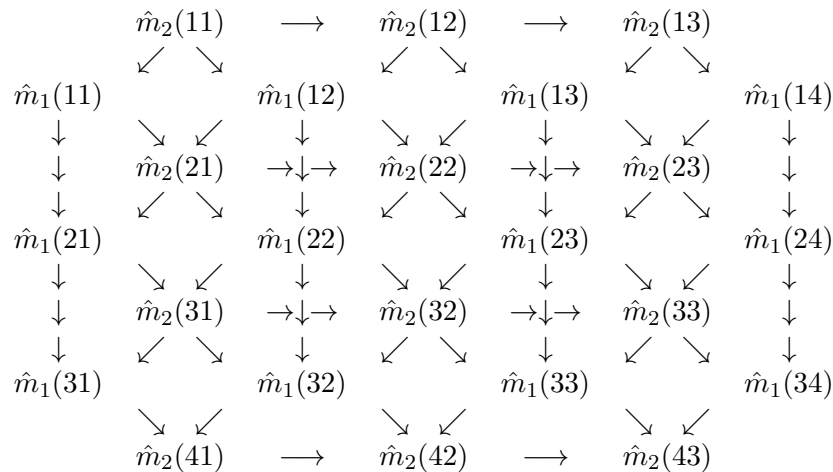


Figure 4.6. The dependency in recovering all messages at receiver 1.

Decoding. In Figure 4.6, we describe the dependency in recovering messages $\{\hat{m}_1(i_1, i_2)\}$ and $\{\hat{m}_2(i_1, i_2)\}$ at receiver 1. Here the arrow $\hat{m}_A \rightarrow \hat{m}_B$ indicates that message m_A should be recovered before message m_B . A valid decoding order is any ordering of $\{\hat{m}_1(i_1, i_2)\} \cup \{\hat{m}_2(i_1, i_2)\}$ such that no message is recovered before any message it depends on. It is clear that a valid decoding order exists if the dependency graph in Figure 4.6 is acyclic. For example, one valid decoding order is top-down row by row and from left to right on each row:

$$\begin{aligned}
& \hat{m}_2(11) \rightarrow \cdots \rightarrow \hat{m}_2(13) \rightarrow \hat{m}_1(11) \rightarrow \cdots \rightarrow \hat{m}_1(14) \\
& \rightarrow \hat{m}_2(21) \rightarrow \cdots \rightarrow \hat{m}_2(23) \rightarrow \hat{m}_1(21) \rightarrow \cdots \rightarrow \hat{m}_1(24) \\
& \rightarrow \hat{m}_2(31) \rightarrow \cdots \rightarrow \hat{m}_2(33) \rightarrow \hat{m}_1(31) \rightarrow \cdots \rightarrow \hat{m}_1(34) \\
& \rightarrow \hat{m}_2(41) \rightarrow \cdots \rightarrow \hat{m}_2(43).
\end{aligned}$$

Another valid decoding order is to follow the diagonals:

$$\begin{aligned}
& \hat{m}_2(11) \rightarrow \hat{m}_1(11) \\
& \rightarrow \hat{m}_2(12) \rightarrow \hat{m}_1(12) \rightarrow \hat{m}_2(21) \rightarrow \hat{m}_1(21) \\
& \rightarrow \hat{m}_2(13) \rightarrow \hat{m}_1(13) \rightarrow \hat{m}_2(22) \rightarrow \hat{m}_1(22) \rightarrow \hat{m}_2(31) \rightarrow \hat{m}_1(31) \\
& \rightarrow \hat{m}_1(14) \rightarrow \hat{m}_2(23) \rightarrow \hat{m}_1(23) \rightarrow \hat{m}_2(32) \rightarrow \hat{m}_1(32) \rightarrow \hat{m}_2(41) \\
& \rightarrow \hat{m}_1(24) \rightarrow \hat{m}_2(33) \rightarrow \hat{m}_1(33) \rightarrow \hat{m}_2(42) \\
& \rightarrow \hat{m}_1(34) \rightarrow \hat{m}_2(43).
\end{aligned}$$

Following a valid decoding order, the decoder applies successive cancellation and sliding-window decoding over two corresponding blocks. With a similar analysis as in the single dimensional SWSC scheme, it can be shown that the rate constraints for successful

decoding at receiver 1 is

$$\begin{aligned} R_1 &< I(U; Y_1) + I(X_1; Y_1 | U, X_2), \\ R_2 &< I(X_2; Y_1 | U). \end{aligned} \tag{4.15}$$

This is the rate region corresponding to layering order $U \rightarrow V \rightarrow X_2 \rightarrow X_1$. The dependency relation at receiver 2 can be derived from Figure 4.6 by swapping $\hat{m}_1(i_1, i_2)$ with $\hat{m}_2(i_2, i_1)$ for $i_1 \in [1 : 3], i_2 \in [1 : 4]$. One can similarly derived the rate constraints for successful decoding at receiver 2

$$\begin{aligned} R_1 &< I(X_1; Y_2 | V), \\ R_2 &< I(V; Y_2) + I(X_2; Y_2 | V, X_1). \end{aligned} \tag{4.16}$$

This is the rate region corresponding to layering order $V \rightarrow U \rightarrow X_1 \rightarrow X_2$. Combining (4.15) and 4.16 establishes Theorem 4.4.1.

Remark 4.4.3 (Decoding delay). *Suppose that the transmission is row by row, that is in the order of $(11) \rightarrow \dots \rightarrow (1, b_2) \rightarrow (21) \rightarrow \dots \rightarrow (2, b_2) \rightarrow \dots \rightarrow (b_1, 1) \rightarrow \dots \rightarrow (b_1, b_2)$. Some thoughts on transmission order and the decoding orders reveal that at receiver 1, the the decoding delay for both messages are b_2 blocks, while at the receiver 2, the decoding delay is one block for message M_1 , and b_2 blocks for message M_2 . Similarly, if the transmission is column by column, the maximum decoding delay is b_1 blocks. In both cases, it is much larger than the single dimensional SWSC scheme, which incurs a delay of two blocks.*

Remark 4.4.4 (Finite-length rate loss). *For a single dimensional 2-2 split, the rate loss is $\frac{1}{b}R_j$ for $j = 1, 2$, where b is the number of blocks. Suppose in the two-dimensional SWSC, we choose $b_1 = b_2$ and thus the block length is $b = b_1b_2$. Then, the rate loss incurred by a two-dimensional 2-2 split is*

$$\frac{1}{\sqrt{b}}R_j$$

for $j = 1, 2$, which is larger than $\frac{1}{6}R_j$ in the single dimensional SWSC.

Now let us fully explore all possible achievable layering orders of the above two-dimensional 2-2 split.

Theorem 4.4.2. *For a 2-2 split $p = p(u, x_1)p(v, x_2)$ and two-dimensional message scheduling as illustrated in Figure 4.5, the achievable rate region is*

$$\bigcup_{i_1=0}^6 \bigcup_{i_2=0}^6 [\mathcal{R}_{1,i_1}(p) \cap \mathcal{R}_{2,i_2}(p)],$$

or equivalently,

$$\bigcap_{j=1}^2 \left(\bigcup_{i=0}^6 \mathcal{R}_{ji}(p) \right),$$

where $\mathcal{R}_{ji}(p), j = 1, 2, i \in [0:6]$, are defined the same as in (4.10) and (4.11).

Remark 4.4.5. *Unlike the single-dimensional SWSC scheme, which only achieves layering orders s_1, s_2, s_3, s_4 , Theorem 4.4.2 states that the two-dimensional SWSC scheme achieves all possible layering orders.*

It is easy to check that the treating interference as noise region $\mathcal{R}_{j0}(p), j = 1, 2$, is achievable. We show the achievability of each rate region $\mathcal{R}_{ji}(p)$ for $j = 1, 2$ and $i \in [1:6]$. The crucial step here is to show the dependency graph among all messages is acyclic and thus a valid decoding order exists. We first note that for the same layering order s_i , the dependency graph remains the same at different receivers $j = 1, 2$. Second, we note that the layering order s_4 can be obtain from s_1 by swapping $U \leftrightarrow V$ and $X_1 \leftrightarrow X_2$. Similarly for $s_2 \leftrightarrow s_3$ and $s_5 \leftrightarrow s_6$. As a consequence, the dependency graph for layering order s_4 (and s_3, s_6 respectively) can be obtained from that of s_1 (and s_2, s_5 respectively) by swapping $\hat{m}_1(i_1, i_2) \leftrightarrow \hat{m}_2(i_2, i_1)$. Therefore, there are three essentially different cases: layering orders s_1, s_2 , and s_5 . The dependency graph corresponding to layering order s_5 is given in Figure 4.6. Now we show the dependency graph corresponding to layering order s_1 in Figures 4.7.

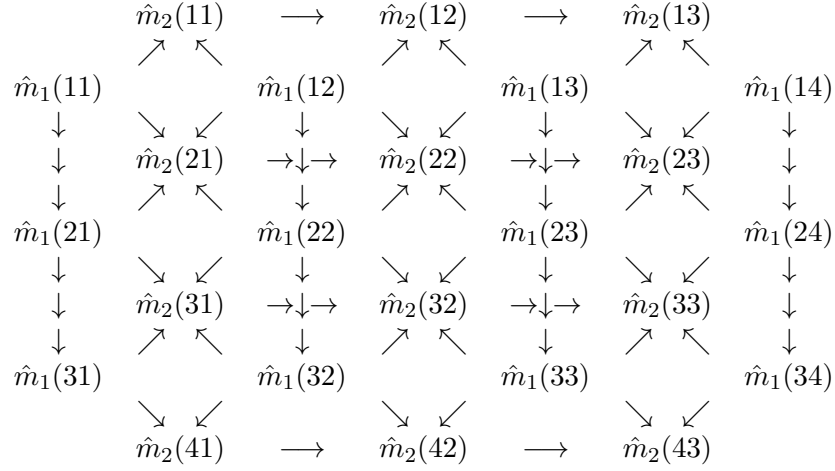


Figure 4.7. The dependency graph corresponds to layering order s_1 .

One valid decoding order corresponds to layering order s_1 is

$$\begin{aligned}
& \hat{m}_1(11) \rightarrow \hat{m}_1(12) \rightarrow \hat{m}_2(11) \rightarrow \hat{m}_1(13) \rightarrow \hat{m}_2(12) \rightarrow \hat{m}_1(14) \rightarrow \hat{m}_2(13) \\
& \rightarrow \hat{m}_1(21) \rightarrow \hat{m}_1(22) \rightarrow \hat{m}_2(21) \rightarrow \hat{m}_1(23) \rightarrow \hat{m}_2(22) \rightarrow \hat{m}_1(24) \rightarrow \hat{m}_2(23) \\
& \rightarrow \hat{m}_1(31) \rightarrow \hat{m}_1(32) \rightarrow \hat{m}_2(31) \rightarrow \hat{m}_1(33) \rightarrow \hat{m}_2(32) \rightarrow \hat{m}_1(34) \rightarrow \hat{m}_2(33) \\
& \rightarrow \hat{m}_2(41) \rightarrow \hat{m}_2(42) \rightarrow \hat{m}_2(43).
\end{aligned}$$

Following a similar analysis as in the single dimensional SWSC scheme, one can show that the rate region $\mathcal{R}_{j1}(p)$ is achievable. We notice that the decoding order on each row above is the same as that of a single dimensional SWSC scheme for a target rate region $\mathcal{R}_{j1}(p)$.

Now we show the dependency graph corresponding to layering orders s_2 in Figure 4.8.

One valid decoding order corresponds to layering order s_2 is

$$\begin{aligned}
& \hat{m}_1(11) \rightarrow \hat{m}_2(11) \rightarrow \hat{m}_1(12) \rightarrow \hat{m}_2(12) \rightarrow \hat{m}_1(13) \rightarrow \hat{m}_2(13) \rightarrow \hat{m}_1(14) \\
& \rightarrow \hat{m}_1(21) \rightarrow \hat{m}_2(21) \rightarrow \hat{m}_1(22) \rightarrow \hat{m}_2(22) \rightarrow \hat{m}_1(23) \rightarrow \hat{m}_2(23) \rightarrow \hat{m}_1(24) \\
& \rightarrow \hat{m}_1(31) \rightarrow \hat{m}_2(31) \rightarrow \hat{m}_1(32) \rightarrow \hat{m}_2(32) \rightarrow \hat{m}_1(33) \rightarrow \hat{m}_2(33) \rightarrow \hat{m}_1(34)
\end{aligned}$$

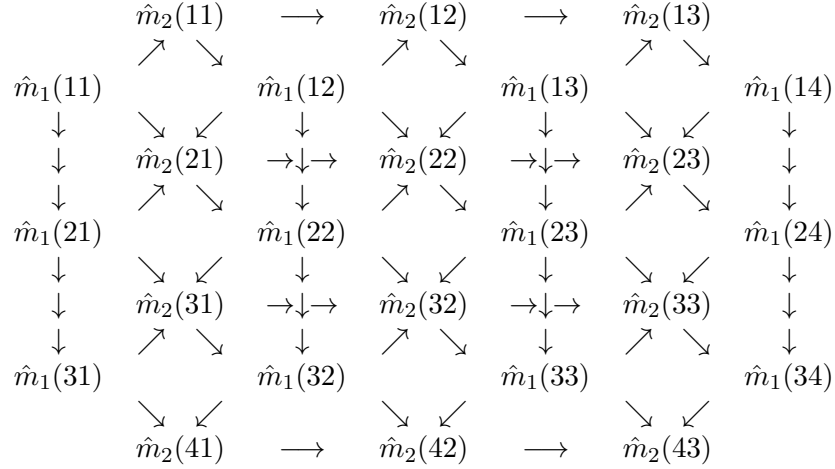


Figure 4.8. The dependency graph corresponds to layering order s_2 .

$$\rightarrow \hat{m}_2(41) \rightarrow \hat{m}_2(42) \rightarrow \hat{m}_2(43).$$

Following a similar analysis as in the single dimensional SWSC scheme, one can show that the rate region $\mathcal{R}_{j_2}(p)$ is achievable. Again, we observe that each row of the above decoding order is the same as that of a single dimensional SWSC scheme for a target rate region $\mathcal{R}_{j_2}(p)$.

This completes the proof of Theorem 4.4.2. As a simple corollary of this theorem, we have the following.

Corollary 4.4.1. *In a 2-sender L -receiver interference networks $p(y_1, y_2, \dots, y_L | x_1, x_2)$, fix a 2-2 split $p = p(u, x_1)p(v, x_2)$ and two-dimensional encoding as illustrated in Figure 4.5, the achievable rate region of the two-dimensional SWSC scheme is given by*

$$\bigcap_{j=1}^L \left(\bigcup_{i=0}^6 \mathcal{R}_{j_i}(p) \right).$$

4.4.2 SWSC achieves the Han–Kobayashi Inner Bound

In this section, we show how a SWSC scheme achieves the Han–Kobayashi inner bound [34] for two-user interference channels.

The original Han–Kobayashi coding scheme in [34] is illustrated in Figure 4.9.

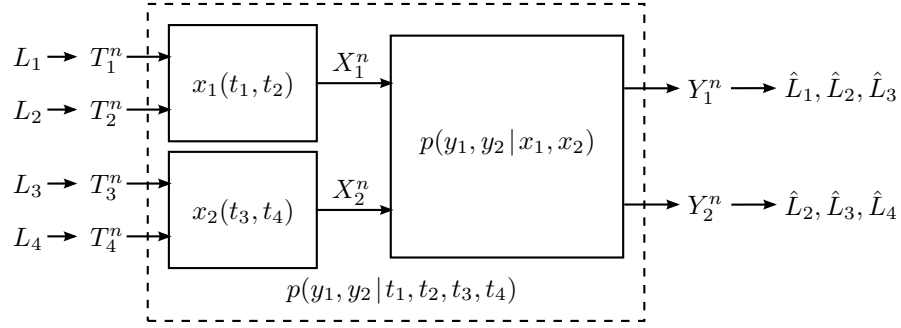


Figure 4.9. Han-Kobayashi coding scheme.

Message M_1 is split into two independent parts (L_1, L_2) and message M_2 is split into two independent parts (L_3, L_4) . Message $L_j, j \in [1 : 4]$, is carried by codeword $T_j^n(L_j)$. Then the channel inputs X_1^n and X_2^n are formed using two symbol-by-symbol mappings $x_1(t_1, t_2)$ and $x_2(t_3, t_4)$. Receiver 1 uniquely decodes $(\hat{L}_1, \hat{L}_2, \hat{L}_3)$ upon receiving Y_1^n , while receiver 2 uniquely decodes $(\hat{L}_2, \hat{L}_3, \hat{L}_4)$ upon receiving Y_2^n . Through the two symbol-by-symbol functions, the original two-user interference channel is transformed into a four-sender two-receiver interference channel with conditional pmf

$$p(y_1, y_2 | t_1, t_2, t_3, t_4) = p(y_1, y_2 | x_1(t_1, t_2), x_2(t_3, t_4)).$$

The Han-Kobayashi inner bound for the two-user interference channel $p(y_1, y_2 | x_1, x_2)$ is given by

$$\bigcup_p \text{Proj}_{4 \rightarrow 2}(\mathcal{R}'_1(p) \cap \mathcal{R}'_2(p)). \quad (4.17)$$

Here the input is of the form $p = (\prod_{j=1}^4 p(t_j))p(x_1 | t_1, t_2)p(x_2 | t_3, t_4)$, where $p(x_1 | t_1, t_2)$ and $p(x_2 | t_3, t_4)$ are $\{0, 1\}$ -valued according to functions $x_1(t_1, t_2)$ and $x_2(t_3, t_4)$. The rate region $\mathcal{R}'_1(p)$ is the set of rate triples (R'_1, R'_2, R'_3) such that

$$\sum_{j \in \mathcal{J}} R'_j \leq I(T_{\mathcal{J}}; Y_1 | T_{[1:3] \setminus \mathcal{J}})$$

for all $\mathcal{J} \subseteq [1 : 3]$ and $T_{\mathcal{J}} := (T_j : j \in \mathcal{J})$ for an index set \mathcal{J} . The rate region $\mathcal{R}'_2(p)$ is

the set of rate triples (R'_2, R'_3, R'_4) such that

$$\sum_{j \in \mathcal{J}} R'_j \leq I(T_{\mathcal{J}}; Y_2 | T_{[2:4] \setminus \mathcal{J}})$$

for all $\mathcal{J} \subseteq [2 : 4]$. The operator $\text{Proj}_{4 \rightarrow 2}$ is to apply the Fourier–Motzkin elimination from the 4-dimensional rate region (R'_1, R'_2, R'_3, R'_4) to the 2-dimensional rate region $(R_1, R_2) = (R'_1 + R'_2, R'_3 + R'_4)$.

Theorem 4.4.3. *The Han–Kobayashi inner bound (4.17) is achievable using a two-dimensional SWSC scheme.*

Remark 4.4.6. *The rate region (4.17) can be improved by coded timesharing [27]. But to simplify notation, here we only present a SWSC scheme that achieves the rate region (4.17). Coded timesharing can be incorporated on top of it later.*

In order to show the achievability of any rate tuple (I_1, I_2, I_3, I_4) in $\mathcal{R}'_1(p) \cap \mathcal{R}'_2(p)$ in (4.17), we need to show there exists a layering order for $\{U_1, U_2, U_3, U_4, V_1, V_2\}$ that achieves the rate triple (I_1, I_2, I_3) in $\mathcal{R}'_1(p)$ and another layering order for $\{U_1, U_2, V_1, V_2, V_3, V_4\}$ that achieves the rate triple (I_2, I_3, I_4) in $\mathcal{R}'_2(p)$. In the following, we show this is indeed the case.

Let us first understand the Han–Kobayashi inner bound (4.17). We note that $\mathcal{R}'_1(p)$ is the standard rate region for a three-user multiple access channel (MAC)

$$p(y_1 | t_1, t_2, t_3) = \sum_{t_4, y_2, x_1, x_2} p(y_1, y_2 | x_1, x_2) p(x_1 | t_1, t_2) p(x_2 | t_3, t_4) p(t_4)$$

and $\mathcal{R}'_2(p)$ is the standard rate region for a three-user MAC

$$p(y_2 | t_2, t_3, t_4) = \sum_{t_1, y_1, x_1, x_2} p(y_1, y_2 | x_1, x_2) p(x_1 | t_1, t_2) p(x_2 | t_3, t_4) p(t_1).$$

The auxiliary region in (R'_1, R'_2, R'_3, R'_4) is the intersection of two three dimensional MAC rate regions, where dimensions R'_2 and R'_3 are in common.

Let us recall the rate-splitting multiple access result by Grant, Rimoldi, Urbanke, and Whiting. Consider a three-user MAC $p(y|x_1, x_2, x_3)$. Fix $p = p(x_1)p(x_2)p(x_3)$, the achievable rate region $\mathcal{R}(p)$ is the set of rate triples (R_1, R_2, R_3) such that

$$\sum_{j \in \mathcal{J}} R_j \leq I(X_{\mathcal{J}}; Y | X_{[1:3] \setminus \mathcal{J}})$$

for all $\mathcal{J} \in [1:3]$.

