CODING SCHEMES FOR PHYSICAL LAYER NETWORK CODING OVER A TWO-WAY RELAY CHANNEL

A Dissertation

by

BRETT MICHAEL HERN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Krishna R. Narayanan
Committee Members,	Scott Miller
	Alex Sprintson
	Anxiao Jiang
Head of Department,	Chanan Singh

August 2013

Major Subject: Electrical Engineering

ABSTRACT

We consider a two-way relay channel in which two transmitters want to exchange information through a central relay. The relay observes a superposition of the transmitted signals from which a function of the transmitted messages is computed for broadcast. We consider the design of codebooks which permit the recovery of a function at the relay and derive information-theoretic bounds on the rates for reliable decoding at the relay.

In the spirit of compute-and-forward, we present a multilevel coding scheme that permits reliable computation (or, decoding) of a class of functions at the relay. The function to be decoded is chosen at the relay depending on the channel realization. We define such a class of reliably computable functions for the proposed coding scheme and derive rates that are universally achievable over a set of channel gains when this class of functions is used at the relay. We develop our framework with general modulation formats in mind, but numerical results are presented for the case where each node transmits using 4-ary and 8-ary modulation schemes. Numerical results demonstrate that the flexibility afforded by our proposed scheme permits substantially higher rates than those achievable by always using a fixed function or considering only linear functions over higher order fields.

Our numerical results indicate that it is favorable to allow the relay to attempt both compute-and-forward and decode-and-forward decoding. Indeed, either method considered separately is suboptimal for computation over general channels. However, we obtain a converse result when the transmitters are restricted to using identical binary linear codebooks generated uniformly at random. We show that it is impossible for this code ensemble to achieve any rate higher than the maximum of the rates achieved using compute-and-forward and decode-and-forward decoding.

Finally, we turn our attention to the design of low density parity check (LDPC) ensembles which can practically achieve these information rates with joint-compute-and-forward message passing decoding. To this end, we construct a class of two-way erasure multiple access channels for which we can exactly characterize the performance of joint-compute-and-forward message passing decoding. We derive the processing rules and a density evolution like analysis for several classes of LDPC ensembles. Utilizing the universally optimal performance of spatially coupled LDPC ensembles with message passing decoding, we show that a single encoder and decoder with puncturing can achieve the optimal rate region for a range of channel parameters.

To my best friend and wife,

Jennifer

ACKNOWLEDGMENTS

I am grateful to God for giving me the opportunity, capability, and desire to pursue a doctoral degree. He has shown my family incredible grace for giving my wife and I our twin girls, Audrey and Avery, and for healing Avery from cancer when modern medicine provided no hope. I am nothing without Him.

I thank my best friend and loving wife, Jennifer. You have worked hard to help support us during my graduate school and raise our daughters while I finish my degree and begin a career. You work so hard to take the best care of our children even when it is difficult. You are an outstanding wife, an amazing mother, and you are the best friend that I've ever had. I love you cute girl!

Dr. Narayanan, thanks for your patience and encouragement for the past seven years of my education, including five years of graduate school. Your depth of knowledge, critical thinking, attention to detail, and value for effective teaching have taught me so much. I have regularly told others that you are the best boss that I am likely to have. Thank you for the opportunity and encouragement to pursue a PhD. Thank you also, for the generous financial support towards my graduate study at Texas A&M and for allowing me to finish my research in Houston when my daughter was sick. Thanks to Dr. Miller for joining my committee, for your practical approach to technical problems, and for our friendly early morning conversations. Thanks to Dr. Sprintson and Dr. Jiang for being on my committee and for your excellent critique and feedback. Thanks to all of my past professors and teachers for giving me a solid knowledge base from which to deepen my technical expertise.

I am grateful to my parents for supporting me throughout my education, teaching me good character, and giving me a love for learning. You have given me an excellent model to follow as Jennifer and I raise our own children.

I am grateful to Jennifer, my parents, Jennifer's parents, Nana Laura, and Nana Helen for helping to watch our twins during my final year of research. I could not have finished my doctorate without your help, encouragement, and support. I thank my whole family, especially my mom and mother in law for patiently trying to understand while I explain my research and encouraging me to keep talking even when you didn't always understand. You really helped me break my patterns of thinking as I finished my degree away from Texas A&M this year. Thanks to my brother's Chris Hern and Chris Konkol for listening to me practice my conference talks while I was away from school. Your interested critique truly helped make them successful. Thanks to Jennifer for always encouraging me to work hard and making it possible for me to do so. Thanks to Jennifer and my dad for attending my defense and to everyone who helped take care of babies so they could come. Thanks to everyone who has prayed for our family and for Avery's healing.

Thanks to my officemates and co-advisees, Engin and Jerry. You made the office a fun place to work. It has been especially enjoyable to become parents at nearly the same time. I wish we could have spent more time in College Station this past year so they could get to know each other. You are both great friends.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

I.1. Background

In 1948, Claude Shannon made a historic breakthrough which has lead to research disciplines known as information and coding theory [1]. Since then a half-century of technical innovations, improvements in computational technology, and economic demands have motivated changes in the way we communicate over long distances [2]. Recently, the number and functionality of portable wireless devices have grown sharply while the available wireless spectrum remains constant. Therefore there is a growing demand for innovations for the efficient use of available spectrum.

The wireless setting presents intriguing challenges because of the shared nature of the wireless medium. A wired network is typically thought of as a large collection of nodes connected with non-interfering point-to-point links or channels. In a wireless network, transmitted messages can be received by all nearby (or connected) nodes, and simultaneous transmissions interfere at all shared destinations. These features are generally referred to as the broadcast and superposition properties of wireless networks. Traditionally, the superposition property is considered a detriment to wireless communication with simultaneous transmissions relegated to orthogonal frequencies or transmission times [3]. Eventually, the superposition property was utilized by code division multiple access (CDMA) schemes which orthogonalize interfering transmissions via digital spreading sequences [3], [4]. In such schemes, the interference is often treated as a noise source with the signal to noise ratio (SNR) and signal to interference ratio (SIR) often appearing in the same equations.

Perhaps surprisingly, the change in our thinking about wireless superposition

was catalyzed by a breakthrough in wired networking called network coding [5], [6]. Network coding was first introduced to optimize network efficiency for wired relay networks with multiple sources and multiple destinations [7] and uses the principle that bandwidth efficiency may be improved if relays are allowed to forward simple functions of their received messages. For the multicast problem, it has been shown that forwarding linear combinations of received packets is sufficient to achieve the capacity for each destination in the network [8], [9].

Physical layer network coding (PLNC), introduced in [10], combines the principle of network coding with the superposition property of wireless channels by only requiring relays to recover functions of the superimposed signals for forwarding. Thus, PLNC comprises a subfield of network coding in which modulation and/or coding techniques are developed so that a relay may recover its desired functions efficiently. For a recent and approachable tutorial/survey of the key ideas behind PLNC with reliable decoding, we refer readers to [6]. For another broad tutorial/survey of PLNC with an eye towards practical implementation, we refer readers to [11].

I.2. Two-Way Relay Channel

In this thesis, we restrict our attention to the simplest network where PLNC techniques have been shown to be effective, namely, the wireless two-way relay (TWR) channel. In this simple network, node A has data to send to node B and vice versa. The relay R is included to assist in this communication, and there is no direct link between nodes A and B. The PLNC approach, shown in Fig. 1, is to allow nodes A and B to transmit simultaneously in the multiple access (MA) stage. Then in the broadcast (BC) stage, the relay broadcasts a function of the received signal to nodes A and B from which they decode their desired messages.

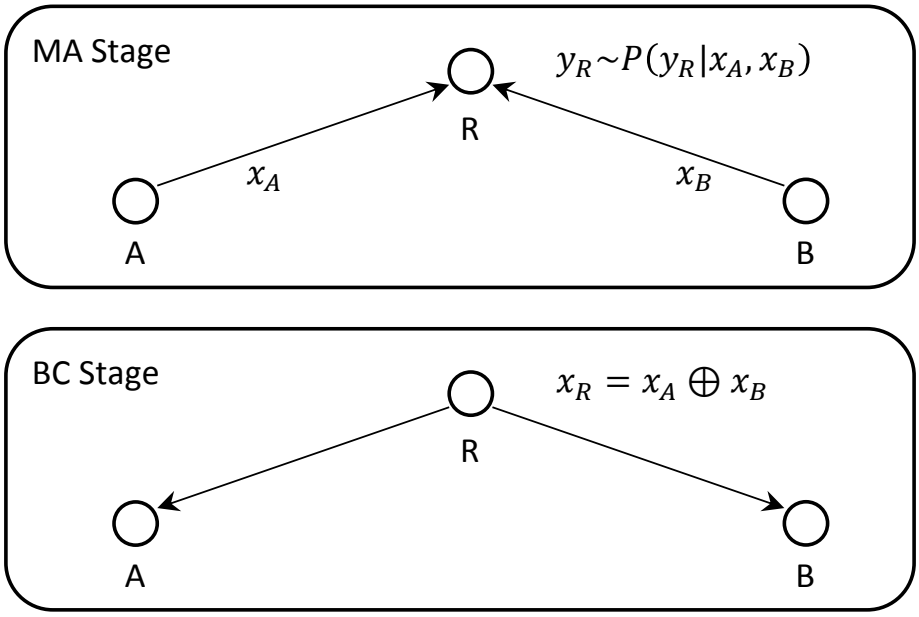


Fig. 1.: System model of a two-way relay channel with PLNC.

PLNC approaches to the TWR problem vary depending on assumptions made about the channel model. A thorough survey of such PLNC techniques is beyond the scope of this thesis. Therefore, we focus on two practical concerns which have inspired our research. First, for networks where noise is a significant impediment, it is desirable to correct errors at the relays so they do not propagate. This motivates the search for coding structures for which reliable physical layer network coding, as described in [6], can be executed efficiently. Second, we consider a network with random fading in which there is no channel state information at the transmitter (CSIT) but we assume there is perfect channel state information at the receiver (CSIR). This is a well accepted model because it is reasonable to expect the relay to estimate the channel parameters from the observed signal [12]. Therefore, we consider modulation/coding schemes which allow the relay to adapt its computed function depending on the channel. Here, we provide a brief review of recent literature which

addresses one or both of these concerns. The recent literature is compared with our results which are discussed in detail in the later chapters.

I.2.1. Reliable PLNC

Decode-and-Forward: A traditional information-theoretic approach to reliably decode a function of the received messages at the relay is to treat the channel as a multiple access channel [6], [13]. Nodes A and B use independent codebooks for encoding such that the relay can reliably decode both received codewords. Then, the relay computes a function of the encoded messages which is re-encoded and broadcast to nodes A and B. This reliable PLNC scheme is called decode-and-forward (DF).

Compute-and-Forward: PLNC achieves its efficient performance by allowing the relay to recover a function of the transmitted messages. For the two-way relay channel, an acceptable function must allow node A (node B) to recover node B's (node A's) message ideally using the knowledge of their own transmitted message. Because of the linearity of the superposition in most wireless channels, a popular approach is for node A and B to use an identical linear codebook over an appropriate field [14], [15], [6]. Then, the linear combination of the transmitted codewords is a member of this linear code and can be reliably decoded at the relay. The message associated with this linearly combined codeword is then re-encoded for broadcast. We refer to this reliable PLNC scheme as compute-and-forward (CF).

Notice that these two decoding paradigms for reliable PLNC, DF and CF, make different assumptions about the codebook structure. The approach that maximizes the achievable rate for reliable computation depends on the structure of the channel and the function to be computed. Research related to the CF paradigm typically focuses on finding functions which are well matched to both the channel and the codebooks used at nodes A and B. If a channel is ideally matched to a network

function, then the channel can be expressed as a noiseless computation of the function followed by a discrete memoryless channel.

To the best of our knowledge, DF and CF provide the best known achievable information rates for the problem of reliable function computation at a relay. This is perhaps surprising because each scheme utilizes a decoding structure which is fundamentally suboptimal for this problem. DF requires the relay to reliably recover all received codewords. Traditional CF operates on a test statistic which discards partial information about the interfering sequences which are not relevant to the desired linearly combined codeword prior to error correction. For CF based on nested lattice codes, the modular operations of lattice decoding decreases information about the observed codeword pair [15]. For CF based on binary multiple access channels, elementwise estimates of the desired xor operation are obtained prior to decoding and are processed by a message passing decoder [16].

Joint-Compute-and-Forward: To address the sub-optimality of CF decoding, several schemes which perform message passing decoding with larger message alphabets have been proposed [17], [16], [18], [19]. We refer to the decoding paradigm used in these papers as joint-compute-and-forward (JCF). The key idea of JCF is to attempt to recover as much information about the observed codeword pair as possible using a joint estimator. Then a simple function is used to combine these estimates into a hard decision about the desired linearly combined codeword.

It makes sense that a decoder which utilizes all information provided by the channel should outperform CF. Indeed, there are simulation results in [17], [16], [18], [19] which indicate that JCF decoding can outperform CF decoding for certain channel parameters. However, these results are based only on simulations and provide little insight into the operation of the JCF decoder. Particularly, the relationship between the achievable rates for DF, CF, and JCF is not clear. For the case where

nodes A and B are restricted to use identical binary linear codebooks, we provide a detailed analysis of the achievable computation rates for general discrete memoryless channels.

I.2.2. TWR Without Fading

For the TWR channel with AWGN without fading, near optimal CF schemes have been designed to maximize the exchange rate in [15], [20], [21]. Building on results from [22], these authors derive an upper bound on the capacity for complex channels of $\log(1+snr)$ and show, that with identical lattice codebooks at nodes A and B and lattice decoding at the relay, a rate of $\log(\frac{1}{2}+snr)$ is achievable, which is optimal at high SNR. This problem has also been studied for the case where there is fading in the channel, but each node perfectly knows the fading coefficients for each network link in [23]. They show that near-optimal performance can be obtained at high SNR if each transmitter inverts its channel prior to transmission. The authors in [24] apply lattices with list decoding to the two way relaying problem with a direct link between nodes A and B. Finally, CF schemes for multiple input multiple output channels have been considered in [25]. Recall that the DF scheme with independent Gaussian codebooks can achieve $\frac{1}{2}\log(1+2snr)$ for the TWR problem with AWGN [26]. Then, these lattice results introduce the notion that CF based schemes perform well at high SNR and DF based schemes work well at lower SNR for wireless channels. Therefore, networks with high SNR are commonly referred to as being interference limited [27] as opposed to noise limited. These CF schemes using lattice codes are most exciting because they introduce the notion that highly structured ensembles can outperform random coding.

I.2.3. TWR With Fading

In practical wireless channels, the transmitted signals are multiplied by random complex channel gains which change the magnitude and phase of the observed baseband waveform. The aforementioned lattice results assume that these channel gains are either fixed to unity or are fully known and invertible at the transmitters. A more realistic assumption is that the complex channel coefficients h_A and h_B are assumed to be perfectly estimated at each receiver but unknown to each transmitter.

For this scenario, the authors in [28] introduce a scheme called denoise-and-forward (DNF) which uses channel dependent denoising functions at the relay to minimize the symbol error probability. These authors focus on the case where nodes A and B use a 4-ary constellation during the multiple access stage. The relay chooses denoising functions so that the distance profile for constellation points with different labels is optimized. Their work showcased the challenge associated with unknown channel coefficients at the transmitters and the advantages of adapting the network coding functions to the channel parameters at the relay. With DNF, the symbol error rate between nodes A and B is improved; however, the relay does not attempt to correct its hard decisions prior to broadcast. Indeed, the denoising functions available to the relay are not required to have representations based on finite field operations, so the structure of the codebook induced at the relay does not favor reliable error correction. In [29], the DNF scheme is extended to be used with trellis codded modulation and Viterbi decoding. This indeed provides some coding gain, however, it is still possible for the relay to forward sequences with errors.

A lattice based CF scheme which allows both adaptation of decoding functions and error correction at the relay and which originally coined the term compute-and-forward has been presented in [26]. In this scheme, the relay decodes an integer

combination of the transmitted codewords, where the integer combination is adapted according to the channel gains. They show that such a scheme can be implemented using nested lattice codes to take advantage of the duality between modulus arithmetic in prime order fields and the modular operations of lattice decoding. Their scheme requires the construction of infinite-dimensional lattice codes which is not practical. The results in [26] are extended in a remarkable way in [30], where an algebraic framework is provided to design lattices over principle ideal domains. Their proposed coding scheme is also based on large-dimensional lattice codes.

In later chapters, we propose a novel scheme for reliable PLNC based on multilevel coding (MLC). Our scheme facilitates reliable computation, function adaptation at the relay, and implementation with practical codes over small finite fields (e.g. binary linear block codes).

I.2.4. Quantize Map Forward

The idea to approximate the superposition of real or complex signals of different signal strength by linear combinations of different bit levels over the binary field was first introduced by the linear deterministic model for arbitrary networks in [27]. In fact, they suggest that these linear combinations can be approximated with the use of a multilevel lattice code as described in [31]. For a large class of linear deterministic networks, the authors in [27] prove that a constant gap to the cut-set bound can be achieved by computing random linear combinations of the signalling levels at each relay. They extend this proof to an equivalent class of Gaussian relay networks by quantizing the signal at the noise level and re-encoding the quantized signal with random Gaussian codebooks at each relay. The resulting scheme, called quantize-map-forward (QMF) is shown to achieve a constant gap to the cut-set upper bound (for the relay Gaussian network) which does not change with SNR. This duality

between the deterministic model and the AWGN model was revisited using lattice codes for transmission and structured mappings at the relay nodes in [32]. The QMF scheme was generalized to wireless networks over general discrete memoryless channels in [33]. Particularly, the noisy network coding scheme of [33] is shown to be a generalization of QMF for Gaussian networks and several other schemes for other types of channels.

The QMF approach provides an additional perspective of the superposition of wireless signals and provides a strategy for approaching the capacity of general networks up to the limit of interference. The QMF scheme does not perform error correction at the relays and, hence, the gap to capacity can grow with the size of the network [27]. The CF strategy is generally more robust to the size of the network because the reliable computation prevents noise propagation. The lack of error correction at the relays in QMF also has implications for the decoding complexity at destination nodes. There are results for a simple relay network in [34] in which a proposed message passing decoder uses a graph structure which emulates the connections within the network. Each QMF destination nodes requires detailed knowledge of the signal interaction and encoding process of each relay within the network. Correcting errors at each relay distributes the computational complexity in a more favorable way. Also, the performance of the QMF strategy for the bi-directional relaying problem has not been studied well for the case when there is no CSIT. A QMF strategy for the TWR channel and for the multi-pair TWR channel has been proposed in [35] and [36] respectively. These schemes require the transmitters to choose code parameters based the channel gains of each interfering node, but they do achieve a constant gap to the cut-set bound for the case of known CSI at the transmitter.

I.3. Overview of Results

The TWR problem has received considerable attention in recent literature. The value of both adaptability to the channel parameters and reliable computation at the relay is well supported. However, few practical schemes which support both of these features exist. In this thesis, we fill in some of the gap between theory and practice for this problem. We present novel information-theoretic analysis of the reliable computation problem for the TWR channel. We also present practical coding schemes with numerically computed or simulated performance which match our theoretical results. The following is a more detailed list of the most important contributions of our work.

- **Multilevel Coding for Adaptive Computation:** We present a reliable PLNC scheme based on MLC. Unlike the coding schemes in [26], [30], our proposed scheme does not require a lattice code. Rather our scheme uses identical linear codebooks over small prime fields (e.g. binary linear codes) for each level of the multilevel code. Therefore, our scheme can be implemented with lower encoding and decoding flexibility. Yet, it facilitates error correction for a larger class of decoding functions than those proposed in [26]. This is because the class of functions for our scheme is derived from the large set of non-singular square matrices over $\mathbb{F}_p$ in place of the set of non-zero elements in large prime order fields. To the best of our knowledge, the use of multilevel encoding in conjunction with the linearity of a small prime order field to facilitate adaptive decoding at a relay is new.

- **Rate Penalties for Structured Ensembles:** Conventional MLC schemes require the use of independent linear codes in each level with careful selection of coding rates [37]. In contrast, the code constructions used in this thesis use

identical linear codes in each level to facilitate adaptive decoding. This requires a non-trivial extension of the achievability theorem from the conventional case. An important result in this thesis is that the use of the same linear code for each encoding level requires that penalty rate constraints must be satisfied for reliable decoding.

- **DF and CF for Improved Adaptation:** Previous research has established the notion for AWGN channels with linear superposition that DF with independent codes is optimal at low SNR while CF with structured (e.g. identical linear) codes is optimal at high SNR. We show that when there is no CSIT, it is better to attempt both CF and DF decoding with structured codes. Particularly, the phase mismatch between the channel gains sometimes makes DF decoding favorable at moderate and even high SNR.

- **Joint-Compute-and-Forward Decoding:** Previous works have introduced a notion of joint decoding for computation especially for message passing decoders. We provide an information-theoretic framework to analyze such joint-compute-and-forward decoders. We study the case where nodes A and B use identical linear codebooks for transmission over a binary-input memoryless multiple access channel. We show that JCF naturally achieves the computation rate for the better of CF or DF. We conversely show that higher rates than those achievable DF or CF (and hence JCF) cannot be achieved with the ensemble of uniformly distributed identical linear codebooks on this channel model.

- **TWEMAC Channel Model:** In order to design LDPC codes and code ensembles which achieve the optimal computation rates for identical linear codebooks, we propose a simplified two-way erasure multiple access (TWEMAC)

channel model. The TWEMAC model may be used to simplify the design and analysis of coding schemes for many problems in wireless communications. We derive and simplify the processing rules for a JCF message passing decoder on a TWEMAC. We subsequently derive efficiently computable expressions for the exact performance of LDPC ensembles with JCF message passing.

- **Spatial Coupling for Computation:** We numerically show that JCF message passing of a spatially coupled LDPC code/ensemble can achieve near-optimal computation rates for identical linear codebooks for several TWEMACs. This is complemented with simulation results of a spatially coupled protograph ensemble for the AWGN channel with binary phase shift keying.

I.4. Organization of Thesis

Throughout this thesis, we focus on the problem of function computation at a relay node suitable for a PLNC solution to the two-way relaying problem. Each chapter uses different assumptions about the channel model to allow us to focus on a different aspect of this problem. In each case, our approach to the problem is to propose or consider an encoding structure for nodes A and B which is suitable for reliable computation over the studied channel. Then, we thoroughly investigate the decoder at the relay. This involves analysis of both the information-theoretic limits for reliable decoding and the details of practical implementation.

In Chapter II, we propose a novel multilevel modulation and coding scheme for the AWGN channel with unknown fading coefficients at the relay. The realistic nature of this channel model invites a detailed analysis of encoding structures which allow the relay to adapt its computed function to the channel realization. The proposed scheme facilitates such adaptation for an larger class of decoding functions

than those achievable using linear functions over higher order finite fields. A careful information-theoretic analysis of the decoder reveals that the proposed encoding structure, which is suitable for flexible decoding, imposes penalty rate constraints for certain channel parameters. Our numerical analysis reveals that decoding flexibility can be significantly improved by allowing the relay to attempt to reliably recover both incoming codewords as in decode-and-forward.

In Chapter III, we consider a simplified encoding structure, namely identical binary linear codebooks at nodes A and B, for the general class of binary-input memoryless multiple access channels. A joint decoding structure which we call joint-compute-and-forward is investigated. Our analysis reveals that the JCF decoder essentially performs either CF or DF decoding which we show is rate optimal for the studied encoding structure.

In Chapter IV, we consider how to design LDPC ensembles which achieve the information-theoretic performance limits for reliable computation. To this end, we develop a simplified class of erasure multiple access channels for which we derive the exact performance of JCF message passing decoding. Numerical results reveal that spatially coupled LDPC ensembles with JCF message passing can achieve the desired performance.

Throughout this thesis, tools and analysis are developed which suggest further research both for the studied computation problem on the two-way relay channel and for more general wireless networks. In Chapter V we discuss our main conclusions and suggest specific ways in which our results may be generalized or extended.

Notation: Throughout this thesis, we will use the following naming conventions. Vectors or sequences will be denoted by underlined variables such as $\underline{x}$. Random variables will be denoted by upper case variables such as X, while their outcomes will be represented by lowercase variables. Matrices will be represented by capital

boldface letters such as $\mathbf{X}$. Subsets will be denoted by capital scripted letters such as $\mathcal{X}$. If a variable is associated with a specific node, this will be indicated by a subscripted capital letter like x_A. We attempt to be consistent in our naming conventions throughout this thesis. However, chapter specific notational issues which may be confusing are mentioned in a brief *note about notation* near the beginning of each chapter.

CHAPTER II

MULTILEVEL CODING SCHEMES FOR COMPUTE-AND-FORWARD WITH FLEXIBLE DECODING*

In the previous chapter, we discussed recent works which use structured code ensembles for reliable computation at rates not achievable with traditional independent encoders. This improved performance generally requires the structure of the multiple access channel to be matched to the desired function. In some cases, the exact nature of the superposition may not be known prior to transmission. Therefore schemes which permit adaptive network coding have been proposed to improve the matching at the receiver. However, few schemes which support both adaptive network coding and reliable computation exist.

In this chapter, we propose a novel multilevel coding scheme which permits the computation of a class of functions at the relay. The function to be computed (or, decoded) is chosen depending on the channel realization. We define such a class of functions and derive rates that are universally achievable over a set of channel gains when this class of functions is used for decoding. The proposed scheme facilitates improved decoding flexibility over previous schemes by providing the relay with a larger class of decoding functions than the set of linear functions over a finite field. The proposed coding scheme is also practically implementable with binary linear codebooks and common modulation schemes. We develop our framework with general modulation formats in mind, but numerical results are presented for the case where each node transmits using 4-ary and 8-ary modulation schemes. Numerical

results demonstrate that the flexibility afforded by our proposed scheme results in substantially higher rates than those achievable by always using a fixed function or considering only linear functions over higher order fields. Additionally, we discover that allowing the relay to attempt to decode both incoming codewords, as in the decode-and-forward paradigm, results in improved decoding flexibility.

This chapter is organized as follows. The key elements of the problem are outlined in Section II.1. Our proposed solution is detailed in Section II.2. An achievable rate for the proposed scheme during the MA stage is given in Section II.3. These rates are numerically determined for an example where nodes A and B transmit using 4-ary and 8-ary constellations in Section II.4. Simulation results for a regular LDPC code are shown to corroborate the information-theoretic results. Key results are reiterated in Section II.5.

Note about notation: In this chapter, we use different variables $\underline{v}$ and $\underline{x}$ to refer to multilevel codewords for channel coding and binary address vectors which are adaptively network coded respectively. This unusual notation is explained in (2.7) and the surrounding discussion. Also in this chapter, superscripts are used primarily to differentiate between different encoding levels of the multilevel encoder.

II.1. Problem Description

Each node in the relay network is assumed to be half-duplex, so communication is split into two stages, a multiple access (MA) stage and a broadcast (BC) stage. We assume perfect symbol synchronization between the transmitters and mainly focus on the MA stage throughout this thesis.

II.1.1. Multiple Access Stage

Nodes A and B each encode their binary messages $\underline{u}_A$ and $\underline{u}_B$ into codewords $\underline{v}_A \in \mathcal{C}_A$ and $\underline{v}_B \in \mathcal{C}_B$ where $\mathcal{C}_A$ and $\mathcal{C}_B$ are the codebooks used at the nodes A and B respectively. These codewords are mapped to sequences of symbols $\underline{s}_A, \underline{s}_B \in \mathcal{Q}^N$ with $|\mathcal{Q}| = 2^\ell$. The relay receives noisy observations of the sum of these symbol sequences according to

$$\underline{y}_R = h_A \underline{s}_A + h_B \underline{s}_B + \underline{w}_R \tag{2.1}$$

where h_A and h_B are complex fading coefficients, and $\underline{w}_R$ is complex additive white Gaussian noise (AWGN). This induces an effective constellation $\mathcal{Q}_R$ at the relay defined by

$$\mathcal{Q}_R = \{q_R \in \mathbb{C} | q_R = h_A q_A + h_B q_B,\ q_A, q_B \in \mathcal{Q}\}. \tag{2.2}$$

II.1.2. Adaptive Decoding at the Relay and the Induced Codebook

The main idea proposed in this chapter is the construction of a coding scheme such that the relay can reliably decode some function of $\underline{v}_A$ and $\underline{v}_B$ for a desired set of channel conditions $\mathcal{H} \subset \mathbb{C}^2$. Specifically, we jointly design codes $\mathcal{C}_A$ and $\mathcal{C}_B$ and a set of decoding functions $\mathcal{F}$ such that, for any $(h_A, h_B) \in \mathcal{H}$, there exists $f \in \mathcal{F}$ such that the relay can reliably decode $f(\underline{v}_A, \underline{v}_B)$ from $\underline{y}_R$. We require that node A (B) must be able to unambiguously decode $\underline{v}_B$ ($\underline{v}_A$) from the output of $f(\underline{v}_A, \underline{v}_B)$ with its knowledge of $\underline{v}_A$ ($\underline{v}_B$) [29]. For a given $f \in \mathcal{F}$, we will define an induced codebook at the relay as the codebook corresponding to f i.e.

$$\mathcal{C}_{f,R} = \{f(\underline{v}_A, \underline{v}_B) | \underline{v}_A \in \mathcal{C}_A,\ \underline{v}_B \in \mathcal{C}_B\}. \tag{2.3}$$

It is important to understand the structure of $\mathcal{C}_{f,R}$ since the probability of error in decoding $f(\underline{v}_A, \underline{v}_B)$ from $\underline{y}_R$ depends on h_A, h_B, and $\mathcal{C}_{f,R}$. The main advantage of

our proposed scheme is that it guarantees that choosing one codebook $\mathcal{C}_A$ and $\mathcal{C}_B$ at the transmitter can result in a good induced codebook $\mathcal{C}_{f,R}$ for a class of functions $\mathcal{F}$. More specifically, it guarantees $\mathcal{C}_{f,R}$ is a member of the ensemble of random coset codes which is an optimal ensemble for achieving the uniformly distributed input information rate for the equivalent channel between $f(\underline{v}_A, \underline{v}_B)$ and $\underline{y}_R$ for all $f \in \mathcal{F}$. We restrict our attention to classes of functions $\mathcal{F}$ which are applied componentwise at the relay. For the uncoded case, the authors in [28] have considered the design of modulation schemes that optimize the performance when the demodulating function is adapted. The broadcast stage is fairly standard and is identical to that considered in [15], [20].

II.2. Proposed Scheme

II.2.1. Multilevel Encoder

The system model for the multilevel encoder for nodes A and B and the channel model for the MA stage is shown in Fig. 2. The encoder at nodes A and B uses MLC with a different coset of the *same linear* code $\mathcal{C}$ used at each bit level. For a detailed description of MLC and achievable rates for the point-to-point channel see [37].

The encoder is described as it pertains to node A to simplify notation. First, the message $\underline{u}_A$ is split into length K sub-vectors $\underline{u}_A^1, ..., \underline{u}_A^\ell$ which form rows of an $\ell \times K$ matrix

$$\mathbf{U}_A = \begin{bmatrix} \underline{u}_A^1 \\ \vdots \\ \underline{u}_A^\ell \end{bmatrix}. \tag{2.4}$$

Each $\underline{u}_A^1, ..., \underline{u}_A^\ell$ is encoded with the same generator matrix $\mathbf{G}$ into codewords $\underline{\gamma}_A^1, ..., \underline{\gamma}_A^\ell$.

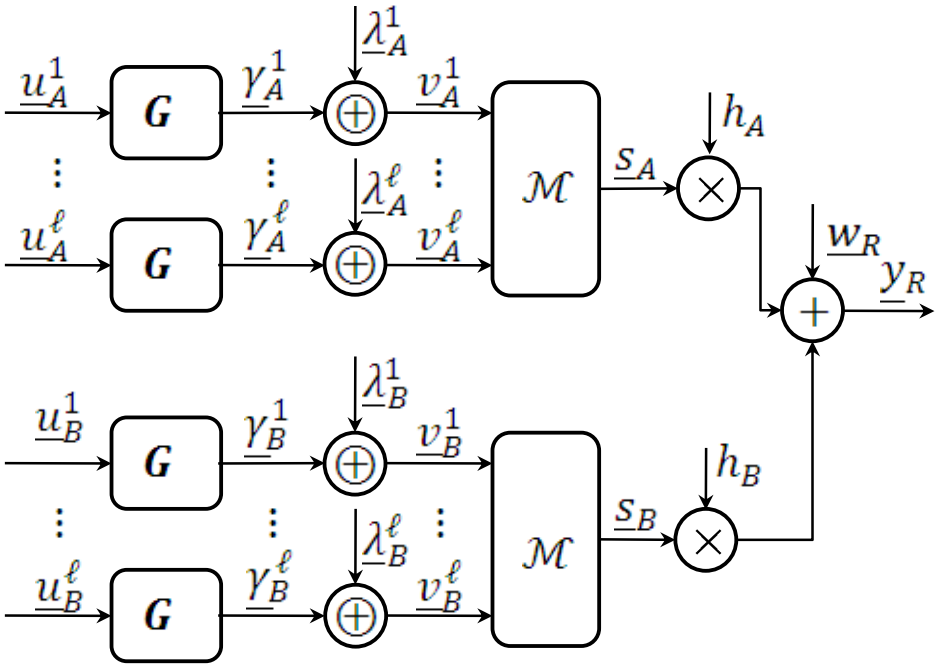


Fig. 2.: Block diagram of MLC coset encoders for MA stage.

These codewords form the rows of an $\ell \times N$ matrix

$$\mathbf{\Gamma}_A = \mathbf{U}_A\mathbf{G} = \begin{bmatrix} \underline{\gamma}_A^1 \\ \vdots \\ \underline{\gamma}_A^\ell \end{bmatrix}. \tag{2.5}$$

Finally, an i.i.d Bernoulli random sequence $\underline{\lambda}_A^k$ is added to each $\underline{\gamma}_A^k$ to symmetrize the input distribution to the channel. The symmetrized codeword $\underline{v}_A^k = \underline{\lambda}_A^k \oplus \underline{\gamma}_A^k$ is therefore a member of a random coset of the original linear code $\mathcal{C}$. The random coset leaders $\underline{\lambda}_A^k$ form the rows of a binary $\ell \times N$ matrix

$$\mathbf{\Lambda}_A = \begin{bmatrix} \underline{\lambda}_A^1 \\ \vdots \\ \underline{\lambda}_A^\ell \end{bmatrix}. \tag{2.6}$$

The resulting coset codewords $\underline{v}_A^k$ form the rows of a binary $\ell \times N$ matrix $\mathbf{X}_A$ given

by

$$\mathbf{X}_A = \mathbf{U}_A\mathbf{G} \oplus \mathbf{\Lambda}_A = \begin{bmatrix} \underline{v}_A^1 \\ \vdots \\ \underline{v}_A^\ell \end{bmatrix} = [\underline{x}_A[1], ..., \underline{x}_A[N]]\,. \tag{2.7}$$

Thus each code $\mathcal{C}_A^k,\ k \in \{1, ..., \ell\}$ will be a different coset of $\mathcal{C}$. The k_{th} row $\underline{v}_A^k$ of $\mathbf{X}_A$ is then a codeword of $\mathcal{C}_A^k$. We use the two variables $\underline{x}_A[n]$ and $\underline{v}_A^k$ to refer to the n_{th} column and k_{th} row of $\mathbf{X}_A$ respectively because it will simplify our notation later. It should be mentioned here that much of the intuition about the main result in the chapter is best obtained by ignoring the fact that cosets are used at each layer and simply considering the use of identical linear codes at each level in the MLC scheme. The coset matrix $\mathbf{\Lambda}_A$ is included to symmetrize the effective channel at the relay (i.e. $\mathbf{\Lambda}_A$ is necessary for the proofs to be correct).

The n_{th} binary address vector $\underline{x}_A[n] \in \mathbb{F}_2^\ell$ maps to a symbol $\underline{s}_A[n] \in \mathcal{Q}$ through the use of a symbol mapping function $\mathcal{M} : \mathbb{F}_2^\ell \to \mathcal{Q}$. An example of such a mapping function is given in Fig. 3 where $\mathcal{Q}$ is the QPSK constellation. As shown, the mapping function is usually derived by partitioning the set of signaling points in $\mathcal{Q}$ into equal sized subsets [37]. Let $\mathcal{S} \subseteq \{1, ..., \ell\}$ be the subset of elements of $\underline{x}_A$ which are fixed. Then let $\mathcal{X}_{\{x_A^k | k \in \mathcal{S}\}} \subseteq \mathbb{F}_2^\ell$ be the set of $\underline{x}_A$'s with the same given values for all elements in $\mathcal{S}$. Then we define the output of $\mathcal{M}(\{x_A^k | k \in \mathcal{S}\}) \subseteq \mathcal{Q}$ according to

$$\mathcal{M}(\{x_A^k | k \in \mathcal{S}\}) = \bigcup_{\underline{b} \in \mathcal{X}_{\{x_A^k | k \in \mathcal{S}\}}} \mathcal{M}(\underline{b}) \tag{2.8}$$

This means that the returned subset of constellation points is the subset whose address vectors are equal to the known bits for all indices, $\mathcal{S}$. The output of $\mathcal{M}(\{x_A^k | k \in \mathcal{S}\})$ is $2^{\ell - |\mathcal{S}|}$ constellation points.