Lemma 4.4.1 ([32]). *Any rate triple (I_1, I_2, I_3) in $\mathcal{R}(p)$ is achievable by splitting two signals of $\{X_1, X_2, X_3\}$ into two layers each, keeping one signal unsplit, and choosing a proper layering order. In particular, there exist a 2-2-1 split $p(u_1)p(u_2)p(v_1)p(v_2)p(x_3)$ and functions $x_1(u_1, u_2), x_2(v_1, v_2)$ and a layering order for $\{U_1, U_2, V_1, V_2, X_3\}$ that achieves (I_1, I_2, I_3) . Moreover, there also exist a 2-1-2 split that achieves $p(s_1)p(s_2)p(t_1)p(t_2)p(x_3)$ and functions $x_1(s_1, s_2), x_3(t_1, t_2)$ and a layering order for $\{S_1, S_2, T_1, T_2, X_3\}$ that achieves the same point.*

Now, fix $p = \prod_{j=1}^4 p(t_j)$ and functions $x_1(t_1, t_2)$ and $x_2(t_3, t_4)$, which determine the regions $\mathcal{R}'_1(p)$ and $\mathcal{R}'_2(p)$. In order to achieve (I_1, I_2, I_3) in $\mathcal{R}'_1(p)$, we keep T_2 unsplit and split T_1 and T_3 into two layers. By Lemma 4.4.1, there exists $p(u_3)p(u_4)p(v_1)p(v_2)p(t_2)$ and functions $t_1(u_3, u_4), t_2(v_1, v_2)$ and a layering order on $\{U_3, U_4, V_1, V_2, T_2\}$ that achieves (I_1, I_2, I_3) . In order to achieve (I_2, I_3, I_4) in $\mathcal{R}'_2(p)$, we keep T_3 unsplit and split T_2 and T_4 into two layers. By Lemma 4.4.1, there exists $p(u_1)p(u_2)p(v_3)p(v_4)$ and functions $t_2(u_1, u_2), t_4(v_3, v_4)$ and a layering order on $\{U_1, U_2, V_3, V_4, T_3\}$ that achieves (I_2, I_3, I_4) . This fully specified the signal splitting

$$\prod_{i=1}^4 p(u_i)p(v_i)$$

and functions

$$x_1(u_1, u_2, u_3, u_4) := x_1(t_1(u_3, u_4), t_2(u_1, u_2))$$

and

$$x_2(v_1, v_2, v_3, v_4) := x_2(t_3(v_1, v_2), t_4(v_3, v_4)).$$

We are ready to describe our SWSC scheme. We will switch from heterogeneous superposition coding as in all above SWSC scheme to homogeneous superposition coding, where independent layers are generated first and mapped to physical channel input through a symbol-by-symbol mapping [22].

Fix the distribution on (U^4, V^4, T^2, X^2) as above. Encoding is done in two-dimension over b^2 blocks. We split the message L_1 into two independent parts (L'_1, L''_1) and message L_4 into two independent parts (L'_4, L''_4) . For $i_1, i_2 \in [1 : b]$, messages $L'_1(i_1, i_2)$, $L''_1(i_1, i_2)$, $L'_4(i_1, i_2)$, and $L''_4(i_1, i_2)$ are carried by the codewords $U_3^n(L'_1)$, $U_4^n(L''_1)$, $V_3^n(L'_4)$, and $V_4^n(L''_4)$ respectively in block (i_1, i_2) . Message L_2 and L_3 are kept unsplit and are carried by (U_1, U_2) and (V_1, V_2) through a two-dimensional SWSC scheduling. That is, message $L_2(i_1, i_2)$ is carried by $U_1^n(L_2)$ in block $(i_1 + 1, i_2)$ and $U_2^n(L_2)$ in block (i_1, i_2) and message $L_3(i_1, i_2)$ is carried by $V_1^n(L_3)$ in block $(i_1, i_2 + 1)$ and $V_2^n(L_3)$ in block (i_1, i_2) , as illustrated in Figure 4.10.

U_1, V_1 U_2, V_2	$i_2 = 1$	2	3	$i_2 = 4$
$i_1 = 1$	1, 1 $l_2(11), l_3(11)$	1, $l_3(11)$ $l_2(12), l_3(12)$	1, $l_3(12)$ $l_2(13), l_3(13)$	1, $l_3(13)$ $l_2(14), 1$
2	$l_2(11), 1$ $l_2(21), l_3(21)$	$l_2(12), l_3(21)$ $l_2(22), l_3(22)$	$l_2(13), l_3(22)$ $l_2(23), l_3(23)$	$l_2(1, 4), l_3(23)$ $l_3(1, b_2), 1$
3	$l_2(21), 1$ $l_2(31), l_3(31)$	$l_2(22), l_3(31)$ $l_2(32), l_3(32)$	$l_2(23), l_3(32)$ $l_2(33), l_3(33)$	$l_2(24), l_3(33)$ $l_2(34), l_3(34)$
$i_1 = 4$	$l_2(31), 1$ 1, $l_3(41)$	$l_2(32), l_3(41)$ 1, $l_3(42)$	$l_2(33), l_3(42)$ 1, $l_3(43)$	$l_2(34), l_3(43)$ 1, 1

Figure 4.10. Two-dimensional message scheduling for L_2 and L_3 .

Based on the layering order that achieves the target point at each receiver, we

need to show a valid decoding order exists. First note that the encoding of L_1 and L_4 are done in the regular rate splitting manner and do not involve coding over multiple blocks. So we only need to check the dependency graph for messages L_2 and L_3 and consider the restricted layering order for $\{U_1, U_2, V_1, V_2\}$. Since we scheduled the two messages with the two-dimensional SWSC, Theorem 4.4.2 ensures that any layering order for $\{U_1, U_2, V_1, V_2\}$ is achievable. Therefore, a valid decoding order exists. This completes the proof of Theorem 4.4.3.

4.5 Discussion

In this Chapter, we presented a random coding scheme that achieves the simultaneous decoding inner bound for the two-user interference channel with successive cancellation and sliding-window decoding. All components involved in the coding scheme have low-complexity implementations using COTS codes. This is an alternative low-complexity coding techniques (in addition to the polar coding technique presented in Chapter 3) for interference management in future wireless networks.

Perhaps interestingly, the way the polar coding scheme and the SWSC scheme resolve incompatibility in interference channels are inherently related. The polar coding scheme produces universality by sending the same bit twice over to bit-channels, while the SWSC scheme provides robustness by sending the same message twice over two blocks. However, it seems that the extensions of the two schemes to more than two-user interference networks are quite different. It would be interesting to fully explore the connections between the two scheme.

Acknowledgment

This chapter is, in part, a reprint of the material in the papers: Lele Wang, Eren Şaşıoğlu, and Young-Han Kim, “Sliding-window superposition coding for interference networks,” to be submitted to *IEEE Transactions on Information Theory*; and Hosung

Park, Young-Han Kim, and Lele Wang, "Interference management via sliding-window superposition coding," *Proceedings of IEEE GLOBECOM*, Austin, TX, December 2014.

Chapter 5

Broadcast Channels

In this chapter, we consider another important random coding scheme, *superposition coding*, in broadcast channels. We compare the two variants of superposition coding schemes, by Cover [22] and by Bergmans [10], under the maximum likelihood decoding. We show that Cover’s superposition coding scheme can achieve rates strictly larger than Bergmans’s scheme. Based on this fact, we then propose a polar coding scheme that achieves the rate region given by Cover’s superposition coding and maximum likelihood decoding.

5.1 Introduction

Superposition coding is one of the fundamental building blocks of coding schemes in network information theory. This idea was first introduced by Cover in 1970 at the IEEE International Symposium on Information Theory, Noordwijk, the Netherlands, in a talk titled “Simultaneous Communication,” and appeared in his 1972 paper [22]. Subsequently, Bergmans [10] adapted Cover’s superposition coding scheme to the general degraded broadcast channel (this scheme is actually applicable to any nondegraded broadcast channel), which establishes the capacity region along with the converse proof by Gallager [29]. Since then, superposition coding has been applied in numerous prob-

lems, including multiple access channels [32], interference channels [14,18,34], relay channels [24], channels with feedback [25,58], and wiretap channels [17,26].

The objective of superposition coding is to communicate two messages simultaneously by encoding them into a single signal in two layers. A “better” receiver of the signal can then recover the messages on both layers while a “worse” receiver can recover the message on the coarse layer of the signal and ignore the one on the fine layer.

There are two variants of the superposition coding idea in the literature, which differ in how the codebooks are generated. The first variant is described in Cover’s original 1972 paper [22] and later in [72] and [23]. Both messages are first encoded independently via separate random codebooks of auxiliary sequences. To send a message pair, the auxiliary sequences associated with each message are then mapped through a symbol-by-symbol superposition function (such as addition) to generate the actual codeword. One can visualize the image of one of the codebooks centered around a fixed codeword from the other as a “cloud” (see the illustration in Figure 5.1(a)). Since all clouds are images of the same random codebook (around different cloud centers), we refer to this variant as *homogeneous superposition coding*. Note that in this variant, the messages are treated equally and the corresponding auxiliary sequences play the same role. Thus, there is no natural distinction between “coarse” and “fine” layers and there are two ways to group the resulting superposition codebook into clouds.

The second variant was introduced in Bergmans’s 1973 paper [10]. Here, the coarse message is encoded into a random codebook of auxiliary sequences. For each auxiliary sequence, a random satellite codebook is generated conditionally independently to represent the fine message. This naturally results in clouds of codewords around each auxiliary sequence. Since all clouds are generated independently, we refer to this variant as *heterogeneous superposition coding*. This is illustrated in Figure 5.1(b).

A natural question is whether these two variants are fundamentally different, and if so, which of the two is preferable. It is known that both variants achieve the capacity region of the degraded broadcast channel [10]. For the two-user-pair interference chan-

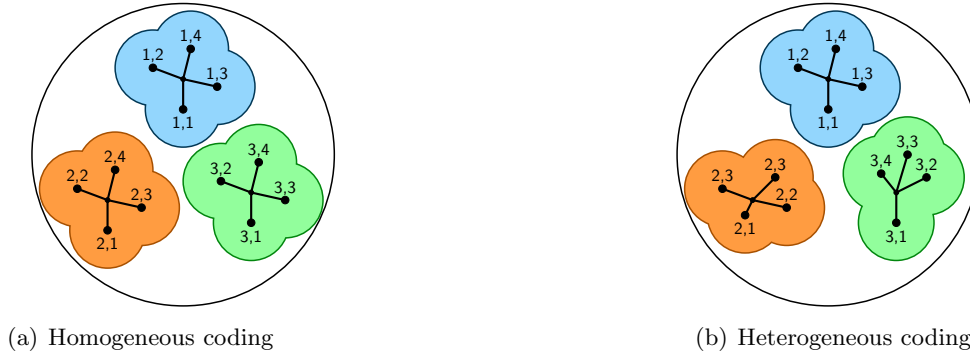


Figure 5.1. Superposition codebooks for which (a) the structure within each cloud is identical and (b) the structure is nonidentical between clouds. Codewords (dots) are annotated by “ m_1, m_2 ”, where m_1 is the coarse layer message and m_2 is the fine layer message.

nel, the two variants again achieve identical rate regions, namely, the Han–Kobayashi inner bound (see [34] for homogeneous superposition coding and [18] for heterogeneous superposition coding). Since heterogeneous superposition coding usually yields a simpler characterization of the achievable rate region with fewer auxiliary random variables, one may be tempted to prefer this variant.

However, we show in this chapter that homogeneous superposition coding always achieves a rate region at least as large as that of heterogeneous superposition coding for two-user broadcast channels, provided that the optimal (maximum likelihood) decoding rule is used. Furthermore, this dominance can be strict in general. Intuitively speaking, homogeneous superposition coding results in more structured interference from the undesired layer and this structure can be exploited under optimal decoding.

5.2 Rate Regions for the Two-Receiver BC

Consider the two-receiver discrete memoryless broadcast channel depicted in Figure 5.2. The sender wishes to communicate message M_1 to receiver 1 and message M_2 to receiver 2. We define a $(2^{nR_1}, 2^{nR_2}, n)$ code by an encoder $x^n(m_1, m_2)$ and two decoders $\hat{m}_1(y_1^n)$ and $\hat{m}_2(y_2^n)$. We assume that the message pair (M_1, M_2) is uniformly distributed over $[1 : 2^{nR_1}] \times [1 : 2^{nR_2}]$, i.e., the messages are independent of each other.

The average probability of error is defined as $P_e^{(n)} = \mathbb{P}\{(M_1, M_2) \neq (\hat{M}_1, \hat{M}_2)\}$. A rate pair (R_1, R_2) is said to be achievable if there exists a sequence of $(2^{nR_1}, 2^{nR_2}, n)$ codes such that $\lim_{n \rightarrow \infty} P_e^{(n)} = 0$.

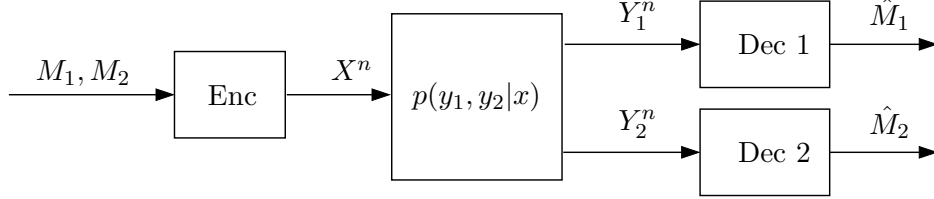


Figure 5.2. Two-receiver broadcast channel.

We now describe the two variants of superposition coding and compare their achievable rate regions for this channel under optimal decoding.

5.2.1 Homogeneous Superposition Coding (*UV* Scheme)

Codebook generation. Fix a pmf $p(u)p(v)$ and a function $x(u, v)$. Randomly and independently generate 2^{nR_1} sequences $u^n(m_1)$, $m_1 \in [1 : 2^{nR_1}]$, from $\prod_{i=1}^n p_U(u_i)$, and 2^{nR_2} sequences $v^n(m_2)$, $m_2 \in [1 : 2^{nR_2}]$ from $\prod_{i=1}^n p_V(v_i)$.

Encoding. To send the message pair (m_1, m_2) , transmit x_1^n , where $x_i = x(u_i(m_1), v_i(m_2))$.

Decoding. Both receivers use simultaneous nonunique decoding, which achieves the same rate region as maximum likelihood decoding [8] for this codebook ensemble. In particular, upon receiving y_1^n , receiver 1 declares that \hat{m}_1 was sent if it is the unique message such that

$$(u^n(\hat{m}_1), v^n(m_2), y_1^n) \in \mathcal{T}_\epsilon^{(n)}$$

for some m_2 . If there is no unique \hat{m}_1 , it declares an error. Similarly, upon receiving y_2^n , receiver 2 declares that \hat{m}_2 was sent if it is the unique message such that

$$(u^n(m_1), v^n(\hat{m}_2), y_2^n) \in \mathcal{T}_\epsilon^{(n)}$$

for some m_1 . If there is no unique \hat{m}_2 , it declares an error. Standard typicality arguments show that receiver 1 will succeed if

$$R_1 < I(U; Y_1) \quad \text{or} \quad \begin{aligned} R_1 + R_2 &< I(X; Y_1), \\ R_1 &< I(X; Y_1 | V), \end{aligned} \quad (5.1)$$

or, equivalently, if

$$\begin{aligned} R_1 &< I(X; Y_1 | V), \\ R_1 + \min\{R_2, I(X; Y_1 | U)\} &< I(X; Y_1). \end{aligned}$$

Similarly, receiver 2 will succeed if

$$R_2 < I(V; Y_2) \quad \text{or} \quad \begin{aligned} R_1 + R_2 &< I(X; Y_2), \\ R_2 &< I(X; Y_2 | U), \end{aligned} \quad (5.2)$$

or, equivalently, if

$$\begin{aligned} R_2 &< I(X; Y_2 | U), \\ R_2 + \min\{R_1, I(X; Y_2 | V)\} &< I(X; Y_2). \end{aligned}$$

The regions for both receivers are depicted in Table 5.1. Letting $\mathcal{R}_{UV}(p)$ denote the set of rate pairs (R_1, R_2) satisfying (5.1) and (5.2) under the given pmf $p(u)p(v)$ and function $x(u, v)$, it follows that the rate region

$$\mathcal{R}_{UV} = \text{co} \left(\bigcup_{p \in \mathcal{P}_{UV}} \mathcal{R}_{UV}(p) \right)$$

is achievable. Here, $\text{co}(\cdot)$ denotes convex hull, and \mathcal{P}_{UV} is the set of distributions of the form $p = p(u)p(v)p(x|u, v)$ where $p(x|u, v)$ represents the function $x(u, v)$.

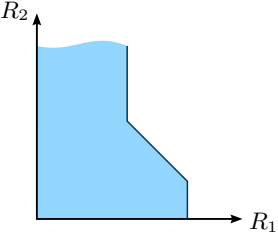
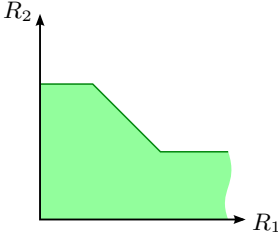
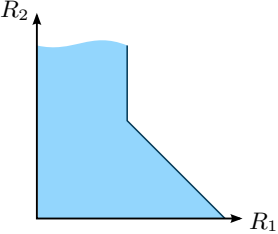
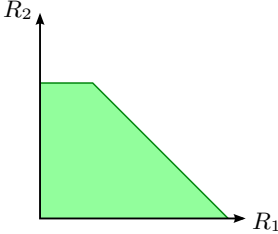
	Receiver 1	Receiver 2
$\mathcal{R}_{UV}(p)$	$R_1 < I(X; Y_1 V),$ $R_1 + \min\{R_2, I(X; Y_1 U)\} < I(X; Y_1)$ 	$R_2 < I(X; Y_2 U),$ $R_2 + \min\{R_1, I(X; Y_2 V)\} < I(X; Y_2)$ 
$\mathcal{R}_{UX}(p)$	$R_1 + \min\{R_2, I(X; Y_1 U)\} < I(X; Y_1)$ 	$R_2 < I(X; Y_2 U),$ $R_1 + R_2 < I(X; Y_2)$ 

Table 5.1. Rate regions for homogeneous and heterogeneous superposition coding.

5.2.2 Heterogeneous Superposition Coding (UX Scheme)

Codebook generation. Fix a pmf $p(u, x)$. Randomly and independently generate 2^{nR_1} sequences $u^n(m_1)$, $m_1 \in [1 : 2^{nR_1}]$ from $\prod_{i=1}^n p_U(u_i)$. For each message $m_1 \in [1 : 2^{nR_1}]$, randomly and conditionally independently generate 2^{nR_2} sequences $x^n(m_1, m_2)$, $m_2 \in [1 : 2^{nR_2}]$ from $\prod_{i=1}^n p_{X|U}(x_i | u_i(m_1))$.

Encoding. To send (m_1, m_2) , transmit $x^n(m_1, m_2)$.

Decoding. Both receivers use simultaneous nonunique decoding, which is rate-optimal as shown below. In particular, upon receiving y_1^n , receiver 1 declares \hat{m}_1 was sent if it is the unique message such that

$$(u^n(\hat{m}_1), x^n(\hat{m}_1, m_2), y_1^n) \in \mathcal{T}_\epsilon^{(n)}$$

for some m_2 . If there is no unique \hat{m}_1 , it declares an error. Similarly, upon receiving y_2^n ,

receiver 2 declares \hat{m}_2 was sent if it is the unique message such that

$$(u^n(m_1), x^n(m_1, \hat{m}_2), y_2^n) \in \mathcal{T}_\epsilon^{(n)}$$

for some m_1 . If there is no unique \hat{m}_2 , it declares an error. Standard arguments show that receiver 1 will succeed if

$$R_1 < I(U; Y_1) \quad \text{or} \quad R_1 + R_2 < I(X; Y_1), \quad (5.3)$$

or, equivalently, if

$$R_1 + \min\{R_2, I(X; Y_1 | U)\} < I(X; Y_1).$$

Similarly, receiver 2 will succeed if

$$R_2 < I(X; Y_2 | U), \quad (5.4)$$

$$R_1 + R_2 < I(X; Y_2).$$

A similar argument to the one in [8] readily shows that the region in (5.3) cannot be improved by using maximum likelihood decoding. It is also shown in the Appendix that applying maximum likelihood decoding does not improve the region in (5.4).

The regions for both receivers are depicted in Table 5.1. Let $\mathcal{R}_{UX}(p)$ denote the set of all rate pairs (R_1, R_2) satisfying both (5.3) and (5.4). Clearly, the rate region

$$\mathcal{R}_{UX} = \text{co}\left(\bigcup_{p \in \mathcal{P}_{UX}} \mathcal{R}_{UX}(p)\right)$$

is achievable. Here, \mathcal{P}_{UX} is the set of distributions of the form $p = p(u, x)$.

If the roles of m_1 and m_2 in codebook generation are reversed, one can also achieve the region $\mathcal{R}_{VX} = \text{co}(\cup_p \mathcal{R}_{VX}(p))$ obtained by swapping Y_1 with Y_2 and R_1 with R_2 in the definition of $\mathcal{R}_{UX}(p)$.