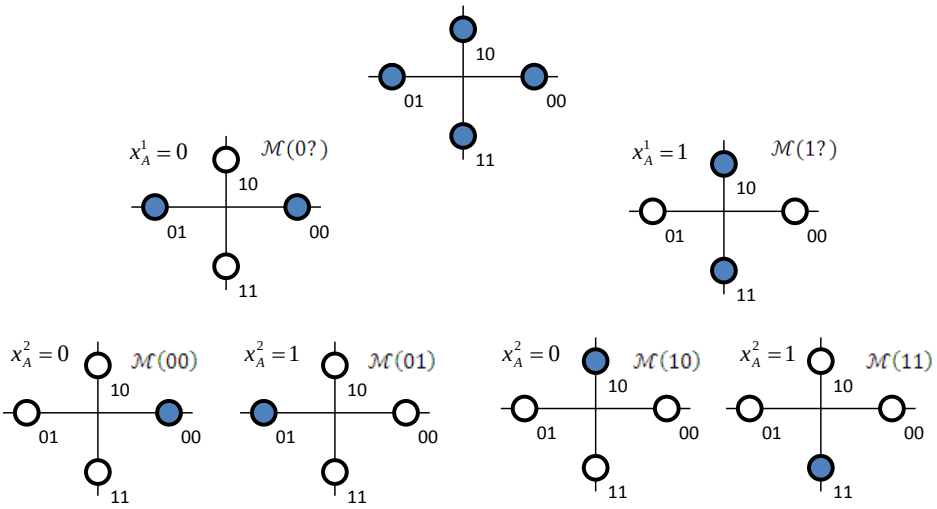


Fig. 3.: Example of MLC address labeling by set partitioning using QPSK.

In the example in Fig. 3 if $\mathcal{S} = \{2\} \subseteq \{1, 2\}$ and $x_A^2 = 1$, then

$$\mathcal{M}(\{x_A^k | k \in \mathcal{S}\}) = \mathcal{M}(\{x_A^2 = 1\}) = \{\mathcal{M}(01), \mathcal{M}(11)\} = \{-1, -j\}.$$

Here, $|\mathcal{S}| = 1$, and $\ell = 2$. Therefore $\mathcal{M}$ returns $2^{2-1} = 2$ constellation points.

II.2.2. Adaptive Decoding at the Relay

As mentioned previously, the goal of the proposed scheme is to allow the relay to decode a function of the transmitted codewords. Similar to the compute-and-forward scheme, our scheme utilizes the linearity of the base code $\mathcal{C}$ and the fact that the relay knows $\mathbf{\Lambda}_A$ and $\mathbf{\Lambda}_B$. If nodes A and B encode their messages as described, the set of decoding functions $\mathcal{F}$ which the relay can use for decoding is defined as follows.

Define $\mathcal{D}$ as the set of $\ell \times \ell$ binary matrices which are invertible over $\mathbb{F}_2$. The

set of functions we consider is given by

$$\mathcal{F} = \left\{ f : \mathbb{F}_2^\ell \times \mathbb{F}_2^\ell \to \mathbb{F}_2^\ell \middle| f(\underline{x}_A, \underline{x}_B) = [\mathbf{D}_A \mathbf{D}_B] \begin{bmatrix} \underline{x}_A \\ \underline{x}_B \end{bmatrix}, \ \mathbf{D}_A, \mathbf{D}_B \in \mathcal{D} \right\}. \tag{2.9}$$

Therefore a given $f \in \mathcal{F}$ is defined by some $\mathbf{D}_A, \mathbf{D}_B \in \mathcal{D}$ from which the relay should attempt to decode a matrix $\mathbf{X}_{f,R}$ given by

$$\mathbf{X}_{f,R} = [\mathbf{D}_A \mathbf{D}_B] \begin{bmatrix} \mathbf{X}_A \\ \mathbf{X}_B \end{bmatrix}. \tag{2.10}$$

Due to the linearity of matrix multiplication, we can express the desired matrix $\mathbf{X}_{f,R}$ as

$$\begin{aligned} \mathbf{X}_{f,R} &= [\mathbf{D}_A \mathbf{D}_B] \begin{bmatrix} \mathbf{X}_A \\ \mathbf{X}_B \end{bmatrix} = [\mathbf{D}_A \mathbf{D}_B] \begin{bmatrix} \mathbf{U}_A \mathbf{G} \oplus \mathbf{\Lambda}_A \\ \mathbf{U}_B \mathbf{G} \oplus \mathbf{\Lambda}_B \end{bmatrix} \\ &= [\mathbf{D}_A \mathbf{D}_B] \begin{bmatrix} \mathbf{U}_A \mathbf{G} \\ \mathbf{U}_B \mathbf{G} \end{bmatrix} \oplus [\mathbf{D}_A \mathbf{D}_B] \begin{bmatrix} \mathbf{\Lambda}_A \\ \mathbf{\Lambda}_B \end{bmatrix} \\ &= [\mathbf{D}_A \mathbf{D}_B] \begin{bmatrix} \mathbf{U}_A \\ \mathbf{U}_B \end{bmatrix} \mathbf{G} \oplus [\mathbf{D}_A \mathbf{D}_B] \begin{bmatrix} \mathbf{\Lambda}_A \\ \mathbf{\Lambda}_B \end{bmatrix} \\ &= \mathbf{U}_{f,R} \mathbf{G} \oplus \mathbf{\Lambda}_{f,R}. \end{aligned} \tag{2.11}$$

Here, we see that the matrix $\mathbf{X}_{f,R}$ can be written in terms of an effective message $\mathbf{U}_{f,R}$ and coset matrix $\mathbf{\Lambda}_{f,R}$ which can be computed separately based on f. Thus each row of $\mathbf{X}_{f,R}$ is a codeword from a different coset code of $\mathcal{C}$. Note that f is applied elementwise to the sequences $\underline{s}_A$ and $\underline{s}_B$.

For clarification, consider the case of $\ell = 2$. Let a function f_1 be defined by $\mathbf{D}_A = \mathbf{D}_B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. Writing the vectors $\underline{x}_A$ and $\underline{x}_B$ as $[x_A^1 \ x_A^2]^T$ and $[x_B^1 \ x_B^2]^T$

respectively, we see that $\underline{x}_{f_1,R} = [x^1_{f_1,R}\ x^2_{f_1,R}]^T$ is given by

$$\underline{x}_{f_1,R} = f_1(\underline{x}_A, \underline{x}_B) = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} x^1_A \\ x^2_A \\ x^1_B \\ x^2_B \end{bmatrix}. \tag{2.12}$$

This corresponds to the binary XOR function given by $f_1(\underline{x}_A, \underline{x}_B) = [x^1_A \oplus x^1_B, x^2_A \oplus x^2_B]^T$.

Define another function f_2 using $\mathbf{D}_A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $\mathbf{D}_B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. This is the rotated-XOR function given by $f_2(\underline{x}_A, \underline{x}_B) = [x^1_A \oplus x^2_B, x^2_A \oplus x^1_B]^T$.

Recall from (2.7), that $\underline{v}^k[n] \Leftrightarrow x^k[n]$. Thus using f_1 at the relay corresponds to decoding $[\underline{v}^1_A \oplus \underline{v}^1_B]$ and $[\underline{v}^2_A \oplus \underline{v}^2_B]$. Similarly, applying f_2 at the relay corresponds to decoding $[\underline{v}^1_A \oplus \underline{v}^2_B]$ and $[\underline{v}^2_A \oplus \underline{v}^1_B]$.

To illustrate the importance of choosing the decoding function f depending on (h_A, h_B), consider an example with

$$\mathcal{Q} = \{1, j, -1, -j\} = \{\mathcal{M}(00), \mathcal{M}(01), \mathcal{M}(11), \mathcal{M}(10)\}$$

(i.e. QPSK with Gray Labeling). Further, let $h_A = 1$ and $h_B = e^{j\theta}$. Then θ is the phase difference between node A and B's channel gains. Consider the decoding functions

$$f_1(\underline{x}_A, \underline{x}_B) = [x^1_A \oplus x^1_B, x^2_A \oplus x^2_B]$$
$$f_2(\underline{x}_A, \underline{x}_B) = [x^1_A \oplus x^2_B, x^2_A \oplus x^1_B].$$

The resulting constellation $\mathcal{Q}_R$ at the relay is shown for different values of θ in Fig. 4. Note that the complex coordinates of the constellation points are exactly the

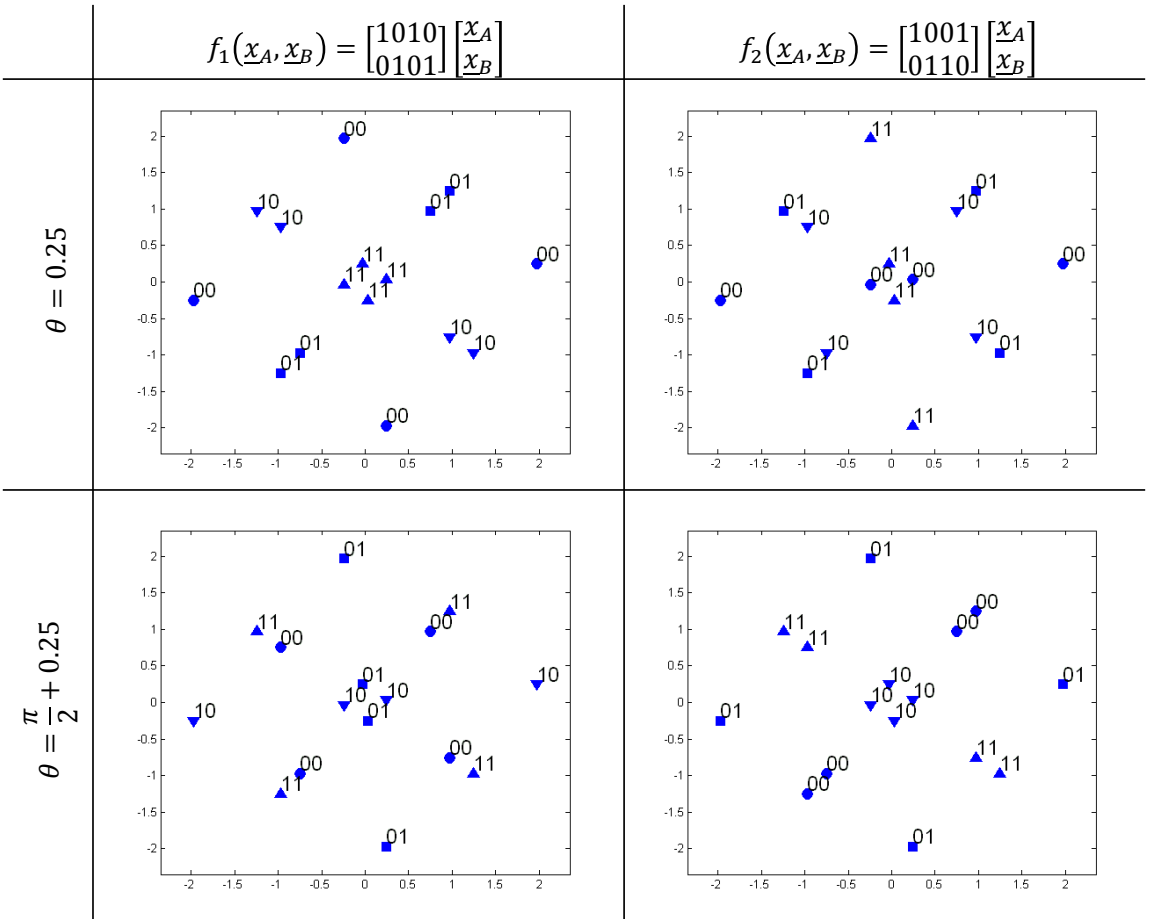


Fig. 4.: Effective constellation at relay for different values of θ.

same, but their labels are different based on θ and $f \in \{f_1, f_2\}$. When $\theta \approx 0$, f_1 appears to ensure larger distances between points with unequal labels than f_2. The situation is reversed when $\theta \approx \frac{\pi}{2}$. This suggests that the performance for a fixed decoding function can vary widely with θ even when both $|h_A|$ and $|h_B|$ are large. This motivating illustration is based on similar examples from [28] which includes a more through analysis of the benefits of function adaptation. We include it here for completeness.

As illustrated in Fig. 4, each $f \in \mathcal{F}$ induces a different mapping function $\mathcal{M}_{f,R} : \mathbb{F}_2^\ell \to \mathcal{Q}_R$. Since $|\mathcal{Q}_R| = 2^{2\ell}$ and each $\underline{x}_{f,R}$ has length ℓ, $\mathcal{M}_{f,R}(\underline{x}_{f,R})$ returns a set of 2^ℓ points similar to (2.8) for the point-to-point case. Particularly this is given by

$$\mathcal{M}_{f,R}(\underline{x}_{f,R}) = \bigcup_{\{(\underline{x}_A, \underline{x}_B) | f(\underline{x}_A, \underline{x}_B) = \underline{x}_{f,R}\}} h_A \mathcal{M}(\underline{x}_A) + h_B \mathcal{M}(\underline{x}_B). \tag{2.13}$$

Similar to the description of $\mathcal{M}$, let $\mathcal{S} \subseteq \{1, ..., \ell\}$ be the subset of elements from $\underline{x}_{f,R}$ which are fixed. Then let $\mathcal{X}_{\{x_{f,R}^k | k \in \mathcal{S}\}} \subseteq \mathbb{F}_2^\ell$ be the set of $\underline{x}_{f,R}$'s with the same given values for all elements in $\mathcal{S}$. Then the output of $\mathcal{M}(\{x_{f,R}^k | k \in \mathcal{S}\}) \subseteq \mathcal{Q}_R$ is defined

$$\mathcal{M}_{f,R}(\{\underline{x}_{f,R}^k | k \in \mathcal{S}\}) = \bigcup_{\underline{b} \in \mathcal{X}_{\{\underline{x}_{f,R}^k | k \in \mathcal{S}\}}} \mathcal{M}_{f,R}(\underline{b}). \tag{2.14}$$

For the example in Fig. 4, $\mathcal{M}_{f,R}(11)$ returns the four constellation points labeled 11 in each figure. Similarly, $\mathcal{M}_{f,R}(\{x_{f,R}^1 = 1\})$ returns the eight constellation points in the union $\mathcal{M}_{f,R}(11) \cup \mathcal{M}_{f,R}(10)$.

In order for nodes A and B to be able to unambiguously decode their desired messages, the authors in [28] show that f must satisfy

$$\begin{aligned} f(\underline{x}_A, \underline{x}_B) &\neq f(\underline{x}'_A, \underline{x}_B) \; \forall \; \underline{x}_A \neq \underline{x}'_A \text{ and } \underline{x}_B \\ f(\underline{x}_A, \underline{x}_B) &\neq f(\underline{x}_A, \underline{x}'_B) \; \forall \; \underline{x}_B \neq \underline{x}'_B \text{ and } \underline{x}_A. \end{aligned} \tag{2.15}$$

The authors in [28] have called this the 'exclusive law'. In this thesis, we call functions which satisfy this property unambiguous. The relay broadcasts the values of the decoded functions $f(\underline{x}_A, \underline{x}_B)$ to nodes A and B. Therefore, if any ambiguous functions are used during the MA stage, nodes A and B will be unable to recover their desired message from $f(\underline{x}_A, \underline{x}_B)$ and their knowledge of their own message.

Lemma II.1. *For any $\mathbf{D}_A, \mathbf{D}_B \in \mathcal{D}$, a decoding function*

$$f(\underline{x}_A, \underline{x}_B) = [\mathbf{D}_A \mathbf{D}_B] \begin{bmatrix} \underline{x}_A \\ \\ \underline{x}_B \end{bmatrix} \tag{2.16}$$

is unambiguous.

Proof. The proof follows from the invertibility of $\mathbf{D}_A$ and $\mathbf{D}_B$. For some $\underline{x}_A$, suppose that there exists $\underline{x}_B \neq \underline{x}'_B$ so that

$$[\mathbf{D}_A \mathbf{D}_B] \begin{bmatrix} \underline{x}_A \\ \\ \underline{x}_B \end{bmatrix} = [\mathbf{D}_A \mathbf{D}_B] \begin{bmatrix} \underline{x}_A \\ \\ \underline{x}'_B \end{bmatrix}.$$

This can be written as

$$\begin{aligned}
\mathbf{D}_A \underline{x}_A \oplus \mathbf{D}_B \underline{x}_B &= \mathbf{D}_A \underline{x}_A \oplus \mathbf{D}_B \underline{x}'_B \\
\mathbf{D}_A \underline{x}_A \oplus \mathbf{D}_A \underline{x}_A \oplus \mathbf{D}_B \underline{x}_B &= \mathbf{D}_A \underline{x}_A \oplus \mathbf{D}_A \underline{x}_A \oplus \mathbf{D}_B \underline{x}'_B \\
\mathbf{D}_B \underline{x}_B &= \mathbf{D}_B \underline{x}'_B \\
\mathbf{D}_B^{-1} \mathbf{D}_B \underline{x}_B &= \mathbf{D}_B^{-1} \mathbf{D}_B \underline{x}'_B \\
\underline{x}_B &= \underline{x}'_B
\end{aligned}$$

which is a contradiction. □

It has been observed in [15] that for some channel gains and signal to noise

ratios, Decode-and-Forward (DF), i.e., decoding

$$\mathbf{X}_{AB} = \begin{bmatrix} \mathbf{X}_A \\ \mathbf{X}_B \end{bmatrix} \tag{2.17}$$

can provide higher information rates than compute-and-forward. After $\mathbf{X}_{AB}$ has been decoded, any of the functions from the set $\mathcal{F}$ in (2.16) can be recovered from $\mathbf{X}_{AB}$.

Therefore, there are two primary methods to recover our desired function. The Compute-and-Forward (CF) decoder directly attempts to recover $\underline{x}_{f,R}$ from the symbolwise estimates

$$P(Y_R[n]|\underline{X}_{f,R}[n]) = \sum_{\{(\underline{x}_A,\underline{x}_B)|f(\underline{x}_A,\underline{x}_B)=\underline{x}_{f,R}[n]\}} P(Y_R[n]|\underline{x}_A,\underline{x}_B). \tag{2.18}$$

The DF decoder attempts to recover $\underline{x}_{f,R}$ by first reliably decoding $\mathbf{X}_{AB}$ from the observations $P(Y_R[n]|\underline{X}_A[n],\underline{X}_B[n])$. For the remainder of the chapter, we will focus on the complete recovery of functions from the set $\mathcal{F}$ from (2.16) with either CF or DF decoding.