It is worth reiterating that the two schemes above differ only in the dependence/independence between clouds around different u^n sequences, and not in the underlying distributions from which the clouds are generated. Indeed, it is well known that the classes of distributions \mathcal{P}_{UX} and \mathcal{P}_{UV} are equivalent in the sense that for every $p(u, x) \in \mathcal{P}_{UX}$, there exists a $q(u)q(v)q(x|u, v) \in \mathcal{P}_{UV}$ such that $\sum_v q(u)q(v)q(x|u, v) = p(u, x)$ (see, for example, [27, p. 626]).

5.3 Main Result

Theorem 5.3.1. *The rate region achieved by homogeneous superposition coding includes the rate region achieved by heterogeneous superposition coding, i.e.,*

$$\text{co}(\mathcal{R}_{UX} \cup \mathcal{R}_{VX}) \subseteq \mathcal{R}_{UV}.$$

Moreover, there are channels for which the inclusion is strict.

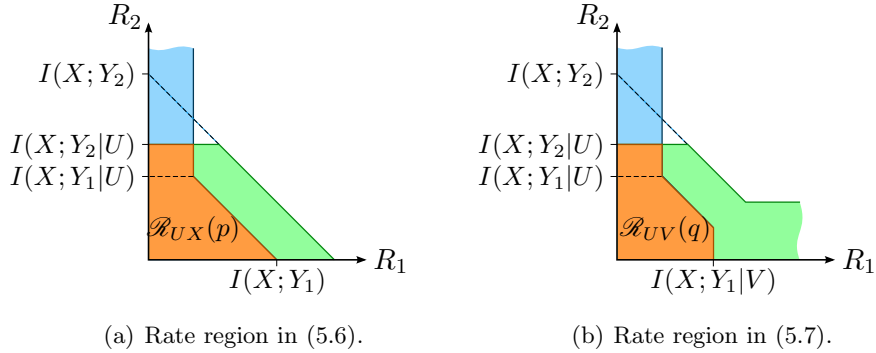


Figure 5.3. Rate regions for the proof of Theorem 5.3.1.

Proof. Due to the convexity of \mathcal{R}_{UV} and the symmetry between the definitions of \mathcal{R}_{UX} and \mathcal{R}_{VX} , it suffices to show that $\mathcal{R}_{UX}(p) \subseteq \mathcal{R}_{UV}$ for all $p \in \mathcal{P}_{UX}$. Fix a $p \in \mathcal{P}_{UX}$. Let $q' \in \mathcal{P}_{UV}$ be such that $U = X$, $V = \emptyset$, and $q'(x) = p(x)$. Let $q'' \in \mathcal{P}_{UV}$ be such that $V = X$, $U = \emptyset$, and $q''(x) = p(x)$. An inspection of (5.1)–(5.4) and Table 5.1 reveals that

$\mathcal{R}_{UV}(q')$ is the set of rate pairs satisfying

$$\begin{aligned} R_2 &= 0, \\ R_1 &< I(X; Y_1), \end{aligned}$$

and $\mathcal{R}_{UV}(q'')$ is the set of rate pairs satisfying

$$\begin{aligned} R_1 &= 0, \\ R_2 &< I(X; Y_2). \end{aligned}$$

It then follows that $\text{co}(\mathcal{R}_{UV}(q') \cup \mathcal{R}_{UV}(q''))$ includes the rate region

$$R_1 + R_2 < \min \{I(X; Y_1), I(X; Y_2)\}. \quad (5.5)$$

We will consider three cases and prove the claim for each.

- If $I(X; Y_1) \geq I(X; Y_2)$ then $\mathcal{R}_{UX}(p)$ reduces to the rate region

$$\begin{aligned} R_2 &< I(X; Y_2 | U), \\ R_1 + R_2 &< I(X; Y_2), \end{aligned}$$

which is included in the rate region in (5.5), and therefore in \mathcal{R}_{UV} .

- If $I(X; Y_1) < I(X; Y_2)$ and $I(X; Y_1 | U) \geq I(X; Y_2 | U)$, then $\mathcal{R}_{UX}(p)$ reduces to the rate region

$$\begin{aligned} R_2 &< I(X; Y_2 | U), \\ R_1 + R_2 &< I(X; Y_1), \end{aligned}$$

which is also included in the rate region in (5.5), and therefore in \mathcal{R}_{UV} .

- If $I(X; Y_1) < I(X; Y_2)$ and $I(X; Y_1 | U) < I(X; Y_2 | U)$, then $\mathcal{R}_{UX}(p)$ reduces to the rate region (see Figure 5.3(a))

$$\begin{aligned} R_2 &< I(X; Y_2 | U), \\ R_1 + \min\{R_2, I(X; Y_1 | U)\} &< I(X; Y_1). \end{aligned} \tag{5.6}$$

Find a $q \in \mathcal{P}_{UV}$ with $q(u, x) = p(u, x)$, and note that $\mathcal{R}_{UV}(q)$ is described by the bounds

$$\begin{aligned} R_2 &< I(X; Y_2 | U), \\ R_1 &< I(X; Y_1 | V), \\ R_1 + \min\{R_2, I(X; Y_1 | U)\} &< I(X; Y_1). \end{aligned} \tag{5.7}$$

Comparing (5.6) with (5.7) (Figure 5.3(b)), one sees that $\mathcal{R}_{UX}(p) \subseteq \text{co}(\mathcal{R}_{UV}(q) \cup \mathcal{R}_{UV}(q'))$. This proves the first claim of the theorem.

To show that the inclusion can be strict, consider the vector broadcast channel with binary inputs (X_1, X_2) and outputs $(Y_1, Y_2) = (X_1, X_2)$. For all $p \in \mathcal{P}_{UX}$, we have from (5.4) that $R_1 + R_2 < I(X_1 X_2; Y_2) < 1$, and thus \mathcal{R}_{UX} is included in the rate region $R_1 + R_2 < 1$, and by symmetry, so is \mathcal{R}_{VX} . Note, however, that the rate pair $(1, 1)$ is achievable by homogeneous superposition coding, setting $U = X_1$ and $V = X_2$. This proves the second claim. \square

5.4 Polar Coding for Cover's Superposition Coding

We now present a polar coding scheme that achieves the rate region \mathcal{R}_{UV} . We note that through the symbol-by-symbol mapping $x(u, v)$, the two-user broadcast channel $p(y_1, y_2 | x)$ can be transformed into a two-user interference channel

$$p(y_1, y_2 | u, v) = P(y_1, y_2 | x(u, v)),$$

where sender $j \in \{1, 2\}$ wishes to communicate an independent message M_j to receiver j , as illustrated in Figure 5.4.

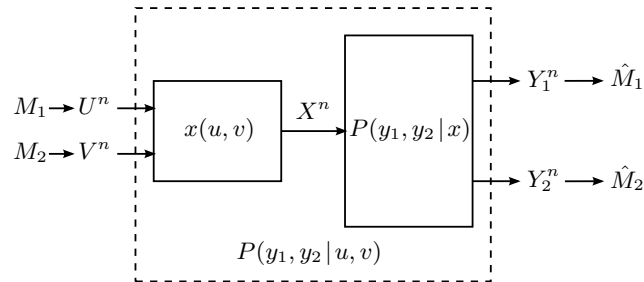


Figure 5.4. Illustration of Cover's superposition coding as coding for two-user interference channels.

Moreover, one can write an equivalent rate region of \mathcal{R}_{UV} as

$$\mathcal{R}_{UV} = \text{co} \left(\bigcup_p \bigcup_{i=1}^4 (\mathcal{R}_{1i}(p) \cap \mathcal{R}_{2i}(p)) \right),$$

where the distribution is of the form $p = p(u)p(v)$ and $x(u, v)$ and $\mathcal{R}_{1i}(p) \cap \mathcal{R}_{2i}(p)$ corresponds to the rate region when the decoder $j = 1, 2$ is required to *uniquely* recover message sets $\hat{M}_{\mathcal{A}_j}$. For $i = 1, 2, 3, 4$, the corresponding message sets are

$$i = 1: \mathcal{A}_1 = \{1\}, \mathcal{A}_2 = \{2\};$$

$$i = 2: \mathcal{A}_1 = \{1, 2\}, \mathcal{A}_2 = \{2\};$$

$$i = 3: \mathcal{A}_1 = \{1\}, \mathcal{A}_2 = \{1, 2\};$$

$$i = 4: \mathcal{A}_1 = \{1, 2\}, \mathcal{A}_2 = \{1, 2\}.$$

Now we can readily apply the polar coding scheme developed in Chapter 3 to achieve each of the region $\mathcal{R}_{1i}(p) \cap \mathcal{R}_{2i}(p)$, $i = 1, 2, 3, 4$.

For example, $\mathcal{R}_{13}(p) \cap \mathcal{R}_{23}(p)$ is the set of (R_1, R_2) such that

$$R_1 < I(U; Y_1),$$

$$R_1 < I(U; Y_2, V),$$

$$R_2 < I(V; Y_2, U),$$

$$R_1 + R_2 < I(U, V; Y_2).$$

To achieve arbitrary point here, one can first find a good point-in-point code for $\mathcal{R}_{13}(p)$ and a monotone permutation for the MAC region $\mathcal{R}_{23}(p)$. Then apply method in 3.4.1 to align the incompatible indices in the code for U^n . This achieves any point in the rate region $\mathcal{R}_{13}(p) \cap \mathcal{R}_{23}(p)$. Similarly for each decoding set, one can design a corresponding polar coding scheme based on the method above. Therefore, the proposed polar coding scheme achieves the optimal rate region given Cover's superposition encoding. The generalization to L -user broadcast channels can be done similarly.

As a side remark, the independence between U and V in Cover's superposition coding is important for transforming the broadcast channel into a two-user interference channel. For general correlated $(U, V) \sim p(u, v)$ as in Marton coding for broadcast channels, one needs different techniques. A method for Marton coding as well as an alternative polar coding scheme for Bergmans's superposition coding can be found in [56] and [31].

5.5 Discussion

In addition to the basic superposition coding schemes presented in Section 5.2, one can consider coded time sharing [27], which could potentially enlarge the achievable rate regions. In the present setting, however, it can be easily checked that coded time sharing does not enlarge \mathcal{R}_{UX} . Thus, the conclusion of Theorem 5.3.1 continues to hold and homogeneous superposition coding with coded time sharing outperforms heterogeneous superposition coding with coded time sharing.

5.6 Appendix: Optimality of the Rate Region in (5.4)

We show that the heterogeneous superposition coding ensemble cannot achieve a rate region larger than the one in (5.4) under any decoding rule. To that end, denote the random codebook by

$$\mathcal{C} = (U^n(1), U^n(2), \dots, X^n(1, 1), X^n(1, 2), \dots).$$

By Fano's inequality,

$$\begin{aligned} H(M_2 | Y_2^n, \mathcal{C}) &= \sum_c \mathbb{P}\{\mathcal{C} = c\} H(M_2 | Y_2^n, \mathcal{C} = c) \\ &\leq \sum_c \mathbb{P}\{\mathcal{C} = c\} [1 + nR_2 \mathbb{P}\{\hat{M}_2 \neq M_2 | \mathcal{C} = c\}] \\ &\leq n\epsilon_n, \end{aligned} \tag{5.8}$$

where $\epsilon_n \rightarrow 0$ as $n \rightarrow \infty$. Thus,

$$\begin{aligned} nR_2 &= H(M_2 | \mathcal{C}, M_1) \\ &\leq I(M_2; Y_2^n | \mathcal{C}, M_1) + n\epsilon_n \\ &= H(Y_2^n | \mathcal{C}, M_1) - H(Y_2^n | \mathcal{C}, M_1, M_2) + n\epsilon_n \\ &\stackrel{(a)}{\leq} nH(Y_2 | U) - nH(Y_2 | X) + n\epsilon_n \\ &= nI(X; Y_2 | U) + n\epsilon_n, \end{aligned}$$

where (a) follows by the definition of the codebook ensemble and the memoryless property.

To see the second inequality, first consider the case

$$R_1 < I(X; Y_2). \tag{5.9}$$

By (5.8), we have

$$\begin{aligned} H(M_1, M_2 | Y_2^n, \mathcal{C}) &= H(M_2 | Y_2^n, \mathcal{C}) + H(M_1 | Y_2^n, \mathcal{C}, M_2) \\ &\leq n\epsilon_n + H(M_1 | Y_2^n, \mathcal{C}, M_2). \end{aligned}$$

To bound the last term, note that given $M_2 = m_2$, the codebook reduces to

$$\mathcal{C}' = (X^n(1, m_2), X^n(2, m_2), X^n(3, m_2), \dots).$$

These codewords are pairwise independent since they do not share common U^n sequences, and thus \mathcal{C}' is a nonlayered random codebook of rate R_1 . Since (5.9) holds, receiver 2 can reliably recover M_1 from $(Y_2^n, \mathcal{C}, M_2)$ by using, for example, a typicality decoder. Thus we have

$$H(M_1 | Y_2^n, \mathcal{C}, M_2) \leq n\epsilon_n.$$

The sum-rate can then be bounded as

$$\begin{aligned} n(R_1 + R_2) &= H(M_1, M_2) \\ &\leq I(M_1, M_2; Y_2^n | \mathcal{C}) + 2n\epsilon_n \\ &\leq nI(X; Y_2) + 2n\epsilon_n. \end{aligned} \tag{5.10}$$

To conclude the argument, suppose that there exists a decoding rule that achieves a rate point (R_1, R_2) with $R_1 \geq I(X; Y_2)$. Then, this decoding rule must also achieve $(R'_1, R'_2) = (I(X; Y_2) - R_2/2, R_2)$ with the heterogeneous superposition coding ensemble, since (R_1, R_2) dominates (R'_1, R'_2) . Note that $R'_1 < I(X; Y_2)$. It thus follows from the discussion above that $R'_1 + R'_2 \leq I(X; Y_2)$, which yields a contradiction.

Acknowledgment

This chapter is, in part, a reprint of the material in the papers: Lele Wang, Eren Şaşođlu, Bernd Bandemer, and Young-Han Kim, “A comparison of superposition coding schemes,” *Proceedings of the IEEE International Symposium on Information Theory*, Istanbul, Turkey, July 2013; and Lele Wang and Eren Şaşođlu, “Polar coding for interference networks,” submitted to *IEEE Transactions on Information Theory*, 2014.

Chapter 6

Relay Channels

In this chapter, polar coding schemes are developed for the decode–forward relaying (digital-to-digital interface) and the compress–forward relaying (analog-to-digital interface) in the three-node relay channel. For decode–forward, a technique based on the recent universal polarization method is applied to create the desired nested structure. For compress–forward, existing methods are generalized to allow arbitrary input distributions and channel statistics. Both schemes achieve full theoretical rates in general relay channels.

6.1 Introduction

Consider the three-node discrete memoryless relay channel $(\mathcal{X}_1 \times \mathcal{X}_2, p(y_2, y_3|x_1, x_2), \mathcal{Y}_2 \times \mathcal{Y}_3)$ that consists of four finite sets $\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}_2, \mathcal{Y}_3$, and a collection of conditional probability mass functions (pmfs) $p(y_2, y_3|x_1, x_2)$ on $\mathcal{Y}_2 \times \mathcal{Y}_3$ (Figure 6.1). The sender wishes to communicate a message M to the receiver with the help of the relay. Let the bold-font letter $\mathbf{x} = (x_1, x_2, \dots, x_N)$ denote the vector of length N . Let $x^i = (x_1, x_2, \dots, x_i)$ for $i \neq N$.

A $(2^{NR}, N)$ code for the relay channel consists of

- a message set $[1 : 2^{NR}] := \{1, 2, \dots, 2^{NR}\}$,

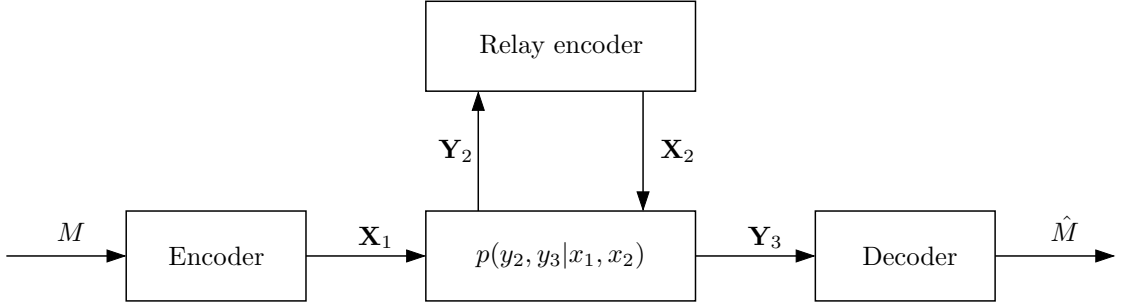


Figure 6.1. Three-node relay channel.

- an encoder that assigns a codeword $\mathbf{x}_1(m)$ to each message $m \in [1 : 2^{NR}]$,
- a relay encoder that assigns a symbol $x_{2i}(y_2^{i-1})$ to each past received sequence y_2^{i-1} at time $i \in [1 : N]$, and
- a decoder that assigns an estimate \hat{m} or an error message e to each received sequence $\mathbf{y}_3 \in \mathcal{Y}_3^N$.

We assume that the message M is uniformly distributed over the message set. The average probability of error is defined as $P_e^{(n)} = \mathbb{P}\{\hat{M} \neq M\}$. A rate R is said to be *achievable* if there exists a sequence of $(2^{NR}, N)$ codes such that $\lim_{N \rightarrow \infty} P_e^{(n)} = 0$. The *capacity* C of the relay channel is the supremum of all achievable rates.

The capacity of the general relay channel is not known. Many information-theoretic relaying techniques have been developed over the past four decades. Among them, two primitive techniques are decode–forward and compress–forward. For decode–forward, the relay treats received sequence \mathbf{y}_2 as digital signals, recovers the message from it, and coherently cooperates with the sender to communicate the message to the receiver. The decode–forward lower bound is given by [24]

$$C \geq R_{\text{DF}} := \max_{p(x_1, x_2)} \min\{I(X_1, X_2; Y_3), I(X_1; Y_2 | X_2)\}. \quad (6.1)$$

In comparison, for compress–forward, the relay treats the received sequence \mathbf{y}_2 as analog signals, applies Wyner–Ziv coding with the receiver’s sequence \mathbf{y}_3 acting as side informa-

tion, and forwards the compression index. The compress–forward lower bound is given by [24]

$$C \geq R_{\text{CF}} := \max I(X_1; \hat{Y}_2, Y_3, X_2), \quad (6.2)$$

where the maximum is over all conditional pmfs $p(x_1)p(x_2)p(\hat{y}_2|x_2, y_2)$ such that

$$I(X_2; Y_3) \geq I(Y_2; \hat{Y}_2|X_2, Y_3).$$

The goal of this chapter is to show the achievability of the two lower bounds using polar coding techniques.

Polar coding schemes for decode–forward have been proposed in *degraded* relay channels, where $X_1 \rightarrow (X_2, Y_2) \rightarrow Y_3$ form a Markov chain [3, 12, 40, 41]. Under the degradation condition, the two polarization processes involved in the decode–forward scheme possess a nested structure. Obtaining such nested structure in general relay channels is nontrivial [35]. This is because the indices of good bit-channels are different for different underlying channels (not universal) by regular polar transform [7]. Recently, two universal polar coding methods were proposed [36, 63], which provide tools to resolve this difficulty. In section 6.2, we propose a novel polar coding scheme that generalizes previous polar coding schemes to arbitrary three-node relay channels. By carefully incorporating the universal polarization techniques in [63], the proposed scheme creates the desired nested structure and thus achieves the decode–forward lower bound (6.1).

A polar coding scheme for compress–forward has been proposed in relay channels with orthogonal receiver components, where $Y_3 = (Y'_3, Y''_3)$ and $p(y_2, y_3|x_1, x_2) = p(y'_3, y_2|x_1)p(y''_3|x_2)$ [12]. The scheme achieves the so called *symmetric* compress–forward rate, which is strictly smaller the compress–forward lower bound (6.2). In Section 6.3, we extend this scheme to general relay channels. While the key ideas in our coding scheme resemble that in [12], the contribution in this part is the probability of error analysis technique that cleans up the unnecessary constraints imposed on the channel statistics and the input distributions and thus establishes a polar coding scheme that achieves the

compress–forward lower bound (6.2) in general relay channels.

6.2 Decode–Forward Relaying

We first write the decode–forward lower bound (6.1) in an alternative way. The proof is skipped due to space limitations.

Lemma 6.2.1. *The decode–forward rate can be expressed as*

$$R_{DF} = \max \min \{ I(X_2; Y_3) + I(V; Y_3, X_2), I(V; Y_2, X_2) \},$$

where the maximum is over all pmfs $p(v)p(x_2)$ and functions $x_1(v, x_2)$ such that $V \sim \text{Unif}[1 : q]$ for some prime q .