II.3. Achievable Information Rates

II.3.1. Achievable Rates for General Discrete Memoryless Channels

So far, we have described the multilevel encoder used by nodes A and B which uses a coset of the same linear code in each level. We have justified this scheme by deriving the set of functions which can be unambiguously decoded at a relay node if this encoder is used. Here, we derive the rates which can be achieved by this scheme for fixed channel parameters. To accomplish this, we will derive the achievable rates for the proposed encoding scheme (where random cosets of the same linear code with

generator matrix $\mathbf{G}$ are used at each level) for a general discrete memoryless channel (DMC) with input alphabet $\mathbb{F}_2^\ell$ in Theorem II.1. Then we extend this to define the achievable rates for decode-and-forward and compute-and-forward using the channel model we have described.

Consider a DMC with ℓ level codeword $\mathbf{X} = \mathbf{UG} \oplus \mathbf{\Lambda}$ as the input and noisy observations $\underline{y}$ as the output, where $P(\underline{Y}|\mathbf{X}) = \prod_{n=1}^{N} P(\underline{Y}[n]|\underline{X}[n])$. Let each element of $\mathbf{U}$, $\mathbf{G}$, and $\mathbf{\Lambda}$ be i.i.d. Bernoulli random variables with parameter $\frac{1}{2}$. Therefore, over the ensemble of codes, each X^k is Bernoulli distributed with parameter $\frac{1}{2}$.

Theorem II.1. *Let $\mathcal{Z}_\ell$ be the set of families of subsets of $\{1, ..., \ell\}$ which are non-empty and disjoint. That is, $\{\mathcal{S}_1, ..., \mathcal{S}_p\} \in \mathcal{Z}_\ell$ if each $\mathcal{S}_i \subseteq \{1, ..., \ell\}$ is non-empty and $\mathcal{S}_i \cap \mathcal{S}_j = \emptyset \ \forall \ i \neq j$. Define $\mathcal{S} = \bigcup_{i=1}^{p} \mathcal{S}_i$ and $\overline{\mathcal{S}} = \{1, ..., \ell\} \setminus \mathcal{S}$. Then the receiver can reliably recover $\mathbf{X}$ from $\underline{y}$ as long as*

$$\mathcal{R} < \min_{\{\mathcal{S}_1, ..., \mathcal{S}_p \in \mathcal{Z}_\ell\}} \frac{1}{p} I(Y; \{X^k | k \in \mathcal{S}\} | \{X^k | k \in \overline{\mathcal{S}}\}, \{X^k \oplus Z_1 | k \in \mathcal{S}_1\}, ..., \{X^k \oplus Z_p | k \in \mathcal{S}_p\}). \tag{2.19}$$

Each $Z_i, i \in \{1, ..., p\}$ is a Bernoulli random variable with parameter $\frac{1}{2}$.

Proof. The detailed proof is provided in Section II.6. However, the key steps in the proof are outlined below.

Our proof uses the standard approach of deriving upper bounds on the probability of error for a jointly typical decoder averaged over a carefully chosen ensemble of codes. The ensemble considered here is obtained by using random cosets of the *same linear* code for each signaling level in the multilevel coding scheme. *The use of a coset of the same linear code in each level is an important ingredient in our proposed scheme since we allow the relay to decode linear combinations of codewords from different signaling levels. This ensures that for each $f \in \mathcal{F}$, $\mathcal{C}_{f,R}^k, \ k \in \{1, ..., \ell\}$ is a coset of $\mathcal{C}$. However, this is also what complicates the proof.* If each node used

an independent linear code, then the rate region is fully characterized by the coding theorem for the multiple access channel. While independent linear codes have been used widely to obtain achievable rates for MLC for the point-to-point channel and the multiple access channel, the former ensemble has not been analyzed in detail in the literature. The key contribution of our proof in Section II.6 is to derive the achievable rates with identical linear codes at each level. □

Corollary II.1. *For the special case of $\ell = 2$, the achievable information rate can be expressed*

$$\mathcal{R} < \min\left\{\frac{1}{2}I(Y;X^1,X^2),\ I(Y;X^1|X^2),\ I(Y;X^2|X^1),\ I(Y;X^1,X^2|X^1\oplus X^2)\right\}. \tag{2.20}$$

Proof. This can be shown by letting the sets $\mathcal{S}_1,...,\mathcal{S}_p \subseteq \{1,2\}$ take each of the values in $\mathcal{Z}_2$. These are given by

$$\begin{aligned}&\{\mathcal{S}_1=\{1\},\mathcal{S}_2=\{2\}\}\\ &\{\mathcal{S}_1=\{1\}\}\\ &\{\mathcal{S}_1=\{2\}\}\\ &\{\mathcal{S}_1=\{1,2\}\}.\end{aligned} \tag{2.21}$$

Notice that the first three terms in (2.20) are also required by the proof for multilevel coding with independent linear codes (i.e. they characterize the multiple access rate region). The last bound is a result of the requirement that each signaling level uses a coset of the same linear code. Note that

$$I(Y;X^1,X^2|X^1\oplus Z_1,X^2\oplus Z_1) = I(Y;X^1,X^2|X^1\oplus X^2).$$

That is, $\{X^1\oplus Z_1, X^2\oplus Z_1\}$ and $\{X^1\oplus X^2\}$ carry the same information about

(X^1, X^2). A detailed proof of this corollary is given in the appendix as part of the proof for Theorem II.1. □

II.3.2. Achievable Rates at the Relay

Theorem II.1 is flexible enough to be applied to general channels, however its primary purpose is to indicate the achievable rates at a relay node which wants to recover a function of the two messages $\mathbf{X}_A$ and $\mathbf{X}_B$ with either CF or DF decoding. Note that for the extension of Theorem II.1 to the multiple access stage, the elements of $\mathbf{G}$, $\mathbf{\Lambda}_A$, and $\mathbf{\Lambda}_B$ are i.i.d. Bernoulli random variables with parameter $\frac{1}{2}$. The following corollaries define the rates at which the relay can reliably recover a function $f \in \mathcal{F}$ of the transmitted messages using CF or DF decoding.

Corollary II.2. *For a fixed $\mathbf{D}_A, \mathbf{D}_B \in \mathcal{D}$ let*

$$\underline{x}_{f,R} = f(\underline{x}_A, \underline{x}_B) = [\mathbf{D}_A \mathbf{D}_B] \begin{bmatrix} \underline{x}_A \\ \underline{x}_B \end{bmatrix}. \tag{2.22}$$

Then the relay can reliably recover $\mathbf{X}_{f,R} = \left[\underline{x}_{f,R}[1], ..., \underline{x}_{f,R}[N]\right]$ at the relay using a compute-and-forward decoder as long as $\mathcal{R}$ satisfies

$$\mathcal{R} < \min_{\{\mathcal{S}_1,...,\mathcal{S}_p\} \in \mathcal{Z}_\ell} \frac{1}{p} I(Y_R; \{X^k_{f,R} | k \in \mathcal{S}\} | \{X^k_{f,R} | k \in \overline{\mathcal{S}}\}, \{X^k_{f,R} \oplus Z_i | k \in \mathcal{S}_i\} \ \forall \ i \in \{1, ..., p\}). \tag{2.23}$$

Corollary II.3. *The relay can reliably recover $\mathbf{X}_{AB}$ at the relay using a decode-and-forward decoder as long as $\mathcal{R}$ satisfies*

$$\mathcal{R} < \min_{\{\mathcal{S}_1,...,\mathcal{S}_p\} \in \mathcal{Z}_{2\ell}} \frac{1}{p} I(Y_R; \{X^k_{AB} | k \in \mathcal{S}\} | \{X^k_{AB} | k \in \overline{\mathcal{S}}\}, \{X^k_{AB} \oplus Z_i | k \in \mathcal{S}_i\} \ \forall \ i \in \{1, ..., p\}). \tag{2.24}$$

The relay can then broadcast any $f \in \mathcal{F}$ to nodes A and B.

II.3.3. Universally Achievable Rate

Since nodes A and B do not have channel state information, it is important that there exists a single code from the ensemble of random coset codes which can achieve the rates from Theorem II.1 for many channel parameters. More rigorously, define $\mathcal{P}_\ell$ as a finite set of DMCs with input alphabet $\mathbb{F}_2^\ell$. We say that a rate $\mathcal{R}$ is universally achievable over $\mathcal{P}_\ell$ if there exists a fixed linear code $\mathcal{C}$ of rate $\mathcal{R}$ and coset matrix $\mathbf{\Lambda}$ such that for each $P(Y|\underline{X}) \in \mathcal{P}_\ell$, $\mathbf{X}$ can be reliably decoded at the receiver. The existence of such a code is verified in Theorem II.2.

Theorem II.2. *Define $\mathcal{R}(P(Y|\underline{X}))$ as the supremum of rates satisfying* (2.19) *for a fixed $P(Y|\underline{X}) \in \mathcal{P}_\ell$. Then any rate $\mathcal{R}_{\mathcal{P}_\ell}$ such that*

$$\mathcal{R}_{\mathcal{P}_\ell} < \min_{P(Y|\underline{X})\in\mathcal{P}_\ell} \mathcal{R}(P(Y|\underline{X})) \tag{2.25}$$

is universally achievable over $\mathcal{P}_\ell$.

Proof. Define $\delta > 0$ as the acceptable probability of error for a finite length code and choose a fixed $\mathcal{R} < \mathcal{R}_{\mathcal{P}_\ell}$. We will first consider an arbitrary $P(Y|\underline{X})$ such that $\mathcal{R} < \mathcal{R}(P(Y|\underline{X}))$. Define Ω^N as the set of linear codes of the form $\mathcal{C}$ of length N and rate $\mathcal{R}$. Thus, by increasing the value of N we form a sequence of ensembles of linear codes Ω. Define $P(Err|\Omega^N)$ as the ensemble average probability of decoding error for the ensemble Ω^N. Define $P(Err|\mathcal{C})$ as the probability of decoding error for a particular $\mathcal{C}$ (averaged over the set of random cosets $\mathbf{\Lambda}$). Define $\Omega_{bad}^N \subset \Omega^N$ as

$$\Omega_{bad}^N = \{\mathcal{C} \in \Omega^N | P(Err|\mathcal{C}) \geq \delta\}. \tag{2.26}$$

Then let $\Omega_{good}^N = \Omega^N \setminus \Omega_{bad}^N$. Define $P(\Omega_{bad}^N) = \frac{|\Omega_{bad}^N|}{|\Omega^N|}$ and $P(\Omega_{good}^N) = \frac{|\Omega_{good}^N|}{|\Omega^N|}$ as the probability that a bad code or good code is selected uniformly at random from Ω^N

respectively. We know that

$$\begin{aligned} P(Err|\Omega^N) &= P(\Omega^N_{bad})P(Err|\Omega^N_{bad}) + P(\Omega^N_{good})P(Err|\Omega^N_{good}) \\ &\geq P(\Omega^N_{bad})\delta + P(\Omega^N_{good})P(Err|\Omega^N_{good}) \\ &\geq P(\Omega^N_{bad})\delta. \end{aligned}$$

The proof of Theorem II.1 relies on showing that $\lim_{N\to\infty} P(Err|\Omega^N) = 0$. Therefore, there exists some N_0 such that for any $N > N_0$,

$$P(\Omega^N_{bad})\delta \leq P(Err|\Omega^N) < \frac{\delta}{\tau} \;\Rightarrow\; P(\Omega^N_{bad}) < \frac{1}{\tau}$$

for some finite $\tau > 2|\mathcal{P}_\ell|$. This means that $|\Omega^N_{bad}| < \frac{|\Omega^N|}{\tau}$. Note that choosing $\tau > 2|\mathcal{P}_\ell|$ is arbitrary but ensures that τ will be "large enough" to complete the proof.

We want to show the existence of some fixed $\mathcal{C} \in \Omega^N$ such that for every $P(Y|\underline{X}) \in \mathcal{P}_\ell$ we have $P(Err|\mathcal{C}) < \delta$. We can apply the steps above to find a set $\Omega^N_{bad}(P(Y|\underline{X}))$ for each $P(Y|\underline{X}) \in \mathcal{P}_\ell$. Since $|\mathcal{P}_\ell|$ is finite, the largest N required by any $P(Y|\underline{X}) \in \mathcal{P}_\ell$ must exist and be a finite integer N_{max}.

Since τ is chosen to be larger than $2|\mathcal{P}_\ell|$, the set

$$\Omega^{N_{max}} \setminus \left\{ \bigcup_{P(Y|\underline{X})\in\mathcal{P}_\ell} \Omega^{N_{max}}_{bad}(P(Y|\underline{X})) \right\}$$

must be non-empty because

$$\sum_{P(Y|\underline{X})\in\mathcal{P}_\ell} |\Omega^{N_{max}}_{bad}(P(Y|\underline{X}))| \leq |\Omega^{N_{max}}|/2.$$

Thus, since at least half of the codes are always good, there exists at least one $\mathcal{C}$ which allows reliable decoding for every $P(Y|\underline{X}) \in \mathcal{P}_\ell$ as long as $\mathcal{R} < \mathcal{R}_{\mathcal{P}_\ell}$.

□

Theorem II.2 can be applied directly to the problem of recovering a function of $\mathbf{X}_A$ and $\mathbf{X}_B$ at the relay using either CF or DF. This results in the following corollaries.

Corollary II.4. *For a fixed $f \in \mathcal{F}$ and (h_A, h_B), define $\mathcal{R}_f(h_A, h_B)$ as the supremum of rates satisfying* (2.23) *where $\underline{x}_{f,R} = f(\underline{x}_A, \underline{x}_B)$. Then define*

$$\mathcal{R}_{CF}(h_A, h_B) = \max_{f \in \mathcal{F}} \mathcal{R}_f(h_A, h_B). \tag{2.27}$$

For any finite set of channel gains, $\mathcal{H} \subset \mathbb{C}^2$, any rate $\mathcal{R}_{\mathcal{H},CF}$ which satisfies

$$\mathcal{R}_{\mathcal{H},CF} < \min_{(h_A,h_B)\in\mathcal{H}} \mathcal{R}_{CF}(h_A, h_B) \tag{2.28}$$

is universally achievable with CF decoding.

Proof. For each $(h_A, h_B) \in \mathcal{H}$, choose the $f \in \mathcal{F}$ to maximize $\mathcal{R}_f(h_A, h_B)$. Then Theorem II.2 can be applied with the set of channels $\mathcal{P}_\ell$ of the form $P(Y_R|\underline{X}_{f,R})$. □

Corollary II.5. *For fixed (h_A, h_B), define $\mathcal{R}_{DF}(h_A, h_B)$ as the supremum of rates satisfying* (2.24). *For any finite set of channel gains $\mathcal{H} \subset \mathbb{C}^2$, any rate $\mathcal{R}_{\mathcal{H},DF}$ such that*

$$\mathcal{R}_{\mathcal{H},DF} < \min_{(h_A,h_B)\in\mathcal{H}} \mathcal{R}_{DF}(h_A, h_B) \tag{2.29}$$

is universally achievable on $\mathcal{H}$ using DF decoding.

Proof. For each $(h_A, h_B) \in \mathcal{H}$, Theorem II.2 can be applied with the set of channels $\mathcal{P}_\ell$ of the form $P(Y|\underline{X}_{AB})$. □

One of the unique features of our scheme is that it allows the relay to choose between CF and DF decoding without requiring any changes at the encoder. This means that the relay can perform such an adaptation without feedback. More importantly, this indicates that the universally achievable rate $\mathcal{R}_{\mathcal{H}}$ for a fixed set of

channel gains $\mathcal{H}$ may be larger than $\max(\mathcal{R}_{\mathcal{H},CF}, \mathcal{R}_{\mathcal{H},DF})$. Particularly, it is possible for $\mathcal{R}_{\mathcal{H},CF}$ and $\mathcal{R}_{\mathcal{H},DF}$ to be limited by different channel gain pairs. The universally achievable rate is defined in *Corollary 6*.

Corollary II.6. *For any finite set of channel gains $\mathcal{H} \subset \mathbb{C}^2$, the relay can reliably decode some function of $\mathbf{X}_A$ and $\mathbf{X}_B$ as long as*

$$\mathcal{R}_{\mathcal{H}} < \min_{(h_A,h_B)\in\mathcal{H}} \max\left(\mathcal{R}_{DF}(h_A,h_B), \mathcal{R}_{CF}(h_A,h_B)\right). \tag{2.30}$$

Proof. We define two sets of channels $\mathcal{P}_\ell$ and $\mathcal{P}_{2\ell}$. For each $(h_A, h_B) \in \mathcal{H}$, choose the $f \in \mathcal{F}$ to maximize $\mathcal{R}_f(h_A, h_B)$. Then, for every $(h_A, h_B) \in \mathcal{H}$ for which $\mathcal{R}_f(h_A, h_B) > \mathcal{R}_{DF}(h_A, h_B)$, Theorem II.2 can be applied to the set of channels $\mathcal{P}_\ell$ of the form $P(Y_R|\underline{X}_{f,R})$. Likewise, for every $(h_A, h_B) \in \mathcal{H}$ for which $\mathcal{R}_{DF}(h_A, h_B) > \mathcal{R}_f(h_A, h_B)$, Theorem II.2 can be applied to the set of channels $\mathcal{P}_{2\ell}$ of the form $P(Y_R|\underline{X}_{AB})$. Each application of Theorem II.2 shows that the majority of linear codes are good. Therefore, there must exist at least one linear code which is good for both sets of channels $\mathcal{P}_\ell$ and $\mathcal{P}_{2\ell}$. □

Note that in order for this problem to be practically interesting, the set $\mathcal{H}$ should be meaningfully defined. It may seem more natural to evaluate our scheme based on the outage probability for a fixed transmission rate, and we include an outage experiment in the numerical results section for completeness. However, we focus on the universally achievable rate formulation because it is more meaningful to discuss function computation at the physical layer when both channel gains are large. If one of the channel gains is small, decode-and-forward can provide near-optimal symmetric information rates (i.e. the rate will be limited by the weak user). In the presence of two strong channels, the outage probability will likely be dominated by the phase differences θ which limit $\mathcal{R}_{\mathcal{H}}$. The $\mathcal{R}_{\mathcal{H}}$ metric leads to natural characterization of

different features of the proposed scheme for a specific input constellation $\mathcal{Q}$. Particularly, the flexibility of the proposed scheme to phase mismatch between nodes A and B can be illustrated. This is especially interesting if we consider a system where the relay is used to provide power control information to nodes A and B as in [23].

II.4. Numerical Results

II.4.1. Results for 4-qam

We begin by presenting some numerical results on the achievable rates with the proposed scheme when $\ell = 2$ and both nodes use 4-qam as the modulation scheme.

II.4.1.1. Extra Rate Constraint

First, we illustrate the effects of the additional rate constraints which are imposed if nodes A and B use cosets of the same linear code for each level. Consider the rate at which the relay can reliably recover the function, $f_1(\underline{x}_A, \underline{x}_B)$ defined in (2.12), using CF decoding when $h_A = 1$, $h_B = e^{j\theta}$, and $snr = 6$ dB. For the proposed scheme, the achievable rate (2.20) is plotted as a function of θ in Fig. 5. For comparison, if the use of identical linear codes for each bit-level did not impose a rate penalty, our scheme could achieve

$$\mathcal{R} < \min\left\{ I(Y_R; X^1_{f_1,R}|X^2_{f_1,R}), I(Y_R; X^2_{f_1,R}|X^1_{f_1,R}), \frac{1}{2} I(Y_R; X^1_{f_1,R}, X^2_{f_1,R}) \right\} \tag{2.31}$$

which corresponds to the multiple access rate region with equal rate codes. This achievable rate is also plotted as a function of θ in Fig. 5. The additional rate constraint $I(Y_R; X^1_{f_1,R}, X^2_{f_1,R}|X^1_{f_1,R} \oplus X^2_{f_1,R})$ required by the proposed scheme is seen to be dominant for values of θ close to $\frac{\pi}{2}$.

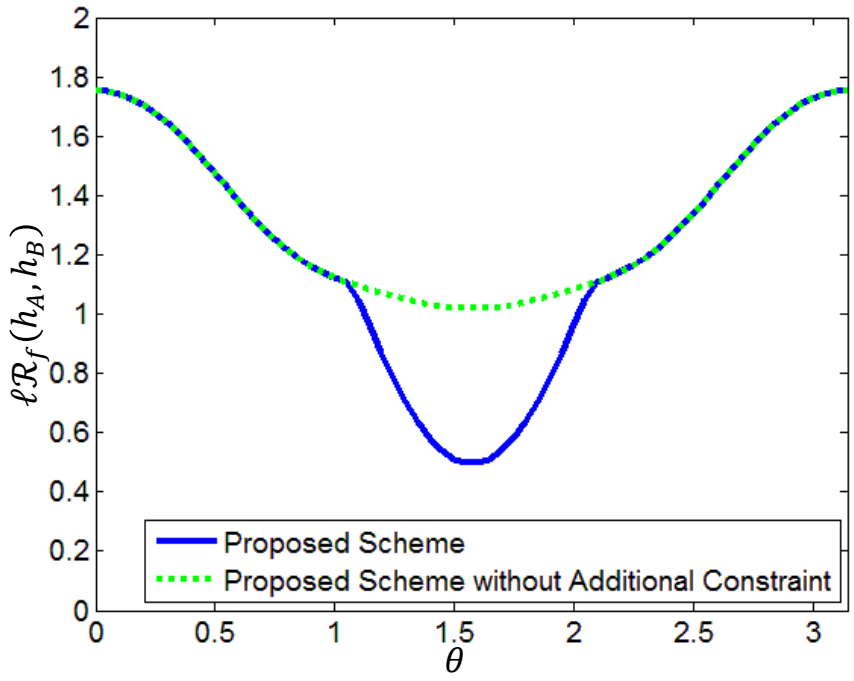


Fig. 5.: Achievable rates for the function f_1. The extra rate constraint $I(Y_R; X^1_{f_1,R}, X^2_{f_1,R} | X^1_{f_1,R} \oplus X^2_{f_1,R})$ must be satisfied if nodes A and B use the proposed MLC scheme.

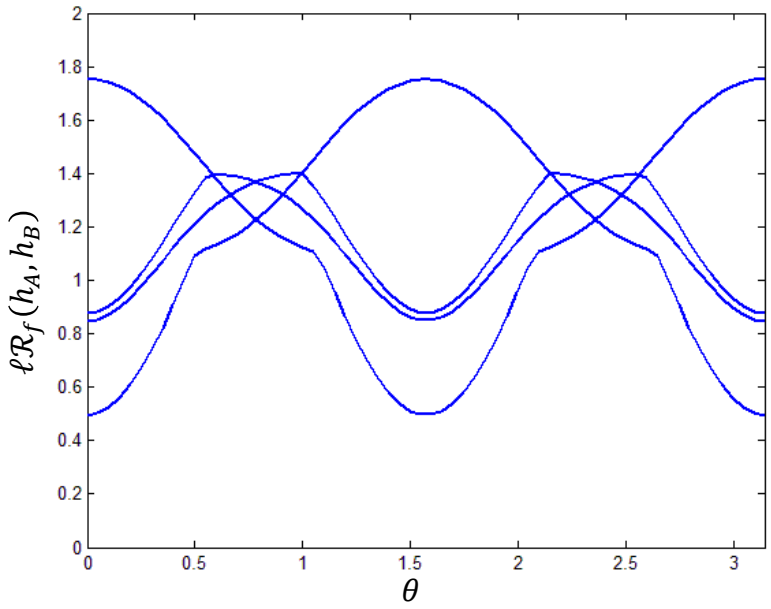


Fig. 6.: $\ell\mathcal{R}_f(h_A, h_B)$ vs. θ for each function $f \in \mathcal{F}$ with the proposed coding scheme using CF decoding.

II.4.1.2. Achievable Information Rates for the Proposed Scheme and Comparison to Coding over GF(4)

In Fig. 6, the achievable rate with the proposed scheme and a CF decoder is shown as a function of θ for each $f \in \mathcal{F}$. It can be seen that for different θ, different functions provide better performance demonstrating the importance of adapting the function at the relay.

Next, we consider the performance of a compute-and-forward scheme using the same linear code $\mathcal{C}_{GF(4)}$ of rate $\mathcal{R}_{GF(4)}$ over $GF(4)$. Specifically, the relay uses the set of decoding functions $\mathcal{F}_{GF(4)}$ corresponding to linear combinations of codewords of the form

$$\underline{v}_R = f(\underline{v}_A, \underline{v}_B) = \alpha\underline{v}_A \oplus \beta\underline{v}_B, \ \alpha, \beta \in \mathbb{F}_4 \backslash \{0\}. \tag{2.32}$$

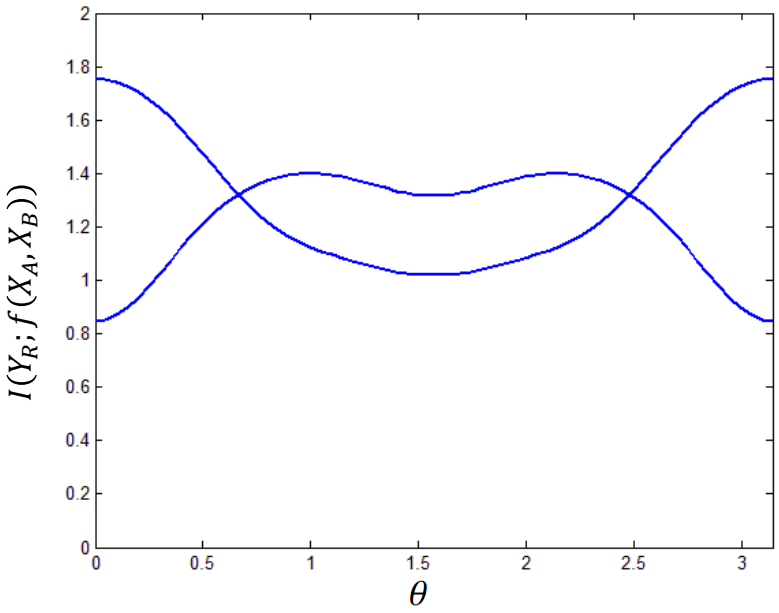


Fig. 7.: $I(Y_R; f(X_A, X_B))$ vs. θ for each linear function $f \in \mathcal{F}_{GF(4)}$ with identical linear codebooks over GF(4) using CF decoding.

Node A can decode $\underline{v}_B$ from $(\underline{v}_A, \underline{v}_R)$ by

$$\underline{v}_B = \beta^{-1}((-\alpha \underline{v}_A) \oplus \underline{v}_R). \tag{2.33}$$

Node B can recover $\underline{v}_A$ similarly. The relay should be able to decode $\underline{v}_R$ reliably as long as there exists some $f \in \mathcal{F}_{GF(4)}$ for which

$$\mathcal{R}_{GF(4)} < I(Y_R; f(X_A, X_B)). \tag{2.34}$$

For, $h_A = 1$, $h_B = e^{j\theta}$, and $snr = 6$ dB, the achievable rates for each function in $\mathcal{F}_{GF(4)}$ are plotted as a function of θ in Fig. 7. There are only two curves because several of the functions in $\mathcal{F}_{GF(4)}$ have the same distance profile for QPSK signaling. From Fig. 6 and Fig. 7, it can be seen that the proposed scheme with the CF decoder provides higher achievable rates than using codes over $GF(4)$ and a CF decoder for

every θ.

II.4.1.3. CF and DF Decoding

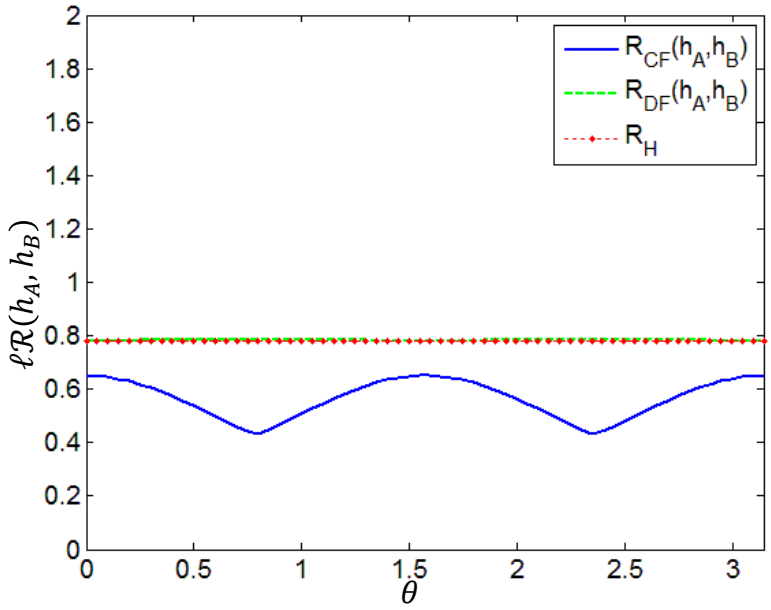
Recall that the proposed scheme facilitates the use of DF decoding at the relay in addition to function adaptation with CF decoding. Therefore, the achievable rates at the relay using CF and DF decoding are plotted as a function of θ in Figs. 8(a), 8(b), 8(c) for $snr = 0,\ 6$, and 12 dB respectively. The CF rates represent the best computation rate for the set of functions $\mathcal{F}$ for the given channel parameters as defined in (2.27). As expected, DF decoding performs better at low SNR, and CF decoding performs better at high SNR. For each of these SNRs, the universally achievable rate $\mathcal{R}_{\mathcal{H}}$ defined in Corollary II.6, is also shown for the $h_A = 1$ and $h_B = e^{j\theta}$ case. Note that this is equivalent to the set of channel gains

$$\mathcal{H} = \{(h_A, h_B) | h_A = e^{j\theta_A}, h_B = e^{j\theta_B}\} \tag{2.35}$$

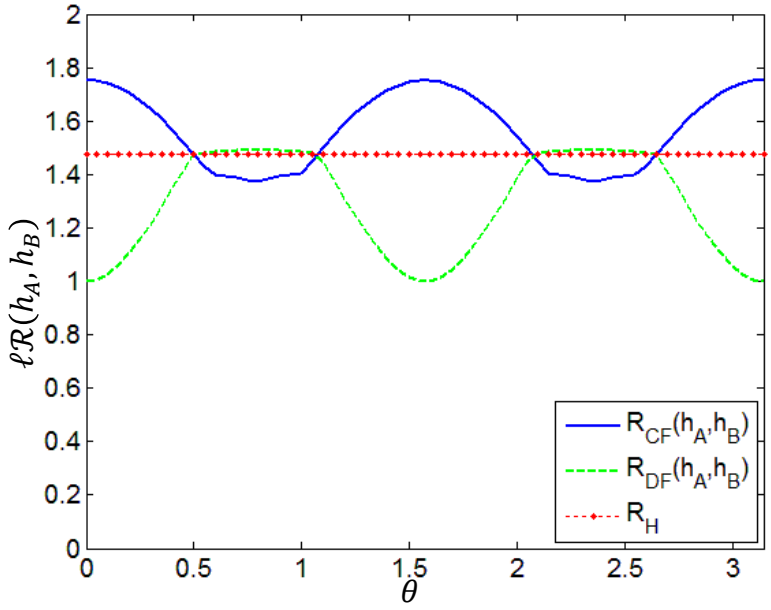
where $\theta_A, \theta_B \in \{0, \frac{\pi}{m}, ..., 2\pi\}$ for a finite integer m and $\theta = \theta_A - \theta_B$. This demonstrates the schemes ability to adapt to changes in the phase difference between h_A and h_B. The universally achievable rates $\mathcal{R}_{\mathcal{H},CF}$, $\mathcal{R}_{\mathcal{H},DF}$, and $\mathcal{R}_{\mathcal{H}}$ for the channel gains (2.35) are plotted as a function of SNR in Fig. 8(d). The universally achievable rate benefits from the ability of the relay to adapt between CF and DF. Particularly, for moderate SNRs we see that $\mathcal{R}_{\mathcal{H}} > \max(\mathcal{R}_{\mathcal{H},CF}, \mathcal{R}_{\mathcal{H},DF})$. This is because CF is limited for θ near $\{\frac{\pi}{4}, \frac{3\pi}{4}\}$, and DF is limited for θ near $\{0, \frac{\pi}{2}\}$. This is depicted in Fig. 8(b).

II.4.1.4. Performance of QPSK for $h_A = 1$ and Arbitrary h_B

In Fig. 9, an image of the achievable rates with the proposed scheme is shown when $h_A = 1$ and $h_B = \Re\{h_B\} + j\Im\{h_B\}$ for various $\Re\{h_B\}$ and $\Im\{h_B\}$ for $snr = 7$ dB and

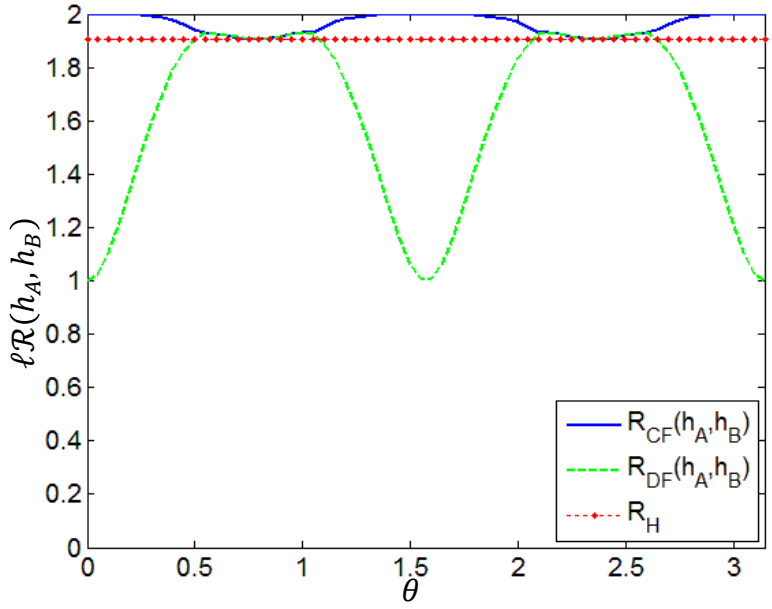


(a) $\ell\mathcal{R}$ vs. θ for $snr = 0$ dB

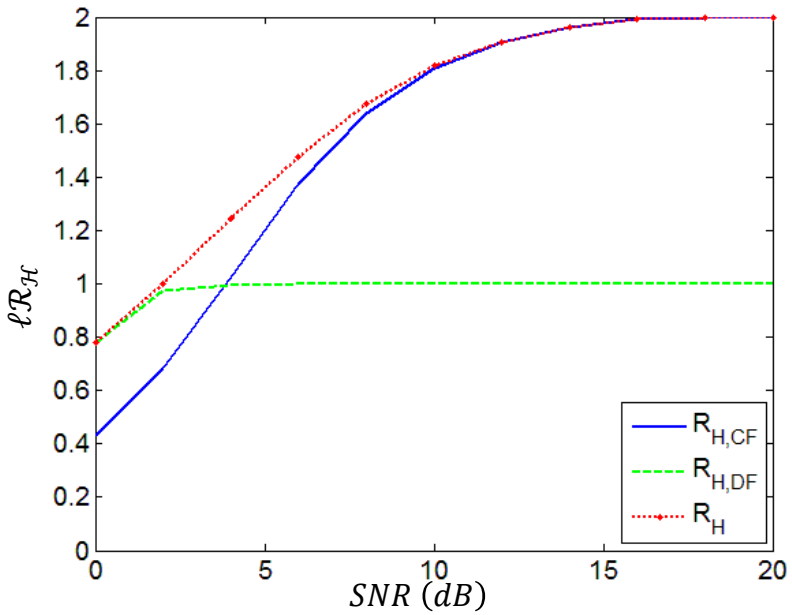


(b) $\ell\mathcal{R}$ vs. θ for $snr = 6$ dB

Fig. 8.: Achievable rates for the proposed scheme when the decoder adaptively chooses between CF and DF decoding.