Proof. By the functional representation lemma [27, Appendix ?], for any conditional pmf $p(x_1|x_2)$, there exists a random variable V independent of X_2 such that X_1 can be represented by a function $x_1(v, x_2)$. If V is not uniform, we let $\tilde{V} \sim \text{Unif}[1 : q]$ for some $q \geq |\mathcal{V}|$ and choose the function $f: \tilde{\mathcal{V}} \rightarrow \mathcal{V}$ to construct V from uniform \tilde{V} . Note that \tilde{V} can be chosen independent of X_2 and X_1 can thus be represented as $x_1(v, x_2) = x_1(f(\tilde{v}), x_2) := \tilde{x}_1(\tilde{v}, x_2)$. Therefore, in expression (6.1), there is no loss of generality in restricting the input pmfs to $V \sim \text{Unif}[1 : q]$ independent of X_2 and functions $x_1(v, x_2)$. Under this input pmf, we have

$$\begin{aligned} I(X_1, X_2; Y_3) &= I(X_2; Y_3) + I(X_1; Y_3 | X_2) \\ &\stackrel{(a)}{=} I(X_2; Y_3) + I(V, X_1; Y_3 | X_2) \\ &\stackrel{(b)}{=} I(X_2; Y_3) + I(V; Y_3 | X_2) \\ &\stackrel{(c)}{=} I(X_2; Y_3) + I(V; Y_3, X_2), \end{aligned}$$

where (a) follows since $V \rightarrow (X_1, X_2) \rightarrow Y_3$ form a Markov chain, (b) follows since X_1 is a function of V and X_2 , and (c) follows since V and X_2 are independent. For a similar

reason, we have $I(X_1; Y_2 | X_2) = I(V; Y_2, X_2)$. \square

In the decode–forward coding scheme, the sender broadcasts the message to both the relay and the receiver. The relay recovers the message and sends the *bin index* of the message to help receiver recover the message. The code design requirements are summarized as follows.

1. We need to assign a bin index $l \in [1 : 2^{NR_2}]$ for each message $m \in [1 : 2^{NR}]$.
2. We need a code to send the bin index l of the message m through the point-to-point channel

$$p(y_3 | x_2) = \sum_{v, x_1, y_2} p(v) p(x_1 | v, x_2) p(y_2, y_3 | x_1, x_2).$$

3. We need a code to broadcast the message m through two point-to-point channels

$$p(y_3, x_2 | v) = \sum_{x_1, y_2} p(x_2) p(x_1 | v, x_2) p(y_2, y_3 | x_1, x_2),$$

$$p(y_2, x_2 | v) = \sum_{x_1, y_3} p(x_2) p(x_1 | v, x_2) p(y_2, y_3 | x_1, x_2),$$

such that the receiver can recover the message if its bin index l is provided as side information and the relay can recover the message without side information.

In degraded relay channels, requirement 3) can be fulfilled easily, as the good bit-channel for $p(y_3, x_2 | v)$ is also good for $p(y_2, x_2 | v)$ using regular polar transform [7]. Unfortunately, the same design is strictly suboptimal in general relay channels. We show in Section 6.2.1 how to create a similar nested index-set structure for good bit-channels using methods in [63]. The coding scheme is presented in Section 6.2.2.

6.2.1 Polarization for Broadcasting

Let us first recall the universal polarization method in [63]. A universal polar transform of rate R is a one-to-one recursive transform such that R fraction of good

channels and $(1 - R)$ fraction of bad channels are created at each recursion. As the number of recursions increases, the good channels become increasingly good, and the bad channels become increasingly bad. The indices of the good channels and bad channels are *independent* of the underlying channels, ensuring universality. Thus, if the good channels converge to perfect channels, we can design a universal code of rate R by putting information bits in the good channels and freezing the bits in the bad channels. In [63], a systematic method to design the universal polar transform of any rational rate is introduced for the class of binary-input memoryless channels. In particular, the following polarization result is established.

Proposition 6.2.1 (Universal polarization [63]). *For any rational rate $R^* \in [0, 1]$, there exists an $N \times N$ invertible matrix T_N over \mathbb{F}_2 and a partition $[1 : N] = \mathcal{G} \uplus \mathcal{N} \uplus \mathcal{O}$ such that*

$$\lim_{N \rightarrow \infty} \frac{1}{N} |\mathcal{G}| = R^* \text{ and } \lim_{N \rightarrow \infty} \frac{1}{N} |\mathcal{N}| = 1 - R^*. \quad (6.3)$$

Let $W: X \rightarrow Y$ be a binary-input memoryless channel with symmetric capacity $I(W)$.

Let $\mathbf{U} = \mathbf{X}T_N$. We have the following properties:

1. If $I(W) \geq R^*$, then for any $i \in \mathcal{G}$,

$$\lim_{N \rightarrow \infty} I(U_i; \mathbf{Y}, U^{i-1}) = 1.$$

2. If $I(W) \leq R^*$, then for any $i \in \mathcal{N}$,

$$\lim_{N \rightarrow \infty} I(U_i; \mathbf{Y}, U^{i-1}) = 0.$$

Remark 6.2.1. *Proposition 6.2.1 continues to hold for any q -ary alphabet with prime q . One can simply change the mutual information computation to base q and apply the modulo- q arithmetics for $\mathbf{U} = \mathbf{X}T_N$.*

Now we are ready to describe the polarization method that fulfills requirement

3). Suppose that for the pmf that attains R_{DF} in Lemma 6.2.1, $I(V; Y_3, X_2) < R_{\text{DF}}$. (Otherwise, one can achieve rate $I(V; Y_3, X_2)$ by transmitting only through the direct link $V \rightarrow (Y_3, X_2)$.) Consider two underlying channels

$$W_1: V \rightarrow (Y_3, X_2) \quad \text{and} \quad W_2: V \rightarrow (Y_2, X_2).$$

Choose $R^* \lesssim R_{\text{DF}}$. We design a concatenated two-stage polarization. For the inner polarization, we apply the universal transform T_{N_1} for q -ary alphabet. By Proposition 6.2.1, the index set partition $[1 : N_1] = \mathcal{G} \uplus \mathcal{N} \uplus \mathcal{O}$ is the same for W_1 and W_2 . For the outer polarization, we concatenate regular polar transform G_{N_2} for q -ary alphabet. Define the set of indices that branch out from \mathcal{G} as

$$\mathcal{R} = \cup_{i \in \mathcal{G}} [(i-1)N_2 + 1 : i \times N_2].$$

For $\beta < 1/2$, define

$$\mathcal{A} = \{i : I(U_i; \mathbf{Y}_3, \mathbf{X}_2, U^{i-1}) > 1 - 2^{-N_2^\beta}\},$$

$$\mathcal{B} = \{i : I(U_i; \mathbf{Y}_2, \mathbf{X}_2, U^{i-1}) > 1 - 2^{-N_2^\beta}\}.$$

Theorem 6.2.1. *For $\mathcal{R}, \mathcal{B}, \mathcal{A}$ defined above,*

$$\lim_{N_1, N_2 \rightarrow \infty} \frac{1}{N_1 N_2} |\mathcal{R} \cap \mathcal{B}| = R^*, \tag{6.4}$$

$$\lim_{N_1, N_2 \rightarrow \infty} \frac{1}{N_1 N_2} |\mathcal{R} \cap \mathcal{B} \cap \mathcal{A}| = I(V; Y_3, X_2). \tag{6.5}$$

Proof. First consider the polarization for channel W_2 . Note that

$$R^* \leq R_{\text{DF}} \leq I(V; Y_2, X_2) = I(W_2),$$

where the last equality follows since $V \sim \text{Unif}[1 : q]$ in Lemma 6.2.1. By property

1) of Proposition 6.2.1, for any $\epsilon > 0$, there exists an N_1' such that after the universal polarization all bit-channels in \mathcal{G} has symmetric capacity as least $1 - \epsilon$. Since the fraction of good bit-channels in \mathcal{R} converges to the average symmetric capacity of channels in \mathcal{R} as $N_2 \rightarrow \infty$, there exists an N_2 such that

$$\frac{|\mathcal{R} \cap \mathcal{B}|}{|\mathcal{R}|} > 1 - \epsilon. \quad (6.6)$$

By equation (6.3), there exists an N_1'' such that

$$(1 + \epsilon)R^* > \frac{1}{N_1'' N_2} |\mathcal{R}| > (1 - \epsilon)R^*.$$

Setting $N_1 = \max\{N_1', N_1''\}$, we have

$$(1 + \epsilon)R^* > \frac{1}{N_1 N_2} |\mathcal{R} \cap \mathcal{B}| > (1 - \epsilon)^2 R^*,$$

establishing (6.4).

For (6.5), consider the polarization for channel W_1 . First, by regular polarization results, there exists an N_2 such that

$$I(V; Y_3, X_2) + \epsilon > \frac{1}{N_1, N_2} |\mathcal{A}| > I(V; Y_3, X_2) - \epsilon.$$

Second, since $I(W_1) = I(V; Y_3, X_2) \leq R^*$, by property 2) of Proposition 6.2.1, there exists an N_1 such that after the universal polarization, all bit-channels in \mathcal{N} has symmetric capacity at most ϵ . Counting the total mutual information of bit-channels in \mathcal{N} , we have

$$\frac{1}{N_1 N_2} |\mathcal{R}^c \cap \mathcal{A}| < \epsilon R^* / (1 - 2^{-N_2^\beta}) < \epsilon$$

Moreover, by (6.6),

$$\frac{|\mathcal{R} \cap \mathcal{A} \cap \mathcal{B}^c|}{N_1 N_2} \leq \frac{|\mathcal{R} \cap \mathcal{A} \cap \mathcal{B}^c|}{|\mathcal{R}|} \leq \frac{|\mathcal{R} \cap \mathcal{B}^c|}{|\mathcal{R}|} < \epsilon.$$

Therefore, $I(V; Y_3, X_2) + \epsilon > \frac{1}{N_1 N_2} |\mathcal{R} \cap \mathcal{B} \cap \mathcal{A}| = \frac{1}{N_1 N_2} (|\mathcal{A}| - |\mathcal{R}^c \cap \mathcal{A}| - |\mathcal{R} \cap \mathcal{A} \cap \mathcal{B}^c|) > I(V; Y_3, X_2) - 3\epsilon$. \square

By construction, the good bit-channel indices $\mathcal{R} \cap \mathcal{B} \cap \mathcal{A}$ for W_1 is a subset of the good bit-channel indices $\mathcal{R} \cap \mathcal{B}$ for W_2 . Theorem 6.2.1 ensures that these sets converge to the right fraction. Now requirement 3) can be fulfilled by sending information in $\mathcal{M} := \mathcal{R} \cap \mathcal{B}$ and freezing the rest bits in $\mathcal{F} := (\mathcal{R} \cap \mathcal{B})^c$. The receiver of channel W_1 can decode if the bits in $\mathcal{S} := \mathcal{R} \cap \mathcal{B} \cap \mathcal{A}^c$ is provided as side information.

6.2.2 Coding Scheme

Now we describe the polar coding scheme that achieves the decode-forward lower bound (6.1). We use b transmission blocks, each consisting of $N = N_1 N_2$ transmissions. A sequence of $(b - 1)$ messages $M_j, j \in [1 : b - 1]$ is sent over these b blocks.

Codebooks and rate. Fix the pmf that attains R_{DF} in Lemma 6.2.1. We use two polar codes. One is a polar code of rate $R_2 = \frac{1}{N} |\mathcal{S}|$ for the point-to-point channel $p(y_3|x_2)$ using the method in [65, Chapter 3]. The existence of such a code is guaranteed since, by Theorem 6.2.1, $\lim_{N \rightarrow \infty} \frac{1}{N} |\mathcal{S}| = R^* - I(V; Y_3, X_2) \leq R_{\text{DF}} - I(V; Y_3, X_2) \leq I(X_2; Y_3)$. Here, we abstract this code by an encoder $\mathbf{x}_2(l)$ and a decoder $\hat{l}(\mathbf{y}_3)$. The other is a polar code for broadcasting a common message over W_1 and W_2 . The polarization transform for this code is given in Section 6.2.1. We locate the message at $U_{\mathcal{M}} := \{U_i : i \in \mathcal{M}\}$. We fix $U_{\mathcal{F}} = u_{\mathcal{F}}$ and reveal its value to all parties. Thus the rate achieved by this coding scheme is $\frac{1}{N} |\mathcal{M}|$, which, by Theorem 6.2.1, can be arbitrarily close to R_{DF} for large N and good choices of R^* .

Bin assignment. The bin index for $U_{\mathcal{M}}$ is $U_{\mathcal{S}}$ (recall $\mathcal{S} \subset \mathcal{M}$). Since the message $U_{\mathcal{M}}$ is uniform, $U_{\mathcal{S}}$ is also uniform.

Encoding. In block $j \in [1 : b - 1]$, the message m_j is carried by $u_{\mathcal{M}}^{(j)}$. The sender computes $\mathbf{v}^{(j)}$ from $\mathbf{u}^{(j)}$ through the polar transform. The sender transmits $x_{1i} = x_1(v_i^{(j)}, x_{2i}(u_{\mathcal{S}}^{(j-1)}))$ at time $i \in [1 : N]$.

Relay encoding. In block $j \in [1 : b]$, upon receiving $\mathbf{y}_2^{(j)}$, the relay sets $\tilde{u}_{\mathcal{F}}^{(j)} = u_{\mathcal{F}}$ and successively recovers $\tilde{u}_{\mathcal{M}}^{(j)}$ as

$$\tilde{u}_i = \arg \max_{u \in \mathcal{U}} p_{U_i | \mathbf{Y}_2, \mathbf{X}_2, U^{i-1}}(u | \mathbf{y}_2^{(j)}, \mathbf{x}_2(\tilde{u}_{\mathcal{S}}^{(j-1)}), \tilde{u}^{i-1}).$$

The relay transmits $\mathbf{x}_2(\tilde{u}_{\mathcal{S}}^{(j)})$ in block $j + 1$.

Decoding. Let the received sequence in block j be $\mathbf{y}_3^{(j)}$. The receiver first recovers $\hat{u}_{\mathcal{S}}^{(j)} = \hat{l}(\mathbf{y}_3^{(j+1)})$. Then it sets $\hat{u}_{\mathcal{F}}^{(j)} = u_{\mathcal{F}}$ and successively recovers $\hat{u}_{\mathcal{M} \setminus \mathcal{S}}^{(j)}$ as

$$\hat{u}_i = \arg \max_{u \in \mathcal{U}} p_{U_i | \mathbf{Y}_3, \mathbf{X}_2, U^{i-1}}(u | \mathbf{y}_3^{(j)}, \mathbf{x}_2(\hat{u}_{\mathcal{S}}^{(j-1)}), \hat{u}^{i-1}).$$

Analysis of the probability of error. For $j \in [1 : b - 1]$, let

$$\mathcal{E}_1(j) = \{\tilde{U}_{\mathcal{M}}^{(j)} \neq U_{\mathcal{M}}^{(j)}\},$$

$$\mathcal{E}_2(j) = \{\hat{U}_{\mathcal{S}}^{(j)} \neq \tilde{U}_{\mathcal{S}}^{(j)}\},$$

$$\mathcal{E}_3(j) = \{\hat{U}_{\mathcal{M} \setminus \mathcal{S}}^{(j)} \neq U_{\mathcal{M} \setminus \mathcal{S}}^{(j)}\}.$$

Then, the average probability of error is bounded by

$$\begin{aligned} & \mathbb{P}\{\hat{U}_{\mathcal{M}}^{(j)} \neq U_{\mathcal{M}}^{(j)}\} \\ & \leq \mathbb{P}\{\mathcal{E}_1(j-1) \cup \mathcal{E}_2(j-1) \cup \mathcal{E}_2(j) \cup \mathcal{E}_3(j)\} \\ & \leq \mathbb{P}\{\mathcal{E}_1(j-1)\} + \mathbb{P}\{\mathcal{E}_2(j-1)\} + \mathbb{P}\{\mathcal{E}_2(j)\} \\ & \quad + \mathbb{P}\{\mathcal{E}_3(j) \cap \mathcal{E}_1^c(j-1) \cap \mathcal{E}_2^c(j-1) \cap \mathcal{E}_2^c(j)\}. \end{aligned}$$

The first term corresponds to the transmission on channel $p(y_2, x_2 | v)$. By the error bound for universal polarization [63], $\mathbb{P}(\mathcal{E}_1(j-1)) = N_1 O(2^{-N_2^\beta})$. The second and third terms correspond to the transmission on channel $p(y_3 | x_2)$. Since we are using a good polar code for this channel, $\mathbb{P}(\mathcal{E}_2(j-1)) = \mathbb{P}(\mathcal{E}_2(j)) = O(2^{-N^\beta})$. To bound the last term, note that $\mathbf{X}_2^{(j-1)}$ is i.i.d. $\sim p(x_2)$ and independent of $\mathbf{V}^{(j)}$. By a similar argument as

in [62, Appendix A], once \mathbf{x}_2 is recovered correctly and $u_{\mathcal{S}}$ is provided to the receiver of channel, $\mathbb{P}\{\mathcal{E}_3(j) \cap \mathcal{E}_1^c(j-1) \cap \mathcal{E}_2^c(j-1) \cap \mathcal{E}_2^c(j)\} = N_1 o(2^{-N_2^\beta})$. Thus, the total error tends to zero for fixed N_1 and $N_2 \rightarrow \infty$.

Complexity. By the complexity results of regular polar codes [7] and universal polar codes [63], the complexity is bounded as $O(N \log N)$.

6.3 Compress–Forward Relaying

In this section, we develop a polar coding scheme that achieves the compress–forward lower bound (6.2). In the compress–forward coding scheme, the relay helps communication by sending a description of its received sequence to the receiver. Because this description is correlated with the received sequence, Wyner–Ziv coding is used to reduce the rate needed to communicate it to the receiver. The code design requirements are summarized as follows.

1. We need a Wyner–Ziv code to compress the source (Y_2, X_2) when the side information (Y_3, X_2) is available at the decoder. Here, the joint pmf is

$$p(y_2, x_2, y_3) = \sum_{x_1} p(x_1) p(x_2) p(y_2, y_3 | x_1, x_2).$$

2. We need a code to send the compression index of the Wyner–Ziv code over the point-to-point channel

$$p(y_3 | x_2) = \sum_{x_1, y_2} p(x_1) p(y_2, y_3 | x_1, x_2).$$

3. We need a code to send the message m over the point-to-point channel

$$p(y_3, \hat{y}_2, x_2 | x_1) = \sum_{y_2} p(x_2) p(\hat{y}_2 | x_2, y_2) p(y_2, y_3 | x_1, x_2).$$

Polar coding for lossy compression with side information available at the decoder (the Wyner–Ziv problem) has been studied for special cases: (i) doubly symmetric binary source [44] and (ii) uniform source [12]. The problem for arbitrary joint distribution has been implicitly addressed in [61]. Here we restate it explicitly for completeness and describe the compress–forward scheme in Section 6.3.2.

6.3.1 Polarization for the Wyner–Ziv coding

Fix the pmf that attains R_{CF} . Let $(\hat{\mathbf{Y}}_2, \mathbf{Y}_2, \mathbf{Y}_3, \mathbf{X}_2)$ be i.i.d. according to $p(\hat{y}_2, y_2, y_3, x_2)$. Let $\mathbf{U} = \hat{\mathbf{Y}}_2 G_N$, where G_N is the polar transform for alphabet of size $|\hat{\mathcal{Y}}_2|$ (refer to [65, Chapter 3] non-binary polar transform). For $\beta < 1/2$, define

$$\begin{aligned}\mathcal{C} &= \{i: H(U_i | \mathbf{Y}_2, \mathbf{X}_2, U^{i-1}) > 1 - 2^{-N^\beta}\}, \\ \mathcal{D} &= \{i: H(U_i | \mathbf{Y}_3, \mathbf{X}_2, U^{i-1}) < 2^{-N^\beta}\}.\end{aligned}$$

Note that the pmf that attains R_{CF} satisfies the Markov chain $\hat{Y}_2 \rightarrow (Y_2, X_2) \rightarrow (Y_3, X_2)$. Thus, for any $i \in \mathcal{C}$, $H(U_i | \mathbf{Y}_3, \mathbf{X}_2, U^{i-1}) \geq H(U_i | \mathbf{Y}_2, \mathbf{X}_2, U^{i-1}) > 1 - 2^{-N^\beta} > 2^{-N^\beta}$, which implies $\mathcal{C} \subseteq \mathcal{D}^c$. Let $\mathcal{I} = (\mathcal{C} \cup \mathcal{D})^c$. Then, the index set is partitioned as $[1 : N] = \mathcal{C} \uplus \mathcal{I} \uplus \mathcal{D}$. By standard polarization results,

$$\begin{aligned}\lim_{N \rightarrow \infty} \frac{1}{N} |\mathcal{I}| &= H(\hat{Y}_2 | Y_3, X_2) - H(\hat{Y}_2 | Y_2, X_2) \\ &= I(\hat{Y}_2; Y_2 | Y_3, X_2).\end{aligned}$$

6.3.2 Coding Scheme

Again, we use b transmission blocks, each consisting of N transmissions. A sequence of $(b - 1)$ messages $M_j, j \in [1 : b - 1]$ is sent over these b blocks.