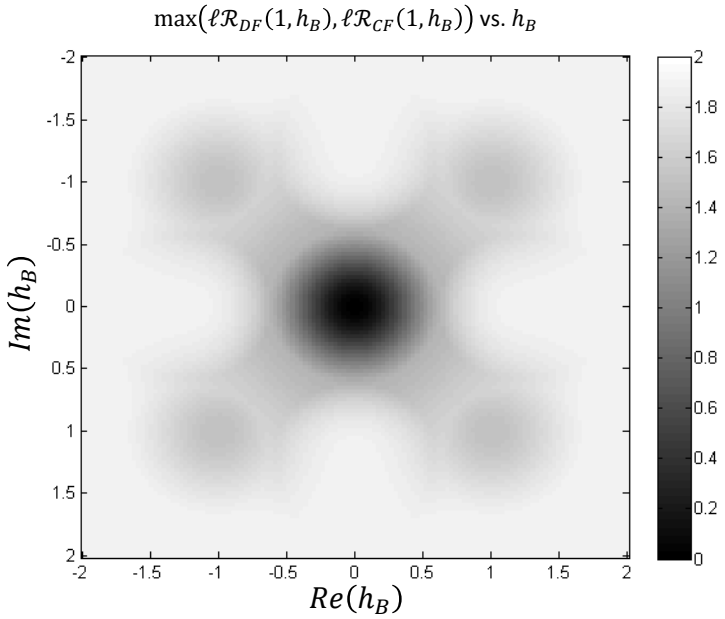


(c) $\ell\mathcal{R}$ vs. θ for $snr = 12$ dB

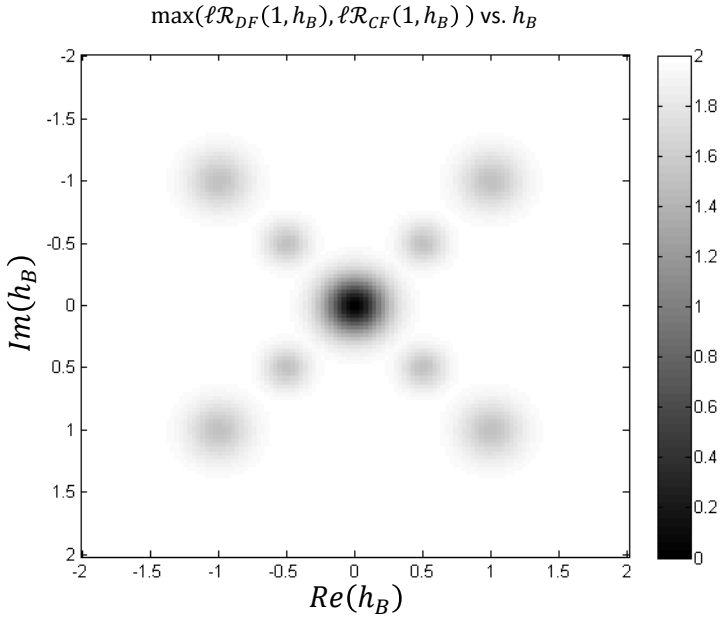


(d) $\ell\mathcal{R}_{\mathcal{H},CF}$, $\ell\mathcal{R}_{\mathcal{H},DF}$, and $\ell\mathcal{R}_{\mathcal{H}}$ vs. SNR

Fig. 8.: Fig. 8: Continued.



(a) Best achievable rate vs. h_B for $snr = 7$ dB



(b) Best achievable rate vs. h_B for $snr = 15$ dB

Fig. 9.: Achievable rates for the proposed scheme with 4-qam when $h_A = 1$ as a function of h_B.

$snr = 15$ dB for the 4-qam signal constellation. It can be seen from the plots that the achievable rates are close to the maximum for a large region of $(\Re\{h_B\}, \Im\{h_B\})$. It can be also seen that the achievable rate is limited to 1.5 bits per symbol for channel gains such that $\frac{h_A}{h_B} \in \{\sqrt{2}e^{j(\frac{\pi}{4}+m\frac{\pi}{2})}, \frac{1}{\sqrt{2}}e^{j(\frac{\pi}{4}+m\frac{\pi}{2})}\}$ for any integer m. This has been observed in [28] which proposes specific 5-ary denoising functions for these channel gains. If we want to achieve reliable decoding at the relay, our scheme requires the careful design of signaling constellations $\mathcal{Q}$ and mapping functions $\mathcal{M}$ to avoid such limits at high SNR. A thorough analysis of such designs is beyond the scope of this thesis.

II.4.1.5. Outage Performance

An outage experiment provides a better perspective to understand the effectiveness of the proposed scheme for a more realistic channel model. Consider a two way relay channel with block fading at the multiple access stage. For this experiment, the channel coefficients h_A and h_B are complex Gaussian random variables with zero mean and unit variance. They are assumed to be independent of each other and from block to block. In our experiment, we test $18,000$ randomly selected channel gain pairs (h_A, h_B) and record the achievable symmetric exchange rate for the MLC encoder with either CF or DF. Our proposed MLC scheme achieves the maximum of CF and DF for each channel gain. For each (h_A, h_B) we also record two upper bounds on the achievable rates based on a trivial cut-set upper bound. We consider the cuts between node A and the relay and between node B and the relay. For each cut, the interference from the other user can be assumed to be perfectly known at the relay. If node A and B each use a Gaussian codebook, the resulting power limited

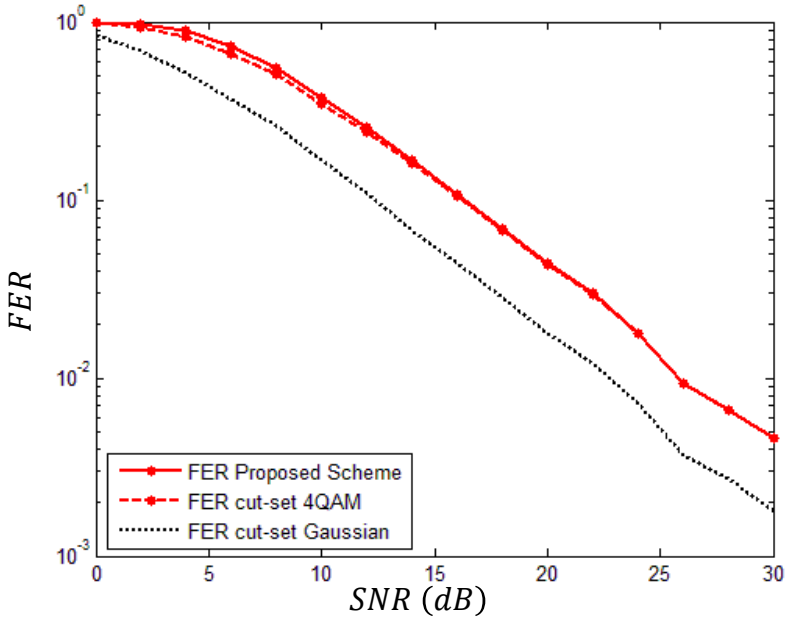
upper bound is expressed

$$\mathcal{R} < \min(\log(1 + |h_A|^2 snr), \log(1 + |h_B|^2 snr)). \tag{2.36}$$

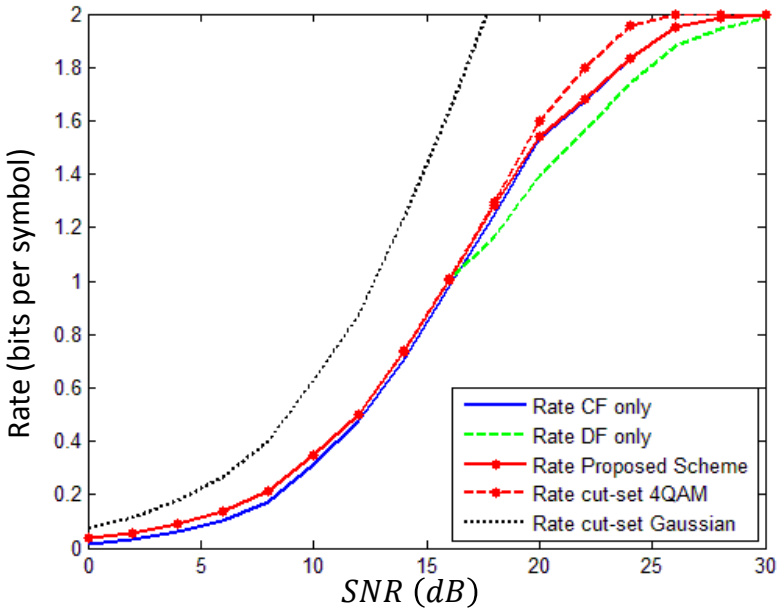
If node A and B each use 4-qam signalling with optimal coding, the resulting cut-set bound is similarly the minimum of the achievable rates for the point-to-point channel between node A and the relay or between node B and the relay. The results of our experiment are plotted in Fig. 10.

In Fig. 10(a), the frame error rate (FER) is plotted as a function of SNR for the outer bounds and the proposed scheme if the code rates are selected to achieve 1.5 bits per symbol spectral efficiency. For the MLC scheme, each code level uses the same linear code of rate $\mathcal{R} = 0.75$. The performance is limited by the weakest channel gain $\min\{|h_A|, |h_B|\}$. Note that the FER for the proposed scheme is very close to the limit imposed by the signalling constellation. This indicates that the proposed scheme handles interference between the users very well for this code rate.

In Fig. 10(b), we fix the acceptable FER as 0.05 and plot the achievable rate as a function of SNR for the outer bounds and for the proposed scheme MLC scheme. We also plot the results for the MLC encoder with only CF or DF decoding or choosing whichever decoder is best as proposed. It is interesting that our performance is quite close to the limit imposed by the signal constellation until there is marked divergence at $\ell\mathcal{R} = 1.5$. This suggests that the distance shortening events identified in Fig. 9 only begin to affect the outage performance at these higher code rates. This contrasts well with the case where the decoder is limited to use DF decoding because any distance shortening events should affect the achievable rate. With DF decoding only, the outage performance is tight for all rates below $\ell\mathcal{R} = 1$. Again, a rigorous design of signalling constellations is beyond the scope of this thesis. However, the relationship between the unavoidable distance shortening events and the outage performance for



(a) FER vs. SNR for 4-qam with 1.5 bit per symbol code rate



(b) Rate (bits per symbol) vs. SNR for 4-qam with a FER of 0.05

Fig. 10.: Outage experiment results with 4-qam signalling.

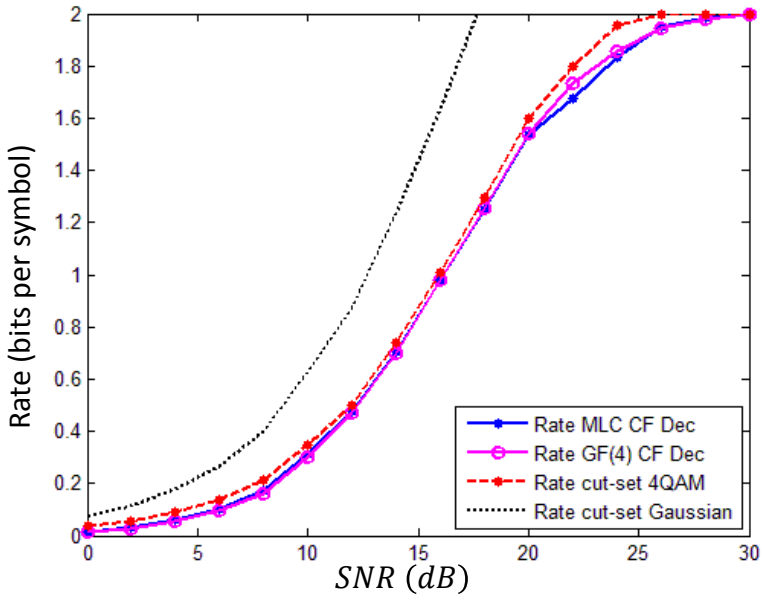
different code rates should have interesting implications for such designs.

In Fig. 11, we compare the outage performance for the proposed MLC scheme with the outage performance if nodes A and B use identical linear codebooks over $GF(4)$. As in Fig. 10(b), we fix the acceptable FER as 0.05, and the achievable rates for both schemes are plotted as a function of SNR. We have previously demonstrated in Figs. 6 and 7 that the proposed MLC scheme facilitates a larger class of decoding functions for CF decoding. It is therefore surprising that coding over $GF(4)$ with CF decoding has better outage performance at high SNR than the proposed MLC scheme with CF decoding. This can be explained by the fact that for coding over $GF(4)$ with CF decoding, only one rate constraint needs to be satisfied to achieve reliable decoding (2.34). For MLC with CF decoding, three additional constraints must be satisfied (2.23) including one penalty constraint associated with the use of identical linear codebooks for each signalling level.

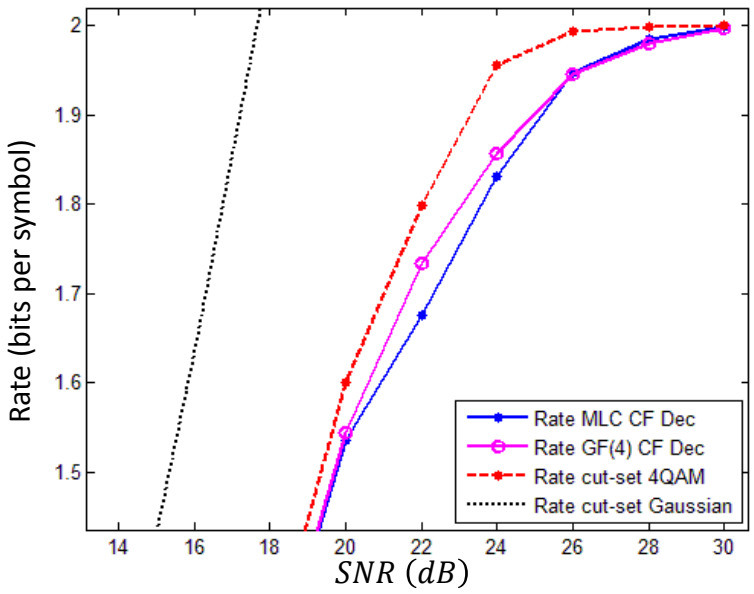
II.4.2. Numerical Results for 8-ary Constellations

We now provide additional illustrative examples using three different 8-ary modulation schemes which are shown in Fig. 12. This means that we use MLC with $\ell = 3$ bit levels, and the scheme supports 168 unique decoding functions for CF decoding at the relay (one for each invertible $\ell \times \ell$ binary matrix).

For the 8-qam constellation, we plot the performance of CF and DF decoding as a function of θ for $h_A = 1$, $h_B = e^{j\theta}$, and $snr = 12$ dB in Fig. 13(a). The universally achievable rate $\mathcal{R}_{\mathcal{H}}$ is plotted as a function of SNR in Fig. 13(b). We note that the high SNR performance is limited for both CF and DF for values of θ near $\{0, \pi\}$. There are many constellation points with unequal labels and nearly the same location in $\mathcal{Q}_R$ for these channel gains. However, in Fig. 13(b), we see that at least one of the CF decoding functions facilitates a 0.5 bits per symbol improvement

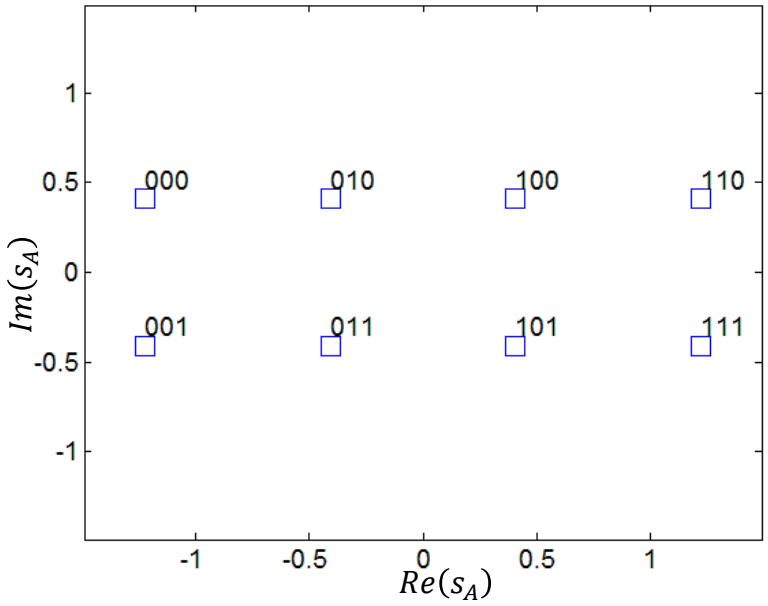


(a) Rate (bits per symbol) vs. SNR for 4-qam with a FER of 0.05

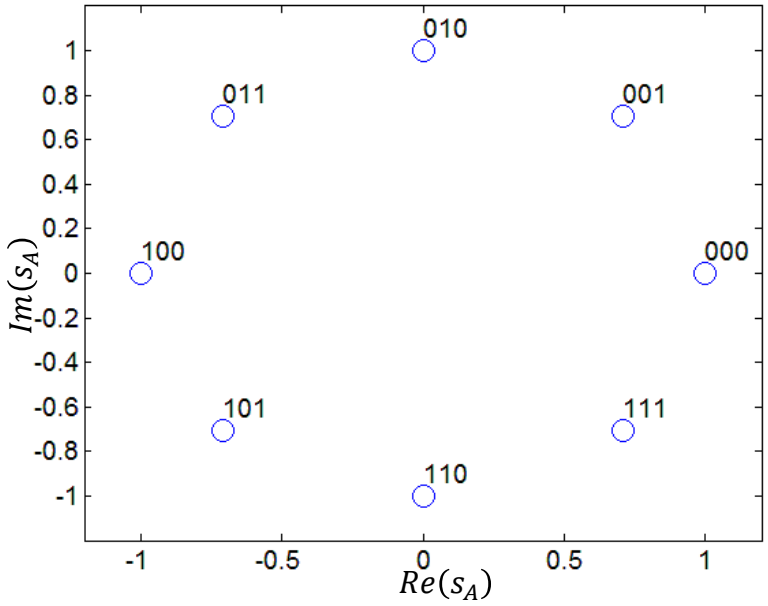


(b) Rate (bits per symbol) vs. SNR for 4-qam with a FER of 0.05 at high SNR

Fig. 11.: Outage experiment results with 4-qam signalling. Proposed multilevel coding scheme compared to coding over GF(4), both with CF decoding.

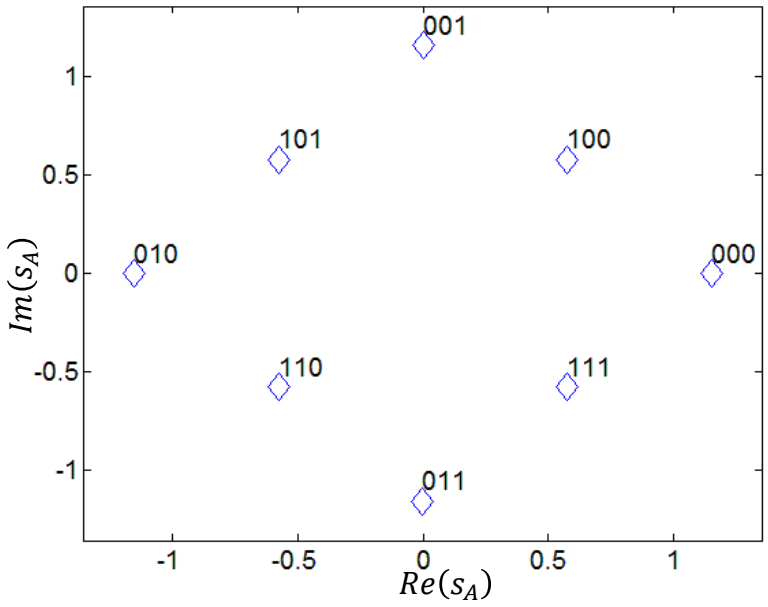


(a) 8-qam signaling constellation



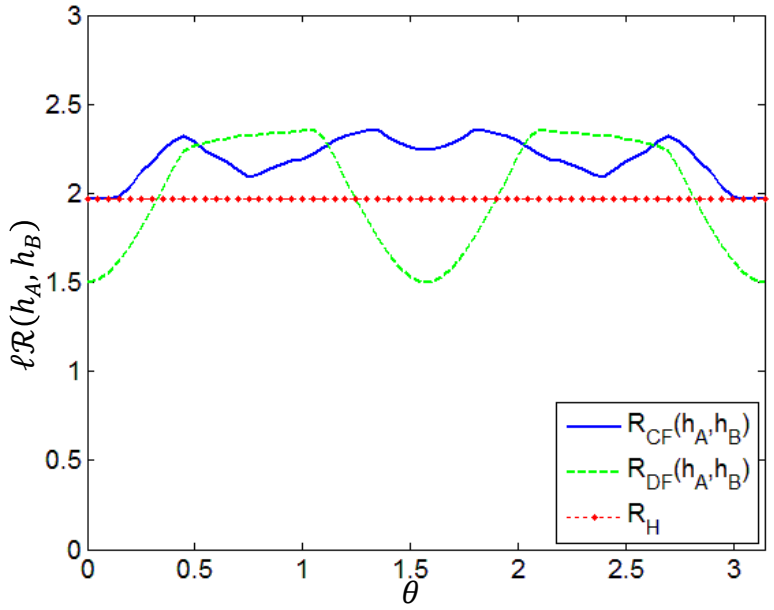
(b) 8-psk signaling constellation

Fig. 12.: 8-ary modulation schemes considered in this section.

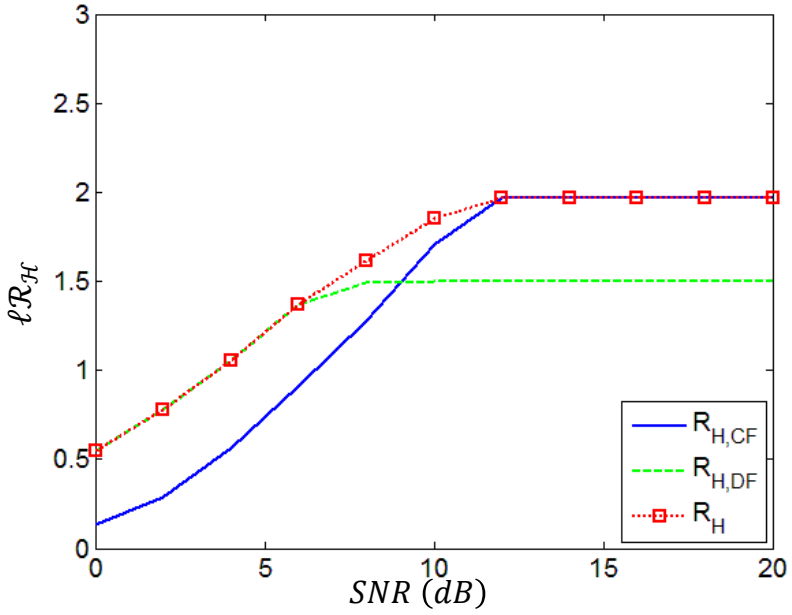


(c) 8-box signaling constellation

Fig. 12.: Fig. 12: Continued.



(a) $\ell\mathcal{R}$ vs. θ for $snr = 12$ dB with 8-qam signaling.



(b) $\ell\mathcal{R}_{\mathcal{H},CF}$, $\ell\mathcal{R}_{\mathcal{H},DF}$, and $\ell\mathcal{R}_{\mathcal{H}}$ vs. SNR with 8-qam signaling.

Fig. 13.: Achievable information rates for the proposed scheme with 8-qam signalling.

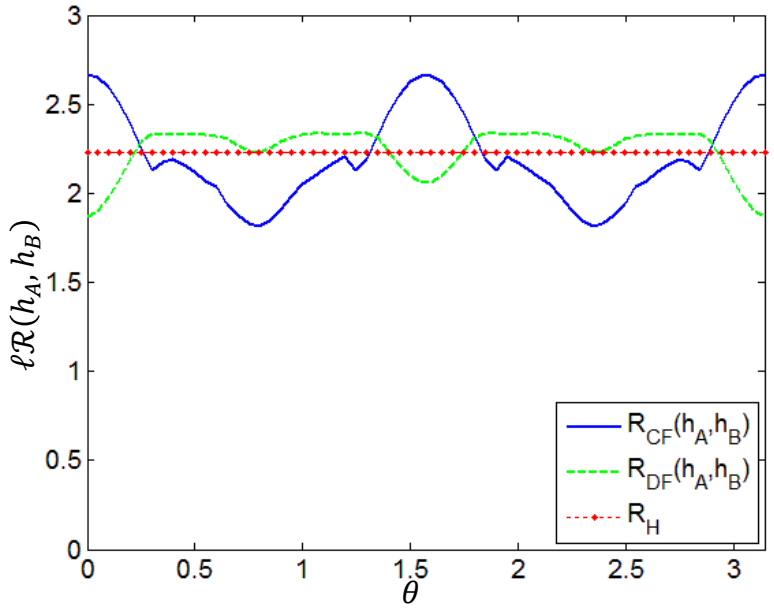
of CF over DF at high SNR, for this constellation.

For the 8-psk constellation, the achievable rate for CF and DF decoding is plotted as a function of θ for $snr = 12$ dB in Fig. 14(a). The universally achievable rate $\mathcal{R}_{\mathcal{H}}$ is plotted as a function of SNR in Fig. 14(b). The performance of the 8-psk signaling constellation are particularly interesting because $\mathcal{R}_{\mathcal{H}} > \max(\mathcal{R}_{\mathcal{H},CF}, \mathcal{R}_{\mathcal{H},DF})$ at high SNR. This is because the limiting values of θ for the tested channel parameters are different for the two decoding methods.

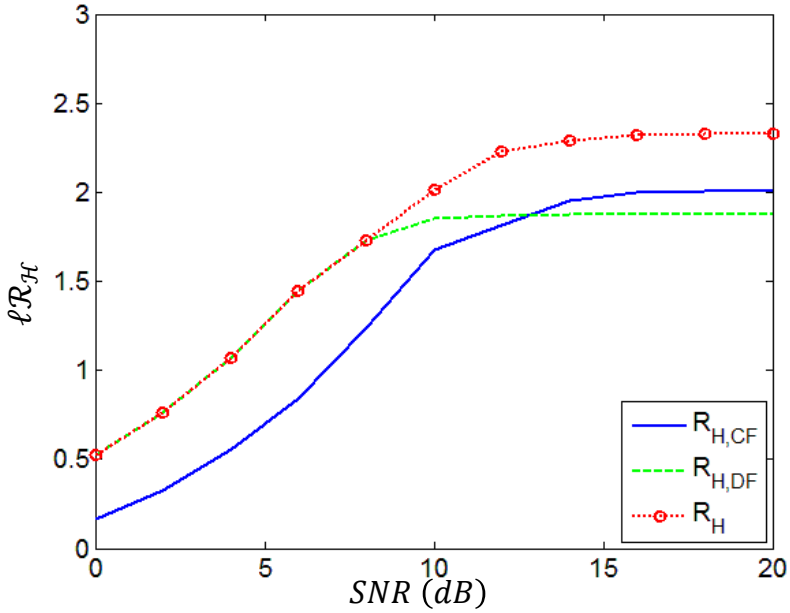
Lastly, we consider the performance of a signalling constellation called 8-box shown in Fig. 12(c). The 8-box constellation is effectively two QPSK constellations with the second constellation rotated by $\frac{\pi}{4}$ and scaled by $\sqrt{2}$. For 8-box, the achievable rate for CF and DF decoding is plotted as a function of θ for $snr = 12$ dB in Fig. 15(a). The universally achievable rate $\mathcal{R}_{\mathcal{H}}$ is plotted as a function of SNR in Fig. 15(b). We see that the $\mathcal{R}_{\mathcal{H}}$ for the 8-box constellation is the best for all SNRs among the tested 8-ary constellations. This is partially because the constellation is defined by points on a lattice in the complex field which assists CF decoding when $|h_A| = |h_B|$. Note that for each of the signalling constellations considered, the universal performance for DF is consistently better at low SNR while the universal performance for CF is better at high SNR. Further, the universally achievable rates are improved by the use of both CF and DF decoders.

II.4.3. Comparison to Upper Bounds and Lattice based Compute-and-Forward

To thoroughly understand the performance of the proposed scheme, it is helpful to compare the results for the given constellations to some upper bounds. For the case when the channel is not known at the transmitter, tight upper bounds on the achievable rates are not known, making this comparison difficult. Nevertheless, a

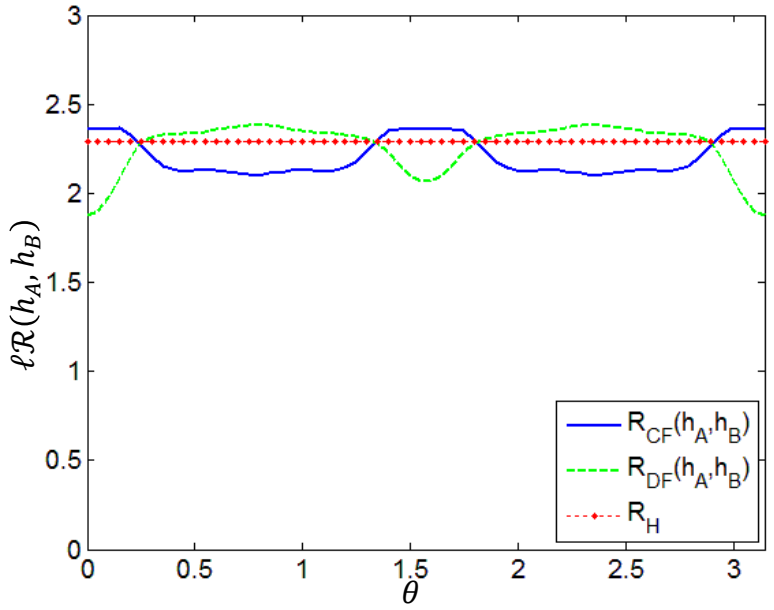


(a) $\ell\mathcal{R}$ vs. θ for $snr = 12$ dB with 8-psk signaling.

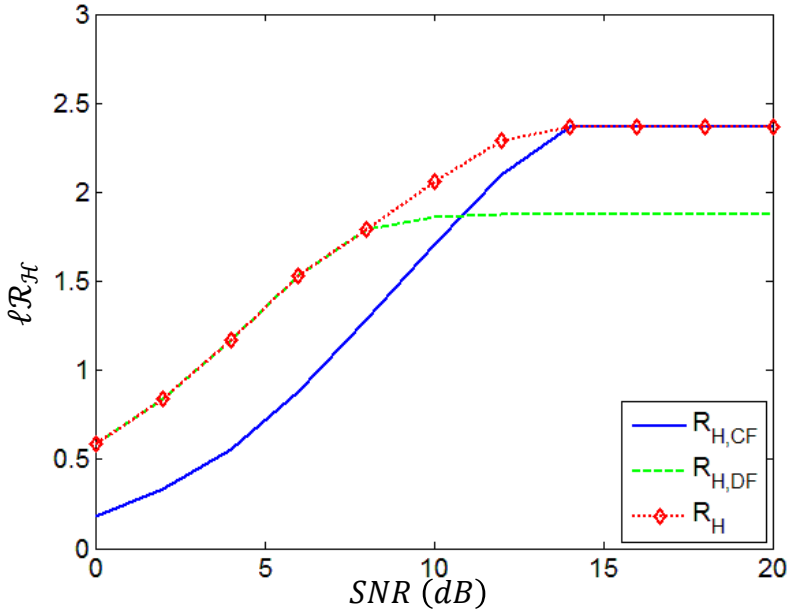


(b) $\ell\mathcal{R}_{\mathcal{H},CF}$, $\ell\mathcal{R}_{\mathcal{H},DF}$, and $\ell\mathcal{R}_{\mathcal{H}}$ vs. SNR with 8-psk signaling.

Fig. 14.: Achievable information rates for the proposed scheme with 8-psk signalling.



(a) $\ell\mathcal{R}$ vs. θ for $snr = 12$ dB with 8-box signaling.



(b) $\ell\mathcal{R}_{\mathcal{H}}$, $\ell\mathcal{R}_{\mathcal{H},DF}$, and $\ell\mathcal{R}_{\mathcal{H},CF}$ vs. SNR with 8-box signaling.

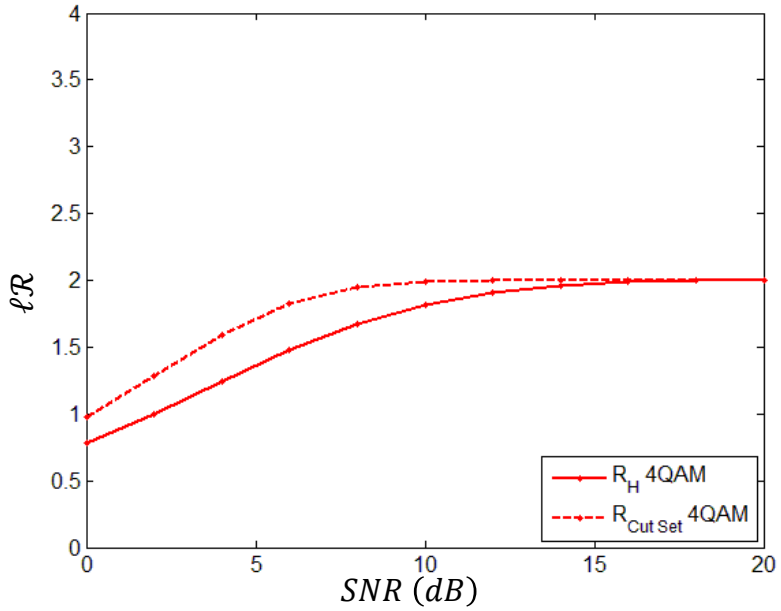
Fig. 15.: Achievable information rates for the proposed scheme with 8-box signalling.

trivial upper bound can be obtained using a cut-set bound. This bound is obtained by considering a cut between node A and the relay and assuming that the interference from the other user is perfectly available at the relay. Further, when computing the information rate for this cut, the specific modulation format used by the nodes is considered. We will refer to this bound as the cut-set bound with known interference. This bound and the performance of the proposed scheme are shown in Fig. 16 for the 4 different modulation formats considered. While there is a gap between the performance of the proposed scheme and the upper bounds, the proposed scheme is asymptotically optimal in the 4-qam case (for the case of $h_A = 1, h_B = e^{j\theta}$) and performs within about 0.5 bits/channel use for the 8-box case.

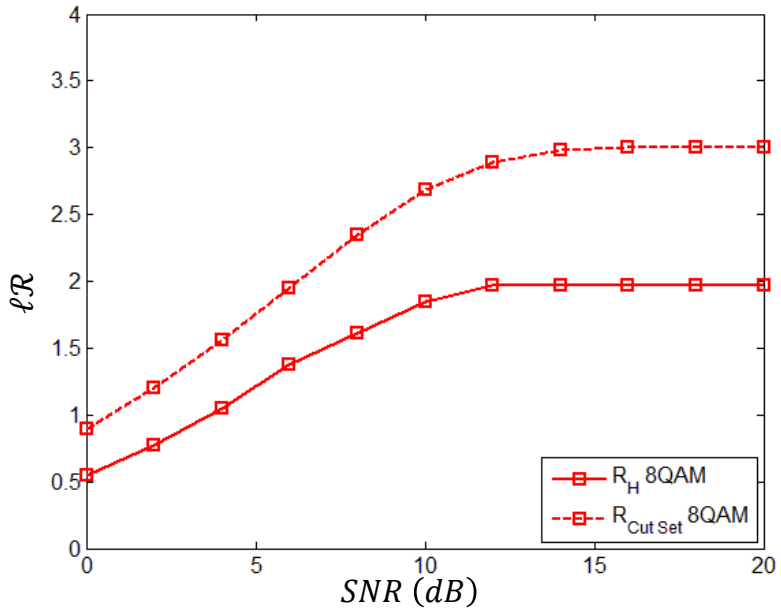
In [26], Nazer and Gastpar derive achievable rates for a compute-and-forward scheme using infinite dimensional lattice codes when channel state information is not available at the transmitter. In Fig. 17, the universally achievable rate with the proposed scheme is compared to that achievable by Nazer and Gastpar's scheme for the $h_A = 1$ and $h_B = e^{j\theta}$ case. It can be seen that the proposed scheme with 4-qam and 8-box modulation schemes outperform the lattice code for low and moderate SNRs for the tested channel gains. At higher SNRs, the performance of the proposed schemes are limited by the cardinality of the 4-qam and 8-box constellations and the lattice based compute-and-forward scheme outperforms the proposed schemes.

II.4.4. Simulation Results

To corroborate our theoretical results, we simulated the performance of a regular (3,6) low density parity check (LDPC) code with message passing decoding. In Fig. 18 the required SNR is plotted as a function of θ for the case where $h_A = 1$ and $h_B = e^{j\theta}$. The theoretical results assume the linear code $\mathcal{C}$ used at each bit level has rate $\mathcal{R} = \frac{1}{2}$. Therefore, the theoretically required SNR for CF refers to the SNR such

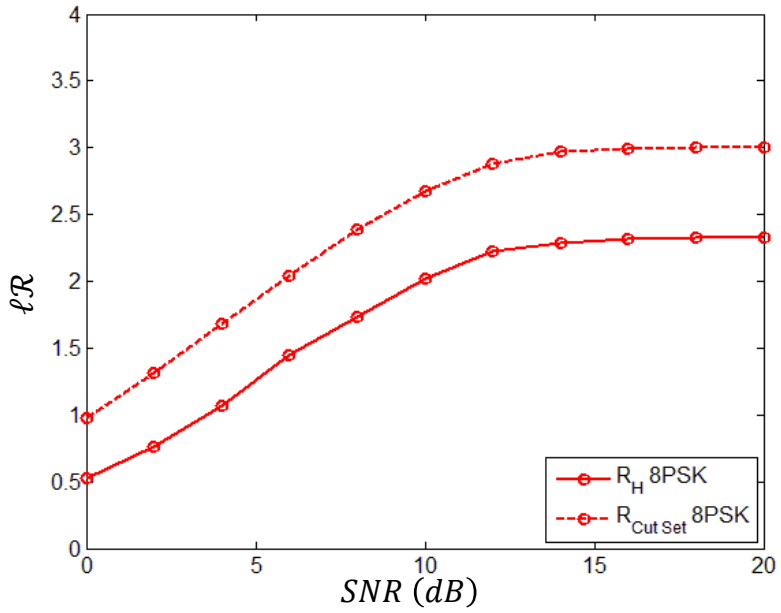


(a) $\ell\mathcal{R}$ vs. SNR with 4-qam signaling.