Codebooks and rate. We use three codes. The first code is a polar code for lossy compression of $(\mathbf{Y}_2, \mathbf{X}_2)$ when the side information $(\mathbf{Y}_3, \mathbf{X}_2)$ is available at the decoder. We let $\mathbf{U} = \hat{\mathbf{Y}}_2 G_N$ as above. The rate of this code is $\hat{R}_2 = \frac{1}{N} |\mathcal{I}| \leq I(\hat{Y}_2; Y_2 | X_2, Y_3) + \delta_1$ for

some small δ_1 . Let the bits $U_{\mathcal{C}}$ be i.i.d. according to $\text{Unif}[1 : |\mathcal{U}|]$. Let $Z \sim \text{Unif}[1 : 2^{n\hat{R}_2}]$ be independent of \mathbf{U} . $(U_{\mathcal{C}}, Z)$ are shared by all parties. We use two point-to-point polar codes for the channels $p(y_3|x_2)$ and $p(y_3, \hat{y}_2, x_2|x_1)$, respectively, using the method in [39]. We abstract the polar code for $p(y_3|x_2)$ as an encoder $\mathbf{x}_2(l)$ and a decoder $\hat{l}(\mathbf{y}_3)$. The rate of the code is $\frac{1}{N}|\mathcal{I}| \leq I(\hat{Y}_2; Y_2|X_2, Y_3) + \delta_1 \leq I(X_2; Y_3) - \delta_2$, where the second inequality follows from the constraint in the compress-forward lower bound (6.2) with small δ_1 and δ_2 . We abstract the polar code for $p(y_3, \hat{y}_2, x_2|x_1)$ as an encoder $\mathbf{x}_1(m)$ and a decoder $\hat{m}(\mathbf{y}_3, \hat{\mathbf{y}}_2, \mathbf{x}_2)$. The rate of the code is $R = R_{\text{CF}} - \delta_3$. Here, $\delta_1, \delta_2, \delta_3 > 0$ can be made arbitrarily small for large N .

Encoding. To send the message $m_j \in [1 : 2^{nR}]$ in block $j \in [1 : b - 1]$, the sender transmits $\mathbf{x}_1(m_j)$.

Relay encoding. Let $(u_{\mathcal{C}}, z)$ be the observed random variable. Let $\mathbf{y}_2^{(j)}$ be the received sequence in block $j \in [1 : b]$. The relay sets $U_{\mathcal{C}}^{(j)} = u_{\mathcal{C}}$. For $i \in \mathcal{C}^c$, it randomly assigns $U_i^{(j)} = u$ with probability $p_{U_i|\mathbf{Y}_2, \mathbf{X}_2, U^{i-1}}(u|\mathbf{y}_2^{(j)}, \mathbf{x}_2(l_{j-1}), u^{i-1})$ for $u \in \mathcal{U}$. Let $l_j = (u_{\mathcal{I}}^{(j)} + z) \bmod 2^{n\hat{R}_2}$. The relay transmits $\mathbf{x}_2(l_j)$ in block $j + 1$.

Decoding. Let the received sequence in block $j \in [1 : b]$ be $\mathbf{y}_3^{(j)}$. The decoder first recovers $\hat{l}_j = \hat{l}(\mathbf{y}_3^{(j+1)})$ and computes $\hat{u}_{\mathcal{I}}^{(j)} = (\hat{l}_j + z) \bmod 2^{n\hat{R}_2}$. Then, it sets $\hat{u}_{\mathcal{C}}^{(j)} = u_{\mathcal{C}}$ and successively recovers bits in \mathcal{D} as

$$\hat{u}_i^{(j)} = \arg \max_{u \in \mathcal{U}} p_{U_i|\mathbf{Y}_3, \mathbf{X}_2, U^{i-1}}(u|\mathbf{y}_3^{(j)}, \mathbf{x}_2(\hat{l}_{j-1}), \hat{u}^{i-1}).$$

Upon recovering $\hat{\mathbf{y}}_2^{(j)} = \hat{\mathbf{u}}^{(j)}G_N$, the decoder declares $\hat{m}_j = \hat{m}(\mathbf{y}_3^{(j)}, \hat{\mathbf{y}}_2^{(j)}, \mathbf{x}_2(\hat{l}_{j-1}))$ as its message estimate.

Analysis of the probability of error. There is some subtlety in the error analysis, since the actual source sequence $(\mathbf{Y}_2, \mathbf{X}_2)$ at the relay are not necessarily i.i.d. as desired in the Wyner–Ziv problem. We denote the desired distribution as

$$P(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}_2, \mathbf{y}_3, \hat{\mathbf{y}}_2, \mathbf{u}) = \mathbb{1}_{\{\mathbf{u}=\hat{\mathbf{y}}_2G_N\}} \left(\prod_{i=1}^N p(x_{1i})p(x_{2i})p(y_{2i}, y_{3i}|x_{1i}, x_{2i})p(\hat{y}_{2i}|x_{2i}, y_{2i}) \right).$$

Let Q be the actual distribution of the above variables

$$Q(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}_2, \mathbf{y}_3, \hat{\mathbf{y}}_2, \mathbf{u}) = Q(\mathbf{x}_1)Q(\mathbf{x}_2)P(\mathbf{y}_3, \mathbf{y}_2 | \mathbf{x}_1, \mathbf{x}_2) \frac{1}{|\mathcal{U}|^{|\mathcal{C}|}} \\ \cdot \prod_{i \in \mathcal{C}^c} P(u_i | \mathbf{y}_2, \mathbf{x}_2, u^{i-1}) \mathbb{1}_{\{\mathbf{u} = \hat{\mathbf{y}}_2 G_N\}}.$$

Proposition 6.3.1. For $\beta < 1/2$,

$$\|P(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}_2, \mathbf{y}_3, \hat{\mathbf{y}}_2, \mathbf{u}) - Q(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}_2, \mathbf{y}_3, \hat{\mathbf{y}}_2, \mathbf{u})\| = O(2^{-N^\beta}),$$

where $\|P(s) - Q(s)\| := \sum_s |P(s) - Q(s)|$.

Proof. For the polar coding ensembles designed in [39],

$$\|Q(\mathbf{x}_1) - P(\mathbf{x}_1)\| = O(2^{-N^\beta}),$$

$$\|Q(\mathbf{x}_2) - P(\mathbf{x}_2)\| = O(2^{-N^\beta}).$$

We write $\tilde{\mathbf{x}} := (\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}_2, \mathbf{y}_3)$. The above implies

$$\begin{aligned} \|Q(\tilde{\mathbf{x}}) - P(\tilde{\mathbf{x}})\| &= \|Q(\mathbf{x}_1, \mathbf{x}_2) - P(\mathbf{x}_1, \mathbf{x}_2)\| \\ &= \sum_{\mathbf{x}_1, \mathbf{x}_2} |Q(\mathbf{x}_1)[Q(\mathbf{x}_2) - P(\mathbf{x}_2)] + P(\mathbf{x}_2)[Q(\mathbf{x}_1) - P(\mathbf{x}_1)]| \\ &\leq \sum_{\mathbf{x}_2} |Q(\mathbf{x}_2) - P(\mathbf{x}_2)| + \sum_{\mathbf{x}_1} |Q(\mathbf{x}_1) - P(\mathbf{x}_1)| = O(2^{-N^\beta}). \end{aligned}$$

Therefore, we have

$$\begin{aligned} \|P(\tilde{\mathbf{x}}, \mathbf{u}, \hat{\mathbf{y}}_2) - Q(\tilde{\mathbf{x}}, \mathbf{u}, \hat{\mathbf{y}}_2)\| &= \|P(\tilde{\mathbf{x}}, \mathbf{u}) - Q(\tilde{\mathbf{x}}, \mathbf{u})\| \\ &= \sum_{\tilde{\mathbf{x}}, \mathbf{u}} P(u_{\mathcal{C}^c} | \tilde{\mathbf{x}}) \left| P(\tilde{\mathbf{x}}) \prod_{i \in \mathcal{C}} P(u_i | \mathbf{y}_2, \mathbf{x}_2, u^{i-1}) - Q(\tilde{\mathbf{x}}) \frac{1}{|\mathcal{U}|^{|\mathcal{C}|}} \right| \\ &\leq \sum_{\tilde{\mathbf{x}}, \mathbf{u}} P(u_{\mathcal{C}^c} | \tilde{\mathbf{x}}) \left| \prod_{i \in \mathcal{C}} P(u_i | \mathbf{y}_2, \mathbf{x}_2, u^{i-1}) - \frac{1}{|\mathcal{U}|^{|\mathcal{C}|}} \right| \end{aligned}$$

$$\begin{aligned}
& + \sum_{\tilde{\mathbf{x}}, \mathbf{u}} P(u_{\mathcal{C}^c} | \tilde{\mathbf{x}}) \frac{1}{|\mathcal{U}|^{|\mathcal{C}|}} \left| P(\tilde{\mathbf{x}}) - Q(\tilde{\mathbf{x}}) \right| \\
& \stackrel{(a)}{\leq} \sum_{i \in \mathcal{C}} \sqrt{2(\log_{|\mathcal{U}|} e)(1 - H(U_i | \mathbf{Y}_2, \mathbf{X}_2, U^{i-1}))} + O(2^{-N^\beta}) \\
& = O(2^{-N^\beta}),
\end{aligned}$$

where (a) follows from Pinsker's inequality. \square

Given Proposition 6.3.1, to bound the error under true distribution Q , we can instead consider the error under distribution P . For $j \in [1 : b]$, define the error events

$$\begin{aligned}
\mathcal{E}(j) &= \{\hat{M}_j \neq M_j\}, & \mathcal{E}_1(j) &= \{\hat{L}_j \neq L_j\}, \\
\mathcal{E}_2(j) &= \{\hat{U}_{\mathcal{I}} \neq U_{\mathcal{I}}\}, & \mathcal{E}_3(j) &= \{\hat{U}_{\mathcal{D}} \neq U_{\mathcal{D}}\}.
\end{aligned}$$

Then the probability of error (averaged over $(U_{\mathcal{C}}, Z)$) is

$$\begin{aligned}
\mathbb{P}(\mathcal{E}(j)) &\leq \mathbb{P}(\mathcal{E}(j) \cup \mathcal{E}_1(j-1) \cup \mathcal{E}_2(j) \cup \mathcal{E}_3(j)) \\
&\leq \mathbb{P}(\mathcal{E}_2(j)) + \mathbb{P}(\mathcal{E}_1(j-1)) \\
&\quad + \mathbb{P}(\mathcal{E}_3(j) \cap \mathcal{E}_2^c(j) \cap \mathcal{E}_1^c(j-1)) \\
&\quad + \mathbb{P}(\mathcal{E}(j) \cap \mathcal{E}_3^c(j) \cap \mathcal{E}_2^c(j) \cap \mathcal{E}_1^c(j-1)).
\end{aligned}$$

For the first term, the compression index $U_{\mathcal{I}}$ is sent through the point-to-point channel $p(y_3|x_2)$ via randomization $L_j = (U_{\mathcal{I}} + Z) \bmod 2^{n\hat{R}_2}$. Note that although the compression index might not be uniform over $[1 : 2^{n\hat{R}_2}]$, after the randomization, $L_j \sim \text{Unif}[1 : 2^{n\hat{R}_2}]$. Thus, the probability of error (averaged over Z) is $\mathbb{P}(\mathcal{E}_2(j)) = \sum_{l_j} p(l_j) \mathbb{P}\{\hat{U}_{\mathcal{I}} \neq U_{\mathcal{I}} | L_j = l_j\} = \frac{1}{2^{n\hat{R}_2}} \sum_{l_j} \mathbb{P}\{\hat{L}_j \neq L_j | L_j = l_j\} = \mathbb{P}(\mathcal{E}_1(j)) = O(2^{-N^\beta})$, where the last equality follows from the error bound for point-to-point polar codes with uniform message [39]. $\mathbb{P}(\mathcal{E}_1(j-1))$ is bounded for the same reason. The third term is bounded by $\sum_{i \in \mathcal{D}} Z(U_i | \mathbf{Y}_2, \mathbf{X}_2, U^{i-1}) = O(2^{-N^\beta})$. Once $\hat{\mathbf{y}}_2(j)$ and $\mathbf{x}_2(\hat{l}_{j-1})$ are recovered correctly,

the last term is bounded as $O(2^{-N^\beta})$, since a good point-to-point polar codes is used over the channel $p(y_3, \hat{y}_2, x_2|x_1)$. Therefore, there must exist a good (u_c, z) such that the error is bounded as $O(2^{-N^\beta})$.

Complexity. By the standard polar coding result, the complexity of this coding scheme is $O(N \log N)$.

6.4 Discussion

In this chapter, we present polar coding schemes that achieve the decode–forward and compress–forward lower bounds in general three-node relay channels. A technique based on the universal polarization method in [63] is applied to resolve the incompatible polarization issue in general relay channels. One could also apply the “chaining construction” in [36] to solve this problem. However, two levels of block Markov coding might be needed as the basic relaying scheme already involves one.

The decode–forward scheme can also be adapted to achieve the partial decode–forward lower bound for the three-node relay channel. The proof technique in the compress–forward scheme ensures that the existing scheme continues to work for “close to i.i.d.” sources. This technique can be useful for other problems, such as Marton coding for broadcast channels. Unlike the polar coding schemes for interference channels [?], which have straightforward extensions to interference networks, here it is nontrivial to extend to more general schemes for relaying networks, such as noisy network coding [50] or distributed decode–forward [49]. We leave this as an open problem.

Acknowledgment

This chapter is, in part, a reprint of the material in the paper: Lele Wang, “Polar coding for relay channels,” *Proceedings of the IEEE International Symposium on Information Theory*, Hong Kong, China, June 2015.

Chapter 7

Channel Coding and Slepian–Wolf Coding

In this chapter, We study the duality between channel coding and Slepian–Wolf coding in the linear coding framework. We show how a code (both its encoder and decoder) for a *symmetric* channel coding problem can be used to design a code for a general Slepian–Wolf problem. Conversely, we show how a code for a *symmetric* Slepian–Wolf problem can be used to design a code for a general channel coding problem. The exact relations between the rates and the probability of errors of the two codes are established.

7.1 Introduction

7.1.1 Channel Coding Problem

A binary-input memoryless channel (BMC) $p(y|x)$ consists of an input alphabet $\mathcal{X} = \{0, 1\}$, a finite output alphabet \mathcal{Y} , and a collection of conditional probability mass functions $p(y|x)$ on \mathcal{Y} for $x \in \{0, 1\}$. We say a BMC $p(y|x)$ is *symmetric* if there exists a permutation $\pi: \mathcal{Y} \rightarrow \mathcal{Y}$ such that $p(y|x) = p(\pi(y)|x \oplus 1)$ for all $y \in \mathcal{Y}$ and $x \in \{0, 1\}$.

A (k, n, ϵ) code (f, ϕ) for the BMC $p(y|x)$ consists of

- a codebook $\mathcal{C} \subseteq \{0, 1\}^n$ of size $|\mathcal{C}| = 2^k$,
- an encoder $f: \mathcal{C} \rightarrow \{0, 1\}^n$ that maps each codeword c^n to a channel input $x^n = f(c^n)$, and
- a decoder $\phi: \mathcal{Y}^n \rightarrow \mathcal{C}$ that assigns a codeword estimate $\hat{c}^n = \phi(y^n)$ to each received sequence y^n .

We assume that C^n is uniform over the codebook \mathcal{C} . The rate of the code is $R_{\text{ch}} = k/n$.

The average probability of error of the code is $\mathbb{P}\{\hat{C}^n \neq C^n\} = \epsilon$.

We say a channel code is *linear* if the codebook \mathcal{C} is such that for any two codewords $c^n, \tilde{c}^n \in \mathcal{C}$, $c^n \oplus \tilde{c}^n \in \mathcal{C}$. Equivalently, a linear code can be defined by its parity check matrix $H_{(n-k) \times n}$. For notational convenience, we introduce the augmented parity check matrix $\bar{H}_{n \times n} = \begin{bmatrix} 0 \\ H \end{bmatrix}$ so that all vectors in this paper are of length n . Thus, the codebook of a linear code can be written as $\mathcal{C} = \{c^n: \bar{H}c^n = 0^n\}$. When a (k, n, ϵ) code (f, ϕ) is linear with associated augmented parity check matrix \bar{H} and $f(c^n) = c^n$, we say it is a (k, n, ϵ) linear code (\bar{H}, ϕ) .

7.1.2 Slepian–Wolf Problem

A Slepian–Wolf problem $p(x, y)$ consists of two finite alphabets $\mathcal{X} = \{0, 1\}$, \mathcal{Y} , and a joint pmf $p(x, y)$ over $\{0, 1\} \times \mathcal{Y}$. The binary memoryless source X with side information Y generates a jointly i.i.d. random process $\{(X_i, Y_i)\}$ with $(X_i, Y_i) \sim p_{X,Y}(x_i, y_i)$. We say a Slepian–Wolf problem $p(x, y)$ is *symmetric* if $X \sim \text{Bern}(1/2)$ and the channel $p(y|x)$ is symmetric.

An (l, n, ϵ) code (g, ψ) for the Slepian–Wolf problem $p(x, y)$ consists of

- an index set $\mathcal{I} \subseteq \{0, 1\}^n$ of size $|\mathcal{I}| = 2^l$,
- an encoder $g: \{0, 1\}^n \rightarrow \mathcal{I}$ that maps each source sequence x^n to an index $s^n = g(x^n)$, and

- a decoder $\psi: \mathcal{I} \times \mathcal{Y}^n \rightarrow \{0, 1\}^n$ that assigns a source estimate $\hat{x}^n = \psi(s^n, y^n)$ to each index s^n and side information sequence y^n .

The rate of the code is $R_{\text{sw}} = (n - k)/n$. The average probability of error of the code is $\mathbb{P}\{\hat{X}^n \neq X^n\} = \epsilon$.

We say a Slepian–Wolf code is *linear* if for any $x^n, \tilde{x}^n \in \{0, 1\}^n$, $g(x^n) \oplus g(\tilde{x}^n) = g(x^n \oplus \tilde{x}^n)$. When an (l, n, ϵ) Slepian–Wolf code (g, ψ) is linear with an encoder defined by matrix multiplication $g(x^n) = \bar{H}x^n$, where $\bar{H}_{n \times n} = \begin{bmatrix} 0 \\ H_{l \times n} \end{bmatrix}$, we say it is an (l, n, ϵ) linear code (\bar{H}, ψ) .

7.1.3 Background

The connection between the channel coding problem and the Slepian–Wolf problem has long been observed in the literature. In [81], Wyner showed that a linear (k, n, ϵ) code for the binary symmetric channel with crossover probability p (BSC(p)) can be used to construct a linear $(n - k, n, \epsilon)$ code for the symmetric Slepian–Wolf problem $p(x, y)$ where $p(y|x)$ is a BSC(p). Since then, several attempts have been made to generalize this observation [2, 15, 16, 21, 30, 44, 55, 59, 68, 70]. In [16], Chen, He, Jagmohan, Lastras-Montano, and Yang related a general Slepian–Wolf problem $p(x, y)$ to a dual channel coding problem $p_{V|U}(v|u)$, where $V = (U \oplus X, Y)$ and $(X, Y) \sim p(x, y)$ is independent of U . Under the maximum *a posteriori* decoding, a linear (k, n, ϵ) code for the dual channel $p(v|u)$ can be used to design a linear $(n - k, n, \epsilon)$ code for the Slepian–Wolf problem $p(x, y)$. Miyake [55] studied this duality for sparse matrix codes with minimum-entropy decoding. Such duality were also established for some low-complexity codes, such as LDPC codes with density evolution decoding [15] and polar codes with successive cancellation decoding [44]. In all of these results except [81], the duality was established only for the encoder, i.e., the encoder of one code is treated as a black box in designing another code. However, one has to specify the decoding rule to analyze the probability of error.

7.1.4 Contributions

In this paper, we investigate whether the duality result can be established for a given encoder and decoder pair. Given a linear (k, n, ϵ) symmetric channel code (\bar{H}, ϕ) , how can we construct a general Slepian–Wolf code and what can we say about its performance (in terms of rate and probability of error)? Conversely, given a linear (l, n, ϵ) symmetric Slepian–Wolf code (\bar{H}, ψ) , how can we construct a general channel code and what can we say about its performance? The motivation is to translate the performance of commercial off-the-shelf codes that are well studied and simulated in one communication scenario into the performance of codes for another communication scenario. From the theoretical point of view, such a linear code duality will generalize most existing results and will unify the analysis.

The main results of this paper are summarized in Figure 7.1. We first show how to construct a linear $(n - k, n, \epsilon)$ symmetric Slepian–Wolf code from a linear (k, n, ϵ) symmetric channel code in Section 7.2.1 and a general $(n - k, n, \epsilon)$ Slepian–Wolf code from a linear $(n - k, n, \epsilon)$ symmetric Slepian–Wolf code in Section 7.2.2. Next we show how to construct a (k, n, ϵ) symmetric channel code from a linear $(n - k, n, \epsilon)$ symmetric Slepian–Wolf code in Section 7.2.3 and a (k, n, ϵ) general channel code from a linear (k, n, ϵ) symmetric channel code in Section 7.2.4. By combining all four results, we establish the duality between the general Slepian–Wolf problem and the general channel coding problem.

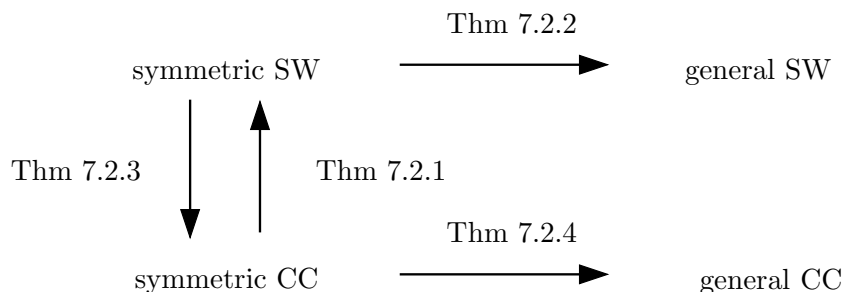


Figure 7.1. A summary of the main results. SW is short for Slepian–Wolf coding and CC is short for channel coding.

7.2 Linear Code Duality

7.2.1 Symmetric Channel Code to Symmetric Slepian–Wolf Code

Suppose that for the symmetric BMC $p(y|x)$ with permutation π , there is a linear (k, n, ϵ) code (\bar{H}, ϕ) . Without loss of generality, assume that the augmented parity check matrix is systematic

$$\bar{H} = \begin{bmatrix} 0 & 0 \\ A & I_{n-k} \end{bmatrix},$$

where A is an $(n-k) \times k$ matrix and I_{n-k} is the $(n-k) \times (n-k)$ identity matrix. The block diagram for this problem is shown in Figure 7.2. We have the average probability of error is given by

$$\mathbb{P}\{\phi(\tilde{R}^n) \neq \tilde{C}^n\} = \epsilon.$$

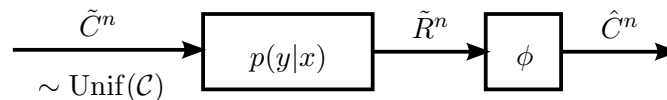


Figure 7.2. A channel code for symmetric BMC $p(y|x)$.

To construct a code for the symmetric Slepian–Wolf problem $p(x, y)$ from the above channel code, we first introduce two building blocks.

The first block, termed *codify*, takes two inputs, a binary sequence x^n and the syndrome $\bar{H}x^n$ of it, and outputs the element-wise modulo-two sum of the two inputs $x^n \oplus \bar{H}x^n$, as depicts in the left part of Figure 7.3. Intuitively, this operation shifts any binary sequence x^n to a codeword, as illustrated in the right part of Figure 7.3. We prove this in Lemma 7.2.1.

Lemma 7.2.1. *For any $x^n \in \{0, 1\}^n$, $x^n \oplus \bar{H}x^n \in \mathcal{C}$.*

Proof. For any $x^n \in \{0, 1\}^n$, we have

$$\begin{aligned}
 \bar{H}(x^n \oplus \bar{H}x^n) &= \bar{H}x^n \oplus \begin{bmatrix} 0 & 0 \\ A & I \end{bmatrix} \bar{H}x^n \\
 &\stackrel{(a)}{=} \begin{bmatrix} 0 \\ Hx^n \end{bmatrix} \oplus \begin{bmatrix} 0 & 0 \\ A & I \end{bmatrix} \begin{bmatrix} 0 \\ Hx^n \end{bmatrix} \\
 &= \begin{bmatrix} 0 \\ Hx^n \end{bmatrix} \oplus \begin{bmatrix} 0 \\ Hx^n \end{bmatrix} \\
 &= 0,
 \end{aligned}$$

where $H = [A, I]$ in (a). Therefore, $x^n \oplus \bar{H}x^n \in \mathcal{C}$. □

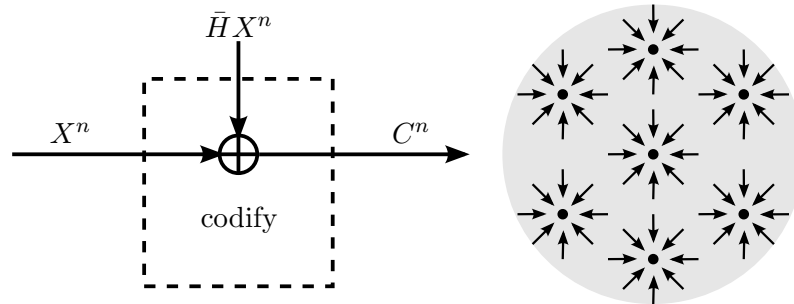


Figure 7.3. The codify block. Left: The block diagram. Right: Illustration of a shift by $\bar{H}X^n$ in $\{0, 1\}^n$ space.

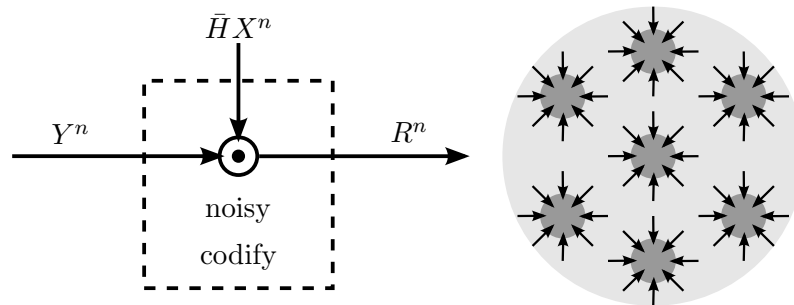


Figure 7.4. The noisy codify block. Left: The block diagram. Right: Illustration of a shift by $\bar{H}X^n$ in \mathcal{Y}^n space.

The second block, termed *noisy codify*, takes two inputs, the noisy observation y^n of the binary sequence x^n and the syndrome $\bar{H}x^n$, and outputs $y^n \oplus \bar{H}x^n$, which is

defined as follows. For $y \in \mathcal{Y}$ and $s \in \{0, 1\}$,

$$y \odot s = \begin{cases} y & \text{if } s = 0 \\ \pi(y) & \text{if } s = 1. \end{cases}$$

Let $y^n \odot s^n$ be the element-wise \odot operation. The left part of Figure 7.4 depicts the block diagram. Similar to the shift in the codify operation, this block takes a *corresponding shift* in \mathcal{Y}^n space and outputs a *noisy version* of the output sequence in the codify block, as illustrated in the right part of Figure 7.4. Lemma 7.2.2 makes this statement rigorous.

Lemma 7.2.2. *Let (X^n, Y^n) be i.i.d. according to $p(x, y)$, where $p(x, y)$ is symmetric under permutation π . Let $C^n = X^n \oplus \bar{H}X^n$ and $R^n = Y^n \odot \bar{H}X^n$. Then,*

$$P\{C^n = c^n, R^n = r^n\} = \frac{1}{2^k} \prod_{i=1}^n p_{Y|X}(r_i | c_i)$$

for every $c^n \in \mathcal{C}$ and $r^n \in \mathcal{Y}^n$.

Proof. Define $\mathcal{S} = \{s^n \in \{0, 1\}^n : s^k = 0^k\}$. For any $c^n \in \mathcal{C}$,

$$\begin{aligned} & P\{C^n = c^n\} \\ &= \sum_{s^n \in \mathcal{S}} P\{X^n \oplus \bar{H}X^n = c^n, \bar{H}X^n = s^n\} \\ &= \sum_{s^n \in \mathcal{S}} P\{X^n = c^n \oplus s^n\} P\{\bar{H}X^n = s^n | X^n = c^n \oplus s^n\} \\ &\stackrel{(b)}{=} \sum_{s^n \in \mathcal{S}} \frac{1}{2^n} \\ &= \frac{1}{2^k}, \end{aligned}$$

where (b) follows since X^n is i.i.d. Bern(1/2) and for any $c^n \in \mathcal{C}$ and $s^n \in \mathcal{S}$, $\bar{H}(c^n \oplus s^n) =$

$$0^n \oplus \begin{bmatrix} 0 & 0 \\ A & I \end{bmatrix} \begin{bmatrix} 0^k \\ s_{k+1}^n \end{bmatrix} = s^n. \text{ Now, for any } r^n \in \mathcal{Y}^n, \text{ consider}$$

$$\begin{aligned} & \mathbb{P}\{R^n = r^n | C^n = c^n\} \\ &= \sum_{s^n \in \mathcal{S}} \mathbb{P}\{\bar{H}X^n = s^n, Y^n \odot \bar{H}X^n = r^n | C^n = c^n\} \\ &= \sum_{s^n \in \mathcal{S}} \mathbb{P}\{\bar{H}X^n = s^n | C^n = c^n\} \\ &\quad \cdot \mathbb{P}\{Y^n \odot s^n = r^n | \bar{H}X^n = s^n, X^n = c^n \oplus s^n\} \\ &= \sum_{s^n \in \mathcal{S}} \mathbb{P}\{\bar{H}X^n = s^n | C^n = c^n\} \\ &\quad \cdot \mathbb{P}\{Y^n = r^n \odot s^n | X^n = c^n \oplus s^n\} \\ &= \sum_{s^n \in \mathcal{S}} \mathbb{P}\{\bar{H}X^n = s^n | C^n = c^n\} \prod_{i=1}^n p_{Y|X}(r_i \odot s_i | c_i \oplus s_i) \\ &\stackrel{(c)}{=} \sum_{s^n \in \mathcal{S}} \mathbb{P}\{\bar{H}X^n = s^n | C^n = c^n\} \prod_{i=1}^n p_{Y|X}(r_i | c_i) \\ &= \prod_{i=1}^n p_{Y|X}(r_i | c_i), \end{aligned}$$

where (c) follows from the symmetry of the channel $p(y|x)$. \square

Lemma 7.2.2 implies that if Y^n is the output of the channel $p(y|x)$ when the channel input is X^n . Then, the output the noisy codify block, R^n , distributes as if it is the output of the same channel $p(y|x)$ when the channel input is the codified sequence C^n . This is illustrated in Figure 7.5. To recover (the top right) X^n from (the bottom left) Y^n , one can go through the path $Y^n \rightarrow R^n \rightarrow C^n \rightarrow X^n$. To get an estimate \hat{C}^n from R^n , one can apply the decoder of the channel code. This explains the three steps—noisy codify, channel decoder, uncodify—in the Slepian–Wolf decoder in Figure 7.6. Moreover, since the noisy codify and uncodify blocks are invertible, the essential error in recovering X^n from Y^n is the same as the error in recovering C^n from R^n .

Now we are ready to construct a code for the symmetric Slepian–Wolf problem

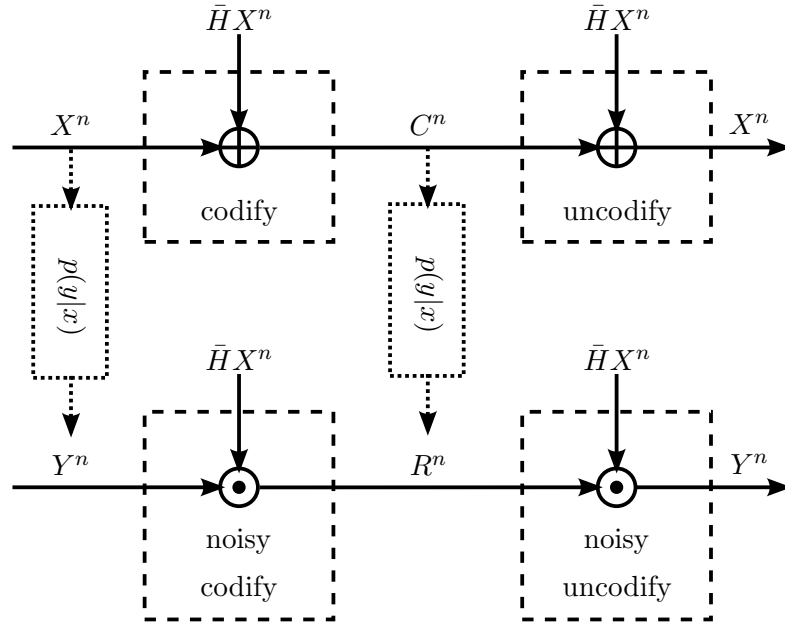


Figure 7.5. Relations of the random variables (X^n, Y^n, C^n, R^n) .

$p(x, y)$. Figure 7.6 illustrates the block diagram.

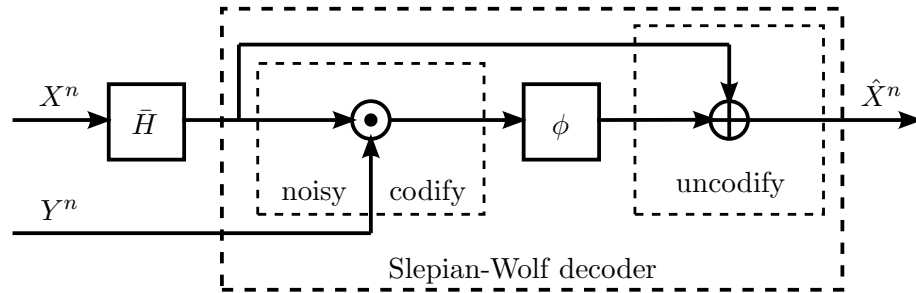


Figure 7.6. The construction of a symmetric Slepian-Wolf code from a symmetric channel code.

Encoding. Upon observing the source sequence x^n , the sender transmits $s^n = \bar{H}x^n$.

Decoding. Upon observing y^n sequence and receiving the index s^n , the decoder declares

$$\hat{x}^n = \phi(y^n \odot s^n) \oplus s^n$$

as the source estimate.

Analysis of the probability of error. We have

$$\begin{aligned}
& \mathbb{P}\{\hat{X}^n \neq X^n\} \\
&= \mathbb{P}\{\phi(Y^n \odot \bar{H}X^n) \oplus \bar{H}X^n \neq X^n\} \\
&= \mathbb{P}\{\phi(Y^n \odot \bar{H}X^n) \neq X^n \oplus \bar{H}X^n\} \\
&= \mathbb{P}\{\phi(R^n) \neq C^n\} \\
&\stackrel{(d)}{=} \mathbb{P}\{\phi(\tilde{R}^n) \neq \tilde{C}^n\} \\
&= \epsilon,
\end{aligned}$$

where (d) follows from Lemma 7.2.2.

This code construction leads to the following conclusion.

Theorem 7.2.1. *From each linear (k, n, ϵ) code for the symmetric BMC $p(y|x)$, one can construct a linear $(n - k, n, \epsilon)$ code for the symmetric Slepian–Wolf problem $p(x, y)$.*

Remark 7.2.1. *By construction, the rate of the Slepian–Wolf code is $R_{sw} = (n - k)/n = 1 - R_{ch}$.*

7.2.2 Symmetric Slepian–Wolf Code to General Slepian–Wolf Code

Now we consider the general Slepian–Wolf problem $p(x, y)$. We show that by introducing common randomness, every general Slepian–Wolf problem can be *symmetrized* by scrambling.

Lemma 7.2.3. *Let $Z \sim \text{Bern}(1/2)$ be independent of (X, Y) . Let $\tilde{X} = X \oplus Z$ and $\tilde{Y} = (Y, Z)$. Then, the Slepian–Wolf problem $p(\tilde{x}, \tilde{y})$ is symmetric.*

Proof. First, we note that $\tilde{X} \sim \text{Bern}(1/2)$. Moreover, for every $x, z \in \{0, 1\}$ and $y \in \mathcal{Y}$, we have

$$p_{\tilde{Y}|\tilde{X}}(y, z|x \oplus z) = \frac{p_{Y,Z,\tilde{X}}(y, z, x \oplus z)}{p_{\tilde{X}}(x \oplus z)}$$

$$\begin{aligned}
& \stackrel{(a)}{=} \frac{p_{Y,Z,X}(y, z, x)}{p_{\tilde{X}}(x \oplus z \oplus 1)} \\
& \stackrel{(b)}{=} \frac{p_{Y,Z,X}(y, z \oplus 1, x)}{p_{\tilde{X}}(x \oplus z \oplus 1)} \\
& = \frac{p_{Y,Z,\tilde{X}}(y, z \oplus 1, x \oplus z \oplus 1)}{p_{\tilde{X}}(x \oplus z \oplus 1)} \\
& = p_{\tilde{Y}|\tilde{X}}(y, z \oplus 1 | x \oplus z \oplus 1),
\end{aligned}$$

where (a) follows since $\tilde{X} \sim \text{Bern}(1/2)$ and (b) follows since $Z \sim \text{Bern}(1/2)$ is independent of (X, Y) . Thus, the Slepian–Wolf problem $p(\tilde{x}, \tilde{y})$ is symmetric under permutation

$$\pi(\tilde{y}) = \pi(y, z) = (y, z \oplus 1).$$

□

In order to design a code for the general Slepian–Wolf problem (X, Y) , we utilize a linear $(n - k, n, \epsilon)$ code (\bar{H}, ψ) for the symmetrized Slepian–Wolf problem $(\tilde{X}, \tilde{Y}) = (X \oplus Z, (Y, Z))$, where $Z \sim \text{Bern}(1/2)$ is independent of (X, Y) . The block diagram of this code is shown in Figure 7.7. The average probability of error of this code is

$$\mathbb{P}\{\psi(\bar{H}\tilde{X}^n, \tilde{Y}^n) \neq \tilde{X}^n\} = \epsilon.$$

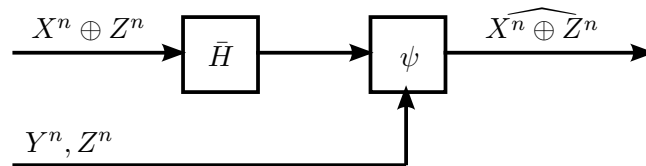


Figure 7.7. A code for the symmetrized Slepian–Wolf problem $p(\tilde{x}, \tilde{y})$.

To construct a code for the general Slepian–Wolf problem $p(x, y)$, we share between the encoder and the decoder a common random sequence Z^n , which is i.i.d. $\text{Bern}(1/2)$ and independent of (X^n, Y^n) , as shown in Figure 7.8.

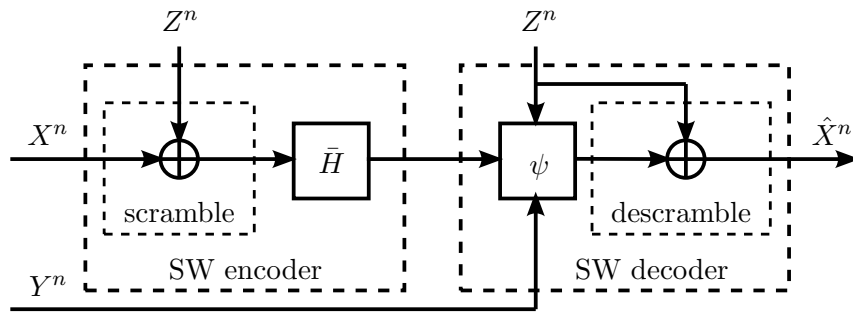


Figure 7.8. The construction of a general Slepian–Wolf code from a symmetric Slepian–Wolf code.

Encoding. Upon observing x^n and z^n , the sender transmits

$$s^n = \bar{H}(x^n \oplus z^n).$$

Decoding. Upon receiving s^n and y^n , the decoder declares

$$\hat{x}^n = \psi(s^n, (y^n, z^n)) \oplus z^n$$

as the source estimate.

Analysis of probability of error. The probability of error averaged over Z^n is

$$\begin{aligned} & \mathbb{P}\{\hat{X}^n \neq X^n\} \\ &= \mathbb{P}\{\psi(\bar{H}(X^n \oplus Z^n), (Y^n, Z^n)) \oplus Z^n \neq X^n\} \\ &= \mathbb{P}\{\psi(\bar{H}\tilde{X}^n, \tilde{Y}^n) \neq \tilde{X}^n\} \\ &= \epsilon. \end{aligned}$$

This code construction leads to the following conclusion.

Theorem 7.2.2. *From each linear $(n - k, n, \epsilon)$ code for the symmetric Slepian–Wolf problem $p(\tilde{x}, \tilde{y})$ as defined above, one can construct an $(n - k, n, \epsilon)$ code for the general Slepian–Wolf problem $p(x, y)$.*

Remark 7.2.2. *By construction, the rate of the general Slepian–Wolf code equals that of the associated symmetric Slepian–Wolf code, $R_{gsw} = l/n = R_{sw}$. Moreover, we note that $H(\tilde{X}|\tilde{Y}) = H(X \oplus Z|Y, Z) = H(X|Y, Z) = H(X|Y)$.*

In the next two sections 7.2.3 and 7.2.4, we show how to construct a general channel code from a symmetric Slepian–Wolf code. Again, we take two steps. We construct first a symmetric channel code and then a general channel code.

7.2.3 Symmetric Slepian–Wolf Code to Symmetric Channel Code

Suppose that for the symmetric Slepian–Wolf problem $p(x, y)$, there is a linear $(n - k, n, \epsilon)$ code (\bar{H}, ψ) , as shown in Figure 7.9. Let $(\tilde{X}^n, \tilde{Y}^n)$ be i.i.d. according to $p_{X,Y}(\tilde{x}, \tilde{y})$. The average probability of error is

$$\mathbb{P}\{\psi(\bar{H}\tilde{X}^n, \tilde{Y}^n) \neq \tilde{X}^n\} = \epsilon.$$

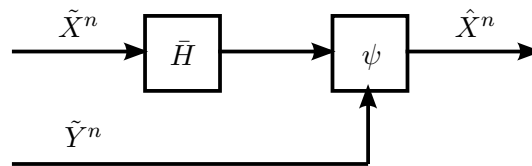


Figure 7.9. A code for symmetric Slepian–Wolf problem $p(x, y)$.

To construct a channel code for the symmetric BMC $p(y|x)$, we share a common random sequence Z^n , which is i.i.d. Bern(1/2) and independent of the message C^n , between the encoder and the decoder. Figure 7.10 illustrates the block diagram.

Encoding. To send $c^n \in \mathcal{C}$, the sender transmits

$$x^n = c^n \oplus z^n.$$

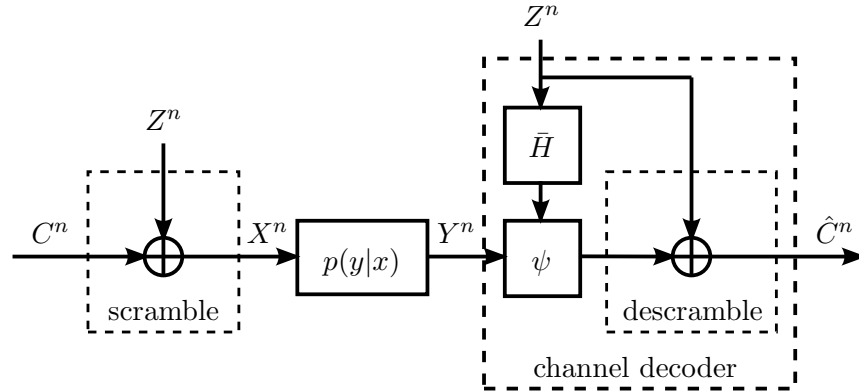


Figure 7.10. The construction of a symmetric channel coding from a symmetric Slepian–Wolf code.

Decoding. Upon receiving y^n , the decoder declares

$$\hat{c}^n = \psi(\bar{H}z^n, y^n) \oplus z^n$$

as the codeword estimate.

Analysis of probability of error. The probability of error averaged over Z^n is bounded as

$$\begin{aligned} \mathbb{P}\{\hat{C}^n \neq C^n\} &= \mathbb{P}\{\psi(\bar{H}Z^n, Y^n) \neq C^n \oplus Z^n\} \\ &\stackrel{(a)}{=} \mathbb{P}\{\psi(\bar{H}X^n, Y^n) \neq X^n\} \\ &\stackrel{(b)}{=} \mathbb{P}\{\psi(\bar{H}\tilde{X}^n, \tilde{Y}^n) \neq \tilde{X}^n\} \\ &= \epsilon, \end{aligned}$$

where (a) follows since $C^n \in \mathcal{C}$ and thus $\bar{H}X^n = \bar{H}C^n \oplus \bar{H}Z^n = \bar{H}Z^n$ and (b) since after scrambling with i.i.d. uniform Z^n sequence, (X^n, Y^n) are identically distributed as $(\tilde{X}^n, \tilde{Y}^n)$ in the Slepian–Wolf problem. Finally, since the probability of error averaged over Z^n is ϵ , there exists a deterministic z^n sequence such that the probability of error is bounded by ϵ .

This code construction leads to the following conclusion.

Theorem 7.2.3. *From each linear $(n - k, n, \epsilon)$ code for the symmetric Slepian–Wolf problem $p(x, y)$, one can construct a (k, n, ϵ) code for the symmetric BMC $p(y|x)$.*

Remark 7.2.3. *By construction, the rate of the channel code $R_{ch} = k/n = 1 - R_{sw}$.*

Remark 7.2.4. *Throughout the construction, we never use the symmetry of the channel $p(y|x)$. Therefore, the same construction works for designing general channel coding from general Slepian–Wolf codes. Due to the uniform dithering, the channel input X is uniform. Thus, the resulting channel code can only achieve up to the symmetric capacity $C_{sym} := I(\text{Bern}(1/2), p(y|x))$ of the BMC $p(y|x)$.*

7.2.4 Symmetric Channel Code to General Channel Code

Now we consider the general channel coding problem $p(y|x)$. Similar to the construction from a symmetric Slepian–Wolf code to a general Slepian–Wolf code, the key technique here is to symmetrize a general channel by scrambling.

Lemma 7.2.4. *Let $\tilde{Z} \sim \text{Bern}(1/2)$ be independent of (X, Y) . Then, the channel*

$$p_{\tilde{Y}, \tilde{Z}|X}(\tilde{y}, \tilde{z}|\tilde{x}) := \frac{1}{2}p_{Y|X}(\tilde{y}|\tilde{x} \oplus \tilde{z})$$

is symmetric.

Proof. The channel $p_{\tilde{Y}, \tilde{Z}|X}(\tilde{y}, \tilde{z}|\tilde{x})$ is symmetric under permutation $\pi(\tilde{y}, \tilde{z}) = (\tilde{y}, \tilde{z} \oplus 1)$ since

$$\begin{aligned} p_{\tilde{Y}, \tilde{Z}|X}(\tilde{y}, \tilde{z}|\tilde{x}) &= \frac{1}{2}p_{Y|X}(\tilde{y}|\tilde{x} \oplus \tilde{z}) \\ &= \frac{1}{2}p_{Y|X}(\tilde{y}|\tilde{x} \oplus 1 \oplus \tilde{z} \oplus 1) \\ &= p_{\tilde{Y}, \tilde{Z}|X}(\tilde{y}, \tilde{z} \oplus 1|\tilde{x} \oplus 1) \end{aligned}$$

for any $\tilde{x}, \tilde{z} \in \{0, 1\}$ and $\tilde{y} \in \mathcal{Y}$. □

Suppose that we have a linear (k, n, ϵ) code (\bar{H}, ϕ) for the symmetrized channel $p(\tilde{y}, \tilde{z}|\tilde{x})$ as illustrated in Figure 7.11. The average probability of error satisfies

$$\mathbb{P}\{\tilde{C}^n \neq \phi(\tilde{Y}^n, \tilde{Z}^n)\} = \epsilon.$$

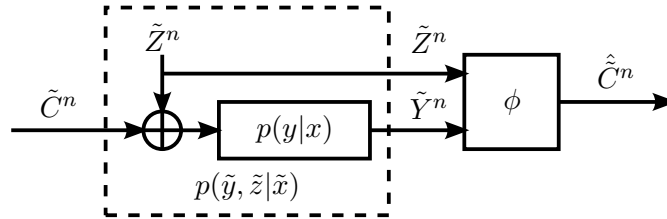


Figure 7.11. Channel coding for symmetric BMC $p(\tilde{y}, \tilde{z}|\tilde{x})$.

To construct a code for the general channel $p(y|x)$, we share between the encoder and the decoder an i.i.d. Bern(1/2) sequence Z^n . The encoding and decoding diagram is shown in Figure 7.12.

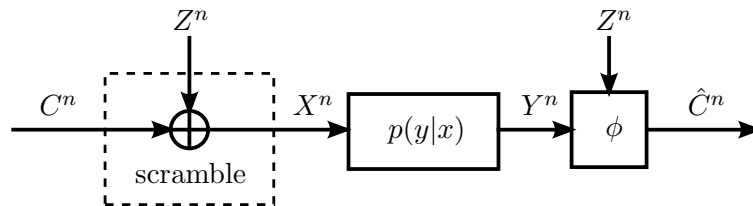


Figure 7.12. The construction of a general channel coding from a symmetric channel code.

Encoding. To send $c^n \in \mathcal{C}$, the sender transmits

$$x^n = c^n \oplus z^n.$$

Decoding. Upon receiving y^n , the decoder declares

$$\hat{c}^n = \phi(y^n, z^n)$$

as the message estimate.

Probability of error analysis. By construction, for all $c^n \in \mathcal{C}$, $y^n \in \mathcal{Y}^n$, and $z^n \in \{0, 1\}^n$, we have

$$\begin{aligned}
& \mathbb{P}\{\tilde{Y}^n = y^n, \tilde{Z}^n = z^n | \tilde{C}^n = c^n\} \\
&= \frac{1}{2^n} \prod_{i=1}^n p_{Y|X}(y_i | c_i \oplus z_i) \\
&= \mathbb{P}\{Z^n = z^n\} \mathbb{P}\{Y^n = y^n | X^n = c^n \oplus z^n\} \\
&\stackrel{(a)}{=} \mathbb{P}\{Z^n = z^n | C^n = c^n\} \\
&\quad \cdot \mathbb{P}\{Y^n = y^n | X^n = c^n \oplus z^n, C^n = c^n\} \\
&= \mathbb{P}\{Z^n = z^n | C^n = c^n\} \\
&\quad \cdot \mathbb{P}\{Y^n = y^n | Z^n = z^n, C^n = c^n\} \\
&= \mathbb{P}\{Y^n = y^n, Z^n = z^n | C^n = c^n\},
\end{aligned}$$

where (a) follows since Z^n is independent of C^n and $C^n \rightarrow X^n \rightarrow Y^n$ form a Markov chain. Therefore, the triples (C^n, Y^n, Z^n) and $(\tilde{C}^n, \tilde{Y}^n, \tilde{Z}^n)$ are identically distributed and the probability of error is

$$\mathbb{P}\{C^n \neq \phi(Y^n, Z^n)\} = \mathbb{P}\{\tilde{C}^n \neq \phi(\tilde{Y}^n, \tilde{Z}^n)\} = \epsilon.$$

This code construction leads to the following conclusion.

Theorem 7.2.4. *From each linear (k, n, ϵ) code for the symmetric BMC $p(\tilde{y}, \tilde{z} | \tilde{x})$ as defined above, one can construct a (k, n, ϵ) code for the general BMC $p(y|x)$.*

Remark 7.2.5. *By construction, the rate of the general channel code is $R_{gch} = k/n = R_{ch}$. We note that*

$$\begin{aligned}
I(\tilde{X}; \tilde{Y}, \tilde{Z}) &= I(\tilde{X}; \tilde{Y} | \tilde{Z}) \\
&= H(\tilde{Y} | \tilde{Z}) - H(\tilde{Y} | \tilde{X}, \tilde{Z})
\end{aligned}$$

$$\begin{aligned}
&\stackrel{(a)}{=} H(\tilde{Y}) - H(\tilde{Y}|\tilde{X} \oplus \tilde{Z}) \\
&= I(\tilde{X} \oplus \tilde{Z}; \tilde{Y}) \\
&= I(\text{Bern}(1/2), p_{Y|X}),
\end{aligned}$$

where (a) follows since $\tilde{X} \sim \text{Bern}(1/2)$ is independent of \tilde{Z} and thus $(\tilde{Y}, \tilde{X} \oplus \tilde{Z})$ is independent of \tilde{Z} . Therefore, we can construct a code for general channel $p_{Y|X}(y|x)$ only up to the symmetric capacity $I(\text{Bern}(1/2), p_{Y|X})$.

Acknowledgment

This chapter is, in part, a reprint of the material in the paper: Lele Wang and Young-Han Kim, “Linear code duality between channel coding and Slepian–Wolf coding,” *Proceedings of the 53th Annual Allerton Conference on Communication, Control, and Computation*, Monticello, Illinois, September 2015.

Chapter 8

Concluding Remarks

We conclude this dissertation with comments for future research directions.

In Chapters 2, 3, 5, and 6, we developed polar coding schemes for various communication scenarios including compound channels, interference channels, broadcast channels, and relay channels. In order to develop a general framework towards the polar Shannon theory, there are several important building blocks to investigate: (i) *Short* length polar coding with universality, as universality is only achieved at very large block-length in the existing schemes; (ii) *Single block* polar coding for Gelfand–Pinsker coding, which has many applications such as joint channel–source coding [54], relaying via hybrid coding [42], and coding for write once memories [79]; (iii) Polar coding with universality for multiple access channels; and (iv) Polar codes construction algorithms for Arikan’s polar splitting scheme [5].

In Chapter 4, we proposed the sliding-window superposition coding scheme, which allowed us to leverage commercial off-the-shelf codes, such as LDPC or turbo codes, to achieve the performance of simultaneous decoding with minimal changes to the encoder and the decoder in existing systems. It would be interesting to explore further whether this scheme can account for the practical constraints in 5G cellular systems. In particular, the finite-blocklength performance, the error propagation effect, the sensitiv-

ity to channel estimation, and the amount of sender-coordination needed are important directions to investigate.

Looking ahead, it would be interesting to consider another crucial technique for wireless systems: an implementation of the capacity-achieving coding scheme for MIMO broadcast channels by Marton. Unlike the current broadcasting techniques that encode different messages for different receivers separately, the capacity-achieving coding scheme by Marton involves a nontrivial joint encoding that “entangles” multiple codewords. In addition to achieving higher rates for broadcast, the low-complexity Marton coding technique can also bring the performance of interference management to the next level, because an interference channel can be transformed into a MIMO broadcast channel via sender-cooperation in the wireless network.

As a long term goal, I wish to better understand the interplay between information theory and coding theory. From a traditional perspective, the random coding scheme, as a standard proof technique in information theory, provides guidance for practical code design. However, in the past decade, several research results, such as compute-and-forward relaying [57], interference alignment [13], distributed compression with coset codes [46], and polar coding for interference channels (Chapter 3), demonstrated that the structured codes can bring strict rate improvement upon state-of-the-art random coding schemes. These results inspired me to fully explore the potential of structured codes and to develop a comprehensive theory of structured coding schemes as well as their performance analysis. Unlike the random coding schemes, which are equipped with ample easy-to-use analysis tools like the packing lemma and the covering lemma, existing analysis tools for structured codes are mostly case-by-case, i.e., depending on the specific communication scenarios, the particular channel statistics, and the underlying code used. Therefore, I plan to develop a set of systematic tools for the performance analysis of structured coding schemes. Chapter 7 took a first step towards this direction.

Appendix A

Polarization Preserves Less Noisy Ordering

Recall that designing a polar code of length 2^n for a channel W consists in finding a set of good channels among $W^{\mathbf{s}}$, $\mathbf{s} \in \{-, +\}^n$, which are defined recursively through

$$W^-(y_1^2|u_1) = \sum_{u_2 \in \{0,1\}} \frac{1}{2} W(y_1|u_1 + u_2) W(y_2|u_2),$$
$$W^+(y_1^2, u_1|u_2) = \frac{1}{2} W(y_1|u_1 + u_2) W(y_2|u_2).$$

A good code of rate $R < I(W)$ can be obtained by picking an R fraction of these channels whose symmetric-capacities $I(W^{\mathbf{s}})$ are largest. Here, we show that a polar code designed in this manner for a channel is also good for all *less noisy* versions of this channel under SC decoding. This result has been established independently in [71]. Here we show it by proving that the less noisy ordering of channels is preserved under polarization. Recall that a channel V is said to be *less noisy* than W if $I(T; Y) \leq I(T; Z)$ for all distributions of the form

$$p(t, x, y, z) = p(x, t) W(y|x) V(z|x), \tag{A.1}$$

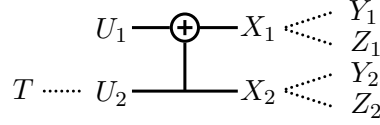


Figure A.1. Dependence graph of the random variables in (A.2).

that is, for all distributions for which $T-X-YZ$ is a Markov chain [45]. Observe that this implies $I(W) \leq I(V)$, and thus will also imply that $I(W^s) \leq I(V^s)$ for all s once we show that polarization preserves the less noisy order. Due to the recursive nature of polarization, it suffices to prove the latter claim for a single step:

Proposition A.0.1. *Let W and V be binary-input channels. If V is less noisy than W , then*

(i) V^+ is less noisy than W^+ ,

(ii) V^- is less noisy than W^- .

Proof. To prove (i), we will show that $I(T; Y_1 Y_2 U_1) \leq I(T; Z_1 Z_2 U_1)$ for all random variables (T, U_1^2, Y_1^2, Z_1^2) that are jointly distributed as

$$\begin{aligned} p(t, u_1^2, y_1^2, z_1^2) \\ = p(t, u_2) W^+(y_1^2, u_1 | u_2) V^+(z_1^2, u_1 | u_2). \end{aligned} \tag{A.2}$$

Note that the channels W^+ and V^+ here share an output, namely U_1 , but this does not affect the mutual informations in question. This assumption on the joint distribution will simplify the proof. Define $X_1 = U_1 + U_2$ and $X_2 = U_2$ (see Figure A.1). We have

$$\begin{aligned} I(T; Y_1 Y_2 U_1) &= I(T; Y_1 Y_2 | U_1) \\ &= I(T; Y_1 | U_1) + I(T; Y_2 | Y_1 U_1) \\ &\leq I(T; Y_1 | U_1) + I(T; Z_2 | Y_1 U_1) \\ &= I(T; Y_1 Z_2 | U_1) \\ &= I(T; Z_2 | U_1) + I(T; Y_1 | Z_2 U_1) \end{aligned}$$

$$\begin{aligned} &\leq I(T; Z_2|U_1) + I(T; Z_1|Z_2U_1) \\ &= I(T; Z_1Z_2|U_1). \end{aligned}$$

To see the first inequality, note that

$$TU_1U_2X_1Y_1Z_1—X_2—Y_2Z_2$$

is a Markov chain. Therefore we have

$$\begin{aligned} &p(t, u_2, x_1, x_2, y_2, z_1, z_2|y_1, u_1) \\ &= p(x_2)p(y_2, z_2|x_2)\frac{p(t, u_1, u_2, x_1, y_1, z_1|x_2)}{p(y_1, u_1)}. \end{aligned}$$

That is, conditioned on $Y_1 = y_1$ and $U_1 = u_1$,

$$TU_2X_1Z_1—X_2—Y_2Z_2$$

is a Markov chain, and therefore so is $T—X_2—Y_2Z_2$. This and the less noisiness of V imply $I(T; Y_2|Y_1 = y_1, U_1 = u_1) \leq I(T; Z_2|Y_1 = y_1, U_1 = u_1)$. Averaging over (y_1, u_1) yields the first inequality. Similarly, for the second inequality, note that

$$TU_1U_2Y_2Z_2—X_1—Y_1Z_1$$

is a Markov chain, and therefore so is $T—X_1—Y_1Z_1$ for every $Z_2 = z_2$ and $U_1 = u_1$. The less noisy relation then implies $I(T; Y_1|Z_2 = z_2, U_1 = u_1) \leq I(T; Z_1|Z_2 = z_2, U_1 = u_1)$. Averaging over (z_2, u_1) yields the inequality.

To prove (ii), we need to show that $I(T; Y_1Y_2) \leq I(T; Z_1Z_2)$ for all (T, U_1, Y_1^2, Z_1^2)

for which

$$\begin{aligned} p(t, u_1, y_1^2) &= q(t, u_1)W^-(y_1^2|u_1) \\ p(t, u_1, z_1^2) &= q(t, u_1)V^-(z_1^2|u_1) \end{aligned} \tag{A.3}$$

We will also define a random variable U_2 such that (T, U_1^2, Y_1^2, Z_1^2) is jointly distributed as

$$\begin{aligned} p(t, u_1^2, y_1^2, z_1^2) &= \frac{1}{2}q(t, u_1)W(y_1|u_1 + u_2)W(y_2|u_2) \\ &\quad \cdot V(z_1|u_1 + u_2)V(z_2|u_2). \end{aligned} \tag{A.4}$$

Observe that this definition is consistent with (A.3), it will simplify the proof. Defining again $X_1 = U_1 + U_2$ and $X_2 = U_2$ (see Figure A.2), we can write

$$\begin{aligned} I(T; Y_1 Y_2) &= I(T; Y_1) + I(T; Y_2 | Y_1) \\ &\leq I(T; Y_1) + I(T; Z_2 | Y_1) \\ &= I(T; Z_2) + I(T; Y_1 | Z_2) \\ &\leq I(T; Z_2) + I(T; Z_1 | Z_2) \\ &= I(T; Z_1 Z_2). \end{aligned}$$

To see the first inequality, note that the distribution in (A.4) implies that

$$TU_1U_2X_1Y_1Z_1-X_2-Y_2Z_2$$

is a Markov chain. Therefore we have

$$\begin{aligned} p(t, u_1, u_2, x_1, x_2, y_2, z_1, z_2 | y_1) \\ = p(x_2)p(y_2, z_2 | x_2) \frac{p(t, u_1, u_2, x_1, y_1)}{p(y_1)}. \end{aligned}$$

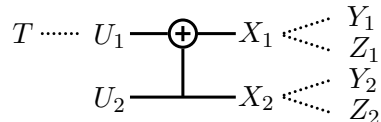


Figure A.2. Dependence graph of the random variables in (A.4).

That is, for any fixed value of Y_1 ,

$$TU_1U_2X_1Z_1—X_2—Y_2Z_2$$

is a Markov chain, and therefore so is $T—X_2—Y_2Z_2$. This and the less noisiness of V imply $I(T; Y_2 | Y_1 = y_1) \leq I(T; Z_2 | Y_1 = y_1)$. Averaging over y_1 yields the first inequality. The proof of the second inequality follows by similar arguments. \square

Note that the choice of the polarization transform and the alphabet size are immaterial to the proof above, and thus the result holds in more generality as long as the polarized channels are appropriately defined.

An interesting question here is whether weaker relations than the less noisy ordering are preserved under polarization. One well-known such relation is the *more capable* relation [45]. A channel $V(z|x)$ is said to be more capable than $W(y|x)$ if $I(X; Y) \leq I(X; Z)$ for all $p(x, y, z) = p(x)W(y|x)V(z|x)$. This ordering is not preserved under polarization, however. To see this, note that in the class of symmetric binary-input channels with a given capacity, the binary symmetric channel W is the least capable [65, Lemma 7.1]. At the same time, $I(V^-) \leq I(W^-)$ for any channel V in this class [65, Lemma 2.1].

Acknowledgment

This appendix is, in part, a reprint of the material in the paper: Eren Şaşoğlu and Lele Wang, “Universal polarization,” accepted for publication in *IEEE Transactions on Information Theory*, 2015.

Appendix B

A Sufficient Condition for Less Noisy Ordering

We establish an easy-to-check sufficient condition for less noisy ordering over the class of binary-input memoryless symmetric (BMS) channels.

Recall that each discrete BMS channel can be decomposed into a convex combination of a collection of binary symmetric channels (BSC), where the BSC with crossover probability $p_i \in [0, 1/2]$ is chosen with probability α_i , $i = 1, \dots, L$. The BMS is fully characterized by its *crossover probability profile*—the random variable T with pmf $p_T(p_i) = \alpha_i$, for $i = 1, \dots, L$.

Definition B.0.1. *The Lorenz curve of a random variable T is defined as*

$$L_T(u) = \int_{1-u}^1 F_T^{-1}(y) dy, \quad 0 \leq u \leq 1,$$

where $F_T^{-1}(y) := \sup\{t: F_T(t) \leq y\}$ is the inverse cdf of T . We say random variable T weakly majorizes random variable S from below if $L_S(u) \leq L_T(u)$ for all $u \in [0, 1]$.

The following is an equivalent condition for weak majorization [53, Chapters 3.C.1.b, 4.B.2].

Lemma B.0.5. *T weakly majorizes S from below if and only if $E(g(S)) \leq E(g(T))$ for all continuous increasing convex functions $g: \mathcal{R} \rightarrow \mathcal{R}$.*

Definition B.0.2. *We say channel $X \rightarrow Y$ is less noisy than channel $X \rightarrow Z$ if $I(U; Z) \leq I(U; Y)$ for all (U, X) for which $U - X - (Y, Z)$ form a Markov chain.*

The main result of this appendix is as follows.

Theorem B.0.5 (A Sufficient Condition). *Consider two BMS channels $X \rightarrow Y$ and $X \rightarrow Z$ with crossover probability profiles T and S respectively. If $(1 - 2T)^2$ weakly majorizes $(1 - 2S)^2$ from below, then channel $X \rightarrow Y$ is less noisy than $X \rightarrow Z$.*

Proof. We first recall the following equivalent characterization of less noisy partial order.

Lemma B.0.6 ([73]). *Channel $X \rightarrow Y$ is less noisy than $X \rightarrow Z$ if and only if the function*

$$f(p(x)) := I(X; Y) - I(X; Z) \tag{B.1}$$

is concave in $p(x)$.

Taking the second derivative with respect to $\gamma := p_X(1)$, for BMS channels, condition (B.1) simplifies to

$$E_T \left(\frac{(1 - 2T)^2}{(1 - \gamma * T)(\gamma * T)} \right) \leq E_S \left(\frac{(1 - 2S)^2}{(1 - \gamma * S)(\gamma * S)} \right) \tag{B.2}$$

for all $\gamma \in (0, 1/2]$, where $a * b = (1 - a)b + a(1 - b)$. Define $k(p) = (1 - 2p)^2$ for $p \in [0, 1/2]$.

Clearly $k(p)$ is one-to-one and its inverse function is $k^{-1}(y) = 0.5(1 - \sqrt{y})$ for $y \in [0, 1]$.

Define

$$g(\gamma, y) = \frac{y}{(1 - \gamma * k^{-1}(y))(y * k^{-1}(y))}$$

for $\gamma \in (0, 1/2]$ and $y \in [0, 1]$. Then condition (B.2) is equivalent as

$$E_{k(T)}(g(\gamma, k(T))) \leq E_{k(S)}(g(\gamma, k(S))) \tag{B.3}$$

for all $\gamma \in (0, 1/2]$. By Lemma B.0.5, it suffices to check that the function $g(\gamma, y)$ is continuous convex increasing in y for every $\gamma \in (0, 1/2]$.

Clearly $g(\gamma, y)$ is continuous. It is increasing in y since for each $\gamma \in (0, 1/2]$,

$$\frac{\partial g(\gamma, y)}{\partial y} = (1 - \gamma * k^{-1}(y))(\gamma * k^{-1}(y)) + \frac{1}{4}(1 - 2(\gamma * k^{-1}(y)))(1 - 2\gamma)\sqrt{y} \geq 0.$$

It is convex in y since

$$\frac{\partial^2 g(\gamma, y)}{\partial y^2} = \frac{1}{4(1 - \gamma * k^{-1}(y))^2(\gamma * k^{-1}(y))^2} > 0.$$

□

Remark B.0.6. *This is, however, not a necessary condition. Let $X \rightarrow Y$ be a BSC(0.11) and $X \rightarrow Z$ be a BMS with crossover probability profile $p(0.08716) = 0.73353$ and $p(0.5) = 0.26647$. One can check that $X \rightarrow Y$ is less noisy than $X \rightarrow Z$ from necessary and sufficient condition (B.2). However, weak majorization is violated since $0.08716 < 0.11$.*

Corollary B.0.1. *Over the class of BMS with the same $E((1 - 2T)^2)$, binary erasure channel (BEC) is the least noisy and BSC is the most noisy.*

Proof. Note that $L_T(1) = E((1 - 2T)^2)$. The claim then follows from an inspection at the Lorenz curves in Figure B.1. □

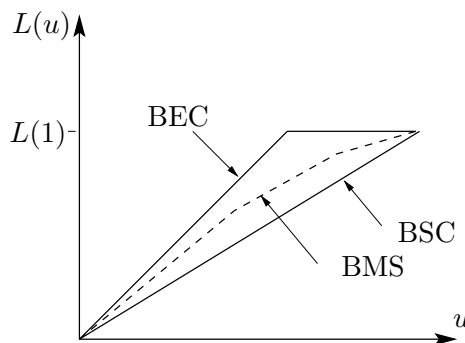


Figure B.1. Lorenz curves for a BSC, a BEC, and a generic BMS channel with the same value at $u = 1$.

Bibliography

- [1] 3GPP TS 36.212, “Multiplexing and channel coding,” Release 12, 2013.
- [2] R. Ahlswede and G. Dueck, “Good codes can be produced by a few permutations,” *IEEE Trans. Inf. Theory*, vol. 28, no. 3, pp. 430–443, May 1982.
- [3] M. Andersson, V. Rathi, R. Thobaben, J. Kliewer, and M. Skoglund, “Nested polar codes for wiretap and relay channels,” *Communications Letters, IEEE*, vol. 14, no. 8, pp. 752–754, Aug. 2010.
- [4] E. Arıkan, “Source polarization,” in *Proc. IEEE Internat. Symp. Inf. Theory*, Austin, TX, Jun. 2010, pp. 899–903.
- [5] —, “Polar coding for the slepian-wolf problem based on monotone chain rules,” in *Proc. IEEE Internat. Symp. Inf. Theory*, Cambridge, MA, 2012, pp. 566–570.
- [6] E. Arıkan and E. Telatar, “On the rate of channel polarization,” in *Proc. IEEE Internat. Symp. Inf. Theory*, Seoul, South Korea, Jul. 2009, pp. 1493–1495.
- [7] E. Arıkan, “Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels,” *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3051–3073, Jul. 2009.
- [8] B. Bandemer, A. El Gamal, and Y.-H. Kim, “Optimal achievable rates for interference networks with random codes,” 2012, preprint available at <http://arxiv.org/abs/1210.4596/>.
- [9] —, “Optimal achievable rates for interference networks with random codes,” 2012. [Online]. Available: <http://arxiv.org/abs/1210.4596/>
- [10] P. P. Bergmans, “Random coding theorem for broadcast channels with degraded components,” *IEEE Trans. Inf. Theory*, vol. 19, no. 2, pp. 197–207, 1973.
- [11] D. Blackwell, L. Breiman, and A. J. Thomasian, “The capacity of a class of channels,” *Ann. Math. Statist.*, vol. 30, no. 4, pp. 1229–1241, 1959.
- [12] R. Blasco-Serrano, R. Thobaben, M. Andersson, V. Rathi, and M. Skoglund, “Polar codes for cooperative relaying,” *Communications, IEEE Transactions on*, vol. 60, no. 11, pp. 3263–3273, Nov. 2012.

- [13] V. Cadambe and S. A. Jafar, "Interference alignment and degrees of freedom of the K -user interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3425–3441, Aug. 2008.
- [14] A. B. Carleial, "Interference channels," *IEEE Trans. Inf. Theory*, vol. 24, no. 1, pp. 60–70, 1978.
- [15] J. Chen, D. ke He, and A. Jagmohan, "The equivalence between Slepian–Wolf coding and channel coding under density evolution," *Communications, IEEE Transactions on*, vol. 57, no. 9, pp. 2534–2540, September 2009.
- [16] J. Chen, D. ke He, A. Jagmohan, L. Lastras-Montano, and E. hui Yang, "On the linear codebook-level duality between Slepian–Wolf coding and channel coding," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5575–5590, Dec 2009.
- [17] Y.-K. Chia and A. El Gamal, "3-receiver broadcast channels with common and confidential messages," in *Proc. IEEE Internat. Symp. Inf. Theory*, Seoul, Korea, June/July 2009, pp. 1849–1853.
- [18] H.-F. Chong, M. Motani, H. K. Garg, and H. El Gamal, "On the Han–Kobayashi region for the interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 3188–3195, Jul. 2008.
- [19] R. Chou and M. Bloch, "Polar coding for the broadcast channel with confidential messages," in *Proc. IEEE Inf. Theory Workshop*, Jerusalem, Israel, April 2015.
- [20] S.-Y. Chung, J. Forney, G.D., T. Richardson, and R. Urbanke, "On the design of low-density parity-check codes within 0.0045 db of the shannon limit," *Communications Letters, IEEE*, vol. 5, no. 2, pp. 58–60, Feb 2001.
- [21] T. Coleman, A. Lee, M. Medard, and M. Effros, "Low-complexity approaches to Slepian–Wolf near-lossless distributed data compression," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3546–3561, Aug 2006.
- [22] T. M. Cover, "Broadcast channels," *IEEE Trans. Inf. Theory*, vol. 18, no. 1, pp. 2–14, Jan. 1972.
- [23] —, "An achievable rate region for the broadcast channel," *IEEE Trans. Inf. Theory*, vol. 21, no. 4, pp. 399–404, 1975.
- [24] T. M. Cover and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 5, pp. 572–584, Sep. 1979.
- [25] T. M. Cover and C. S. K. Leung, "An achievable rate region for the multiple-access channel with feedback," *IEEE Trans. Inf. Theory*, vol. 27, no. 3, pp. 292–298, 1981.
- [26] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. 24, no. 3, pp. 339–348, 1978.
- [27] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge: Cambridge University Press, 2011.

- [28] O. Fawzi and I. Savov, "Rate-splitting in the presence of multiple receivers," 2012. [Online]. Available: <http://arxiv.org/abs/1207.0543>
- [29] R. G. Gallager, "Capacity and coding for degraded broadcast channels," *Probl. Inf. Transm.*, vol. 10, no. 3, pp. 3–14, 1974.
- [30] J. Garcia-Frias, "Compression of correlated binary sources using turbo codes," *Communications Letters, IEEE*, vol. 5, no. 10, pp. 417–419, Oct 2001.
- [31] N. Goela, E. Abbe, and M. Gastpar, "Polar codes for broadcast channels," *IEEE Trans. Inf. Theory*, vol. 61, no. 2, pp. 758–782, Feb 2015.
- [32] A. J. Grant, B. Rimoldi, R. Urbanke, and P. A. Whiting, "Rate-splitting multiple access for discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 873–890, 2001.
- [33] T. Gulcu and A. Barg, "Achieving secrecy capacity of the wiretap channel and broadcast channel with a confidential component," in *Proc. IEEE Inf. Theory Workshop*, Jerusalem, Israel, April 2015.
- [34] T. S. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 49–60, 1981.
- [35] S. H. Hassani, S. B. Korada, and R. Urbanke, "The compound capacity of polar codes," in *Proc. 47th Ann. Allerton Conf. Commun. Control Comput.*, Oct. 2009, pp. 16–21.
- [36] S. H. Hassani and R. Urbanke, "Universal polar codes," 2013. [Online]. Available: <http://arxiv.org/abs/1307.7223>
- [37] S. Hassani, S. Korada, and R. Urbanke, "The compound capacity of polar codes," in *Proc. 47th Ann. Allerton Conf. Commun. Control Comput.*, Monticello, IL, Sep. 2009, pp. 16–21.
- [38] S. Hassani and R. Urbanke, "Universal polar codes," in *Proc. IEEE Internat. Symp. Inf. Theory*, Honolulu, HI, Jul. 2014, pp. 1451–1455.
- [39] J. Honda and H. Yamamoto, "Polar coding without alphabet extension for asymmetric models," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 7829–7838, Dec. 2013.
- [40] D. Karas, K. Pappi, and G. Karagiannidis, "Smart decode-and-forward relaying with polar codes," *Wireless Comm. Letters, IEEE*, vol. 3, no. 1, pp. 62–65, Feb. 2014.
- [41] M. Karzand, "Polar codes for degraded relay channels," in *Int. Zurich Seminar on Communications*, Feb. 2012, pp. 59–62.
- [42] Y.-H. Kim, S. H. Lim, and P. Minero, "Relaying via hybrid coding," in *Proc. IEEE Internat. Symp. Inf. Theory*, St. Petersburg, Russia, Aug. 2011, pp. 1946–1950.
- [43] S. B. Korada, "Polar codes for channel and source coding," Ph.D. Thesis, Ecole Polytechnique Fédérale de Lausanne, Switzerland, Sep. 2009.

- [44] S. Korada and R. Urbanke, "Polar codes are optimal for lossy source coding," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1751–1768, Apr. 2010.
- [45] J. Körner and K. Marton, "Comparison of two noisy channels," in *Topics in Information Theory (Colloquia Mathematica Societatis János Bolyai, Keszthely, Hungary, 1975)*, I. Csiszár and P. Elias, Eds. Amsterdam: North-Holland, 1977, pp. 411–423.
- [46] D. Krithivasan and S. S. Pradhan, "Lattices for distributed source coding: Jointly Gaussian sources and reconstruction of a linear function," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5628–5651, Dec. 2009.
- [47] S. Kudekar, T. Richardson, and R. Urbanke, "Spatially coupled ensembles universally achieve capacity under belief propagation," 2012. [Online]. Available: <http://arxiv.org/abs/1201.2999>
- [48] J. Lee, H. Kwon, and I. Kang, "Interference mitigation in mimo interference channel via successive single-user soft decoding," in *Proc. UCSD Inf. Theory Appl. Workshop*, 2012, pp. 180–185.
- [49] S. H. Lim, K. T. Kim, and Y.-H. Kim, "Distributed decode-forward for multicast," in *Proc. IEEE Internat. Symp. Inf. Theory*, June 2014, pp. 636–640.
- [50] S. H. Lim, Y.-H. Kim, A. El Gamal, and S.-Y. Chung, "Noisy network coding," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 3132–3152, May 2011.
- [51] W. Liu, S. H. Hassani, and R. Urbanke, "The least degraded and the least upgraded channel with respect to a channel family," 2013. [Online]. Available: <http://arxiv.org/abs/1304.5150>
- [52] H. MahdaviFar and A. Vardy, "Achieving the secrecy capacity of wiretap channels using polar codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6428–6443, Oct. 2011.
- [53] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of Majorization and Its Applications*. New York: Springer Science+Business Media, LLC, 2011.
- [54] P. Minero, S. H. Lim, and Y.-H. Kim, "Joint source–channel coding via hybrid coding," in *Proc. IEEE Internat. Symp. Inf. Theory*, St. Petersburg, Russia, Aug. 2011, pp. 884–888.
- [55] S. Miyake, "Coding theorems for point-to-point communication systems using sparse matrix codes," Ph.D. Thesis, The University of Tokyo, Tokyo, Japan, 2010.
- [56] M. Mondelli, S. Hassani, I. Sason, and R. Urbanke, "Achieving marton's region for broadcast channels using polar codes," *Information Theory, IEEE Transactions on*, vol. 61, no. 2, pp. 783–800, Feb 2015.
- [57] B. Nazer and M. Gastpar, "Compute-and-forward: Harnessing interference through structured codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6463–6486, Oct 2011.

- [58] L. H. Ozarow, "The capacity of the white Gaussian multiple access channel with feedback," *IEEE Trans. Inf. Theory*, vol. 30, no. 4, pp. 623–629, 1984.
- [59] S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): design and construction," *IEEE Trans. Inf. Theory*, vol. 49, no. 3, pp. 626–643, Mar 2003.
- [60] B. Rimoldi and R. Urbanke, "A rate-splitting approach to the Gaussian multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 42, no. 2, pp. 364–375, Mar 1996.
- [61] A. Sahebi and S. Pradhan, "Polar codes for some multi-terminal communications problems," in *Proc. IEEE Internat. Symp. Inf. Theory*, June 2014, pp. 316–320.
- [62] E. Sasoglu, E. Telatar, and E. Yeh, "Polar codes for the two-user multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6583–6592, Oct. 2013.
- [63] E. Şaşoğlu and L. Wang, "Universal polarization," 2013, submitted to *IEEE Trans. Inf. Theory*. [Online]. Available: <http://arxiv.org/abs/1307.7495>
- [64] —, "Universal polarization," in *Proc. IEEE Internat. Symp. Inf. Theory*, June 2014, pp. 1456–1460.
- [65] E. Şaşoğlu, "Polar coding theorems for discrete systems," Ph.D. Thesis, Ecole Polytechnique Fédérale de Lausanne, Switzerland, Sep. 2011.
- [66] E. Şaşoğlu, E. Telatar, and E. Arıkan, "Polarization for arbitrary discrete memoryless channels," 2009. [Online]. Available: <http://arxiv.org/abs/0908.0302>
- [67] H. Sato, "On the capacity region of a discrete two-user channel for strong interference," *IEEE Trans. Inf. Theory*, vol. 24, no. 3, pp. 377–379, May 1978.
- [68] D. Schonberg, S. Pradhan, and K. Ramchandran, "Distributed code constructions for the entire Slepian–Wolf rate region for arbitrarily correlated sources," in *Signals, Systems and Computers, 2004. Conference Record of the Thirty-Seventh Asilomar Conference on*, vol. 1, Nov 2003, pp. 835–839 Vol.1.
- [69] C. E. Shannon, "A mathematical theory of communication," *Bell System Tech. J.*, vol. 27, no. 3, pp. 379–423, 27(4):623–656, 1948.
- [70] V. Stankovic, A. Liveris, Z. Xiong, and C. Georghiades, "On code design for the slepian-wolf problem and lossless multiterminal networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1495–1507, April 2006.
- [71] D. Sutter and J. M. Renes, "Universal polar codes for more capable and less noisy channels and sources," 2013. [Online]. Available: <http://arxiv.org/abs/1312.5990>
- [72] E. C. van der Meulen, "Random coding theorems for the general discrete memoryless broadcast channel," *IEEE Trans. Inf. Theory*, vol. 21, no. 2, pp. 180–190, 1975.
- [73] M. van Dijk, "On a special class of broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. 43, no. 2, pp. 712–714, Mar. 1997.

- [74] L. Wang and E. Şaşoğlu, “Polar coding for interference networks,” 2014, submitted to *IEEE Trans. Inf. Theory*. [Online]. Available: <http://arxiv.org/abs/1401.7293>
- [75] L. Wang, “Polar coding for relay channels,” in *Proc. IEEE Internat. Symp. Inf. Theory*, Hong Kong, China, June 2015.
- [76] L. Wang and E. Sasoglu, “Polar coding for interference networks,” in *Proc. IEEE Internat. Symp. Inf. Theory*, June 2014, pp. 311–315.
- [77] L. Wang, E. Sasoglu, and Y.-H. Kim, “Sliding-window superposition coding for interference networks,” in *Proc. IEEE Internat. Symp. Inf. Theory*, Honolulu, HI, June 2014, pp. 2749–2753.
- [78] Y.-P. Wei and S. Ulukus, “Polar coding for the general wiretap channel,” in *Proc. IEEE Inf. Theory Workshop*, Jerusalem, Israel, April 2015.
- [79] J. K. Wolf, A. D. Wyner, J. Ziv, and J. Körner, “Coding for write-once memory,” *Bell System Tech. J.*, vol. 63, no. 6, pp. 1089–1112, 1984.
- [80] Y. Wu, S. Shamai, and S. Verdú, “Degrees of freedom of the interference channel: A general formula,” in *Proc. IEEE Internat. Symp. Inf. Theory*, Aug. 2011, pp. 1362–1366.
- [81] A. D. Wyner, “Recent results in the Shannon theory,” *IEEE Trans. Inf. Theory*, vol. 20, no. 1, pp. 2–10, 1974.
- [82] A. Yedla, P. Nguyen, H. Pfister, and K. Narayanan, “Universal codes for the gaussian mac via spatial coupling,” in *Proc. 49th Ann. Allerton Conf. Commun. Control Comput.*, 2011, pp. 1801–1808.
- [83] Y. Zhao, C. W. Tan, A. Avestimehr, S. Diggavi, and G. Pottie, “On the maximum achievable sum-rate with successive decoding in interference channels,” *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3798–3820, Jun. 2012.