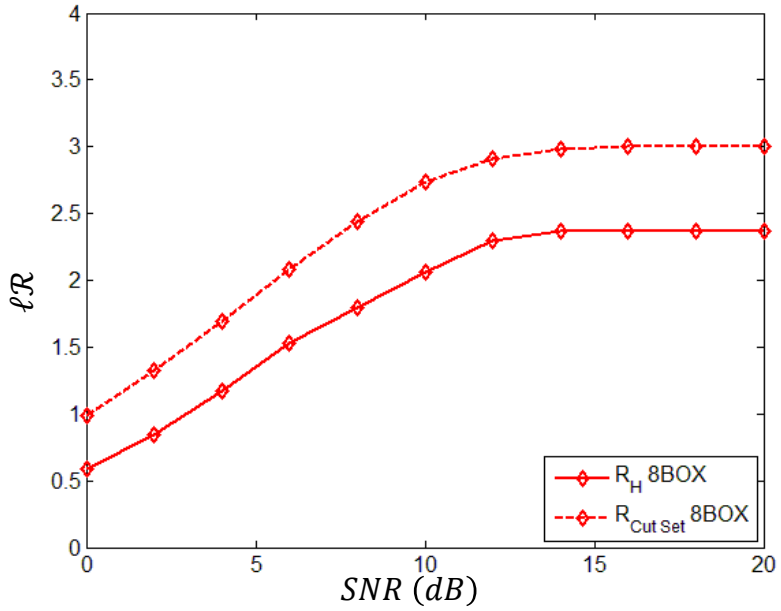


(b) $\ell\mathcal{R}$ vs. SNR with 8-qam signaling.

Fig. 16.: Universally achievable rates with the proposed scheme compared to the cut-set bound with known interference.



(c) $\ell\mathcal{R}$ vs. SNR with 8-psk signaling.



(d) $\ell\mathcal{R}$ vs. SNR with 8-box signaling.

Fig. 16.: Fig. 16.: Continued.

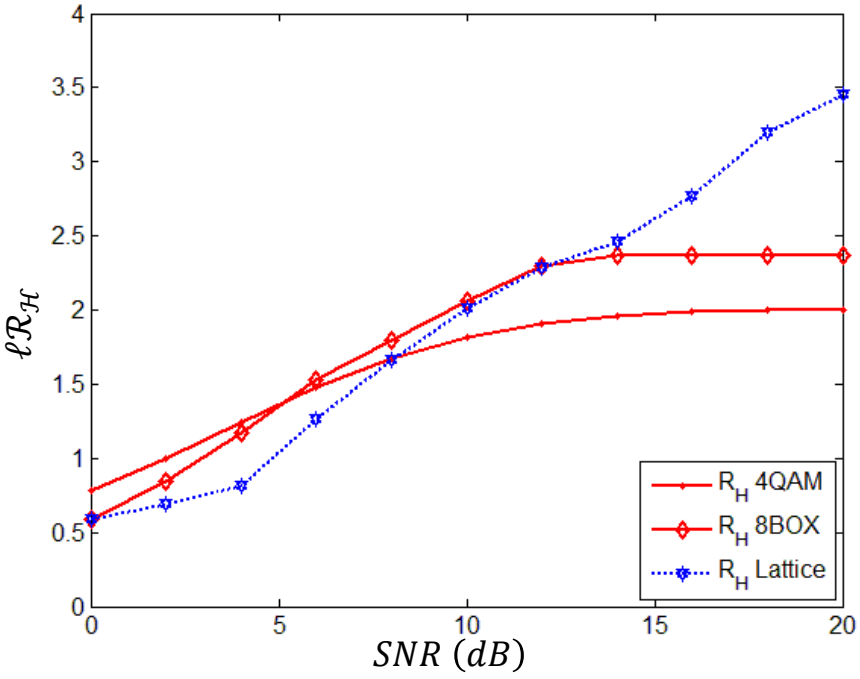


Fig. 17.: $\ell\mathcal{R}_{\mathcal{H}}$ vs. SNR for 4-qam and 8-box signaling compared to $\mathcal{R}_{\mathcal{H}}$ for lattice based compute-and-forward.

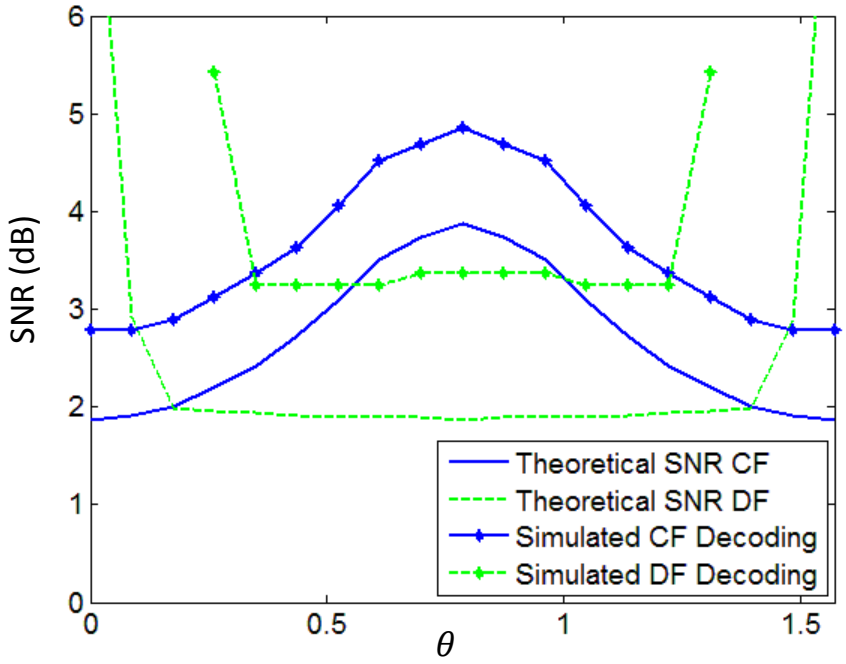


Fig. 18.: Required SNR (dB) vs. θ to reliably decode a rate $\ell\mathcal{R} = 1$ code using the proposed scheme MLC scheme. The theoretical results are compared to simulation results for a long $(3, 6)$ LDPC code.

that $\mathbf{X}_{f,R}$ can be reliably decoded for at least one $f \in \mathcal{F}$. The theoretically required SNR for DF refers to the SNR such that $\mathbf{X}_{AB}$ can be reliably decoded. Any function can be subsequently computed. For the simulated results, we simulated up to 200 transmissions of a (3,6) LDPC codeword of length $100,000$ for each tested SNR and θ. The simulated curves therefore represent the SNR for which there were zero bit errors after 200 simulated transmissions. This had to be satisfied for at least one $f \in \mathcal{F}$ for CF decoding. The messages from both node A and B had to be recovered without error for DF decoding.

For a point-to-point AWGN channel using binary phase shift keying (BPSK), it has been shown in [38] that the required SNR for a (3,6) LDPC code with iterative

decoding is about 1 dB away from the Shannon limit for BPSK signaling and AWGN. The simulated SNR are approximately 0.85 to 1.5 dB larger than the theoretically required SNR for the tested values of θ for the best of CF and DF which suggests that the achievable rates discussed in this chapter can be approached with practical coding schemes.

Note that to achieve the theoretical limit imposed by (2.19) using structured codes, it will be necessary to design practical coding schemes which universally achieve capacity for many channel conditions. The conventional approach to achieving capacity with a message passing decoder has been to optimize the code ensemble depending on the channel conditions. However, it has recently been discovered that, for certain classes of spatially coupled LDPC codes, a message passing decoder can achieve the performance of an optimal decoder for the underlying code ensemble [39], [40]. This has been shown to hold universally for a large class of binary-input channels [41] including the binary-input MAC channel with AWGN and fading [42], [43]. For these reasons, it appears that the class of spatially coupled LDPC codes may allow us to achieve the information rates derived in this chapter with practical decoding complexity. This is investigated for a simplified channel model in Chapter IV.

II.5. Concluding Remarks

In this chapter, we have proposed a multilevel coding scheme which is more practical and facilitates improved decoding flexibility over previous schemes based on infinite dimensional lattice codes. Our scheme can be implemented with practical binary linear codebooks and modulation schemes. Our analysis of the achievable rates with the proposed scheme reveals that the identical linear codebooks for each signalling

level imposes penalty rate constraints. Specifically, our construction introduces a class of error events which are not present with independent codebooks. A detailed proof is provided in the following Appendix II.6. Our numerical results for show that decoding flexibility can be significantly improved if the relay attempts direct computation of the desired functions (called the CF decoder) or attempts to decode all incoming codewords (called the DF decoder) depending on the channel. This fact, along with the notion that the achievable computation rate is the suitable metric for system performance, should have important implications for the design of better signalling constellations.

A logical step for further research might be to investigate ways to optimize the signaling constellations for reliable PLNC. We discuss some ways we think this should be approached with future work in Chapter V. However, as we have found in this chapter, there are still things we do not know about the limits for reliable decoding when the transmitters use structured code ensembles suitable for joint channel and network coding. Particularly, we do not have a converse result to match the achievability proof in this chapter. We address this for a general channel model with identical binary linear codebooks in Chapter III. Then, we consider how to design practical coding schemes to achieve our theoretical performance in Chapter IV. These results are consolidated and concluded in Chapter V.

II.6. Appendix: Proof of Theorem II.1

II.6.1. Additional Notation

It will be notationally convenient to refer to variables associated with different messages according to a single integer index. Therefore, we define the integer j as the binary expansion of the message $\mathbf{U}_j$. Recall that $\mathbf{U}_j$ is an $\ell \times K$ binary matrix, so

$j \in \{0, ..., 2^{K\ell} - 1\}$. We use $\mathbf{U}_t$ to refer to the true message sent from the transmitter.

The relay will attempt to reliably decode $\mathbf{X}_t$ from $\underline{y}$ using a joint typicality decoder. Thus an error occurs if either $\mathbf{X}_t$ is not jointly typical with $\underline{y}$ or if some incorrect message $\mathbf{X}_j,\ j \neq t$ is jointly typical with $\underline{y}$.

II.6.2. Pairwise Independence of Codewords

Here we provide a brief analysis of the ensemble of coset codes used by the transmitter. Both of the following lemmas appear as part of the proof of Gallager's Coding Theorem for Random Parity Check Codes [44]. We include these proofs because the intuition behind some of the steps is used for other parts of the proof.

Lemma II.2. *Let each element of* $\mathbf{G}$ *and* $\underline{\lambda}^k$ *be i.i.d. Bernoulli random variables with parameter* $\frac{1}{2}$*. Then we have*

$$P(\underline{V}_j^k = \underline{v}_j^k) = \frac{1}{2^N} \ \forall \ \underline{v}_j^k \in \mathbb{F}_2^N. \tag{2.37}$$

That is, the codeword $\underline{v}_j^k \in \mathbb{F}_2^N$ *associated with message vector* $\underline{u}_j^k \in \mathbb{F}_2^K$ *can take any value with uniform probability over the ensemble of random coset codes.*

Proof. For a fixed $\mathbf{G}$ and $\underline{u}_j^k$, the output of the linear encoder $\underline{\gamma}_j^k = \underline{u}_j^k \mathbf{G}$ must take some value in $\mathbb{F}_2^N$. Since $\underline{\lambda}^k$ can take any value with equal probability we have

$$P(\underline{V}_j^k = \underline{v}_j^k) = P(\underline{\Lambda}^k = \underline{v}_j^k \oplus \underline{\gamma}_j^k) = \frac{1}{2^N} \ \forall \ \underline{v}_j^k \in \mathbb{F}_2^N. \tag{2.38}$$

□

Lemma II.3. *Let each element of* $\mathbf{G}$ *and* $\underline{\lambda}^k$ *be i.i.d. Bernoulli random variables*

with parameter $\frac{1}{2}$*. Then for any* $j' \neq j$ *such that* $\underline{u}_j^k \neq \underline{u}_{j'}^k$*, we have*

$$P(\underline{V}_j^k = \underline{v}_j^k, \underline{V}_{j'}^k = \underline{v}_{j'}^k) = P(\underline{V}_j^k = \underline{v}_j^k)P(\underline{V}_{j'}^k = \underline{v}_{j'}^k)$$
$$= \frac{1}{2^{2N}} \ \forall \ \underline{v}_j^k, \underline{v}_{j'}^k \in \mathbb{F}_2^N. \tag{2.39}$$

Particularly, the codewords $\underline{v}_j^k$ *and* $\underline{v}_{j'}^k$ *associated with* $\underline{u}_j^k$ *and* $\underline{u}_{j'}^k$ *respectively are independent and uniformly distributed over* $\mathbb{F}_2^N$*.*

Proof. Suppose that $\underline{u}_j^k$ and $\underline{u}_{j'}^k$ differ in position m, and let $\underline{g}_i, \ i \in \{1, ..., K\}$ refer to the i_{th} row of $\mathbf{G}$. Then for any set of rows

$$\underline{g}_1, ..., \underline{g}_{m-1}, \underline{g}_{m+1}, ..., \underline{g}_K$$

there is some $\underline{g}_m$ which gives $\underline{v}_j^k \oplus \underline{v}_{j'}^k = \underline{\gamma}_j^k \oplus \underline{\gamma}_{j'}^k$ any fixed value. By the construction of $\mathbf{G}$ and Lemma II.2, $\underline{g}_m$ and $\underline{v}_j^k$ can take any value with uniform probability. We can conclude that

$$\begin{aligned} &P(\underline{V}_j^k = \underline{v}_j^k, \underline{V}_{j'}^k = \underline{v}_{j'}^k | \underline{u}_j^k \neq \underline{u}_{j'}^k) \\ &= P(\underline{V}_j^k = \underline{v}_j^k | \underline{u}_j^k \neq \underline{u}_{j'}^k) P(\underline{V}_{j'}^k = \underline{v}_{j'}^k | \underline{v}_j^k, \underline{u}_j^k \neq \underline{u}_{j'}^k) \\ &= P(\underline{\Lambda}^k = \underline{v}_j^k \oplus \underline{\gamma}_j^k | \underline{u}_j^k \neq \underline{u}_{j'}^k) P(\underline{G}_m = \underline{v}_{j'}^k \oplus \underline{v}_j^k | \underline{v}_j^k, \underline{u}_j^k \neq \underline{u}_{j'}^k) \\ &= \frac{1}{2^N} \frac{1}{2^N} = \frac{1}{2^{2N}} \ \forall \ \underline{v}_j^k, \underline{v}_{j'}^k \in \mathbb{F}_2^N. \end{aligned} \tag{2.40}$$

□

The key idea behind each proof is the same. In Lemma II.2, we see that the uniform distribution of $\underline{\lambda}^k$ implies the uniform distribution of $\underline{v}_j^k$. In Lemma II.3, we see that the uniform distribution of $\mathbf{G}$ implies the independence of codewords corresponding to distinct messages.

II.6.3. Analysis of Error Probability

The relay uses a joint typicality decoder to decode $\mathbf{X}_t$ from $\underline{y}$. For some fixed $\epsilon > 0$, define $\mathcal{A}_\epsilon^N$ as the set of $(\mathbf{X}_j, \underline{y})$ pairs which satisfy the definition of joint typicality given in [13]. The set $\mathcal{A}_\epsilon^N$ is referred to as the jointly typical set. Let the event $E_j,\ j \in \{0, ..., 2^{K\ell} - 1\}$ be the event $(\mathbf{X}_j, \underline{y}) \in \mathcal{A}_\epsilon^N$. The probability of error given that the codeword corresponding to t is observed by the receiver can be expressed

$$P(Err|t) = P\left(\overline{E}_t \cup \bigcup_{j \in \{0,...,2^{K\ell}-1\} \setminus \{t\}} E_j \middle| t\right) \tag{2.41}$$

Applying the union bound, we get

$$P(Err|t) \leq P\left(\overline{E}_t|t\right) + \sum_{j \in \{0,...,2^{K\ell}-1\} \setminus \{t\}} P(E_j|t). \tag{2.42}$$

Recall that $\underline{y} \sim P(\underline{Y}|\mathbf{X}_t)$. Thus by the joint asymptotic equipartition property (AEP) we have that for any $\epsilon > 0$,

$$P(\overline{E}_t|t) < \epsilon \tag{2.43}$$

for sufficiently large N.

The proof of the channel coding theorem for the general discrete memoryless channel in [13] relies on upper bounding $P(E_j|t)$ using the joint AEP. This is not straightforward here because $\mathbf{X}_j,\ j \neq t$ and $\underline{y}$ are not independent with the same marginals for certain classes of error events. For example, if $\ell = 2$, we could have

$$\begin{aligned} \mathbf{U}_j &= \mathbf{U}_t \oplus \begin{bmatrix} \underline{e}_u \\ \underline{0} \end{bmatrix} \\ \Rightarrow \mathbf{X}_j &= \mathbf{X}_t \oplus \begin{bmatrix} \underline{e}_v \\ \underline{0} \end{bmatrix} \end{aligned} \tag{2.44}$$

for some $\underline{e}_u \in \mathbb{F}_2^K \setminus \{0\}$ and $\underline{e}_v \in \mathbb{F}_2^N \setminus \{0\}$. This means that $\underline{v}_j^1 \neq \underline{v}_t^1$ but $\underline{v}_j^2 = \underline{v}_t^2$. Thus for this class of error events $\mathbf{X}_j$ and $\mathbf{X}_t$ are not independent. Note that this class of error events is handled by the proof of the coding theorem for the multiple access channel [37], [13], and [45]. In the coding theorem proof for the multiple access channel, it is possible for the receiver to correctly decode a codeword from one transmitter while making an error in decoding the codeword from a second transmitter. This has the same effect as correctly decoding the codeword on one level of a multilevel encoder while making an error in decoding the codeword transmitted on the second level.

Unfortunately, choosing to use a coset of the *same* linear codes at each bit level introduces a new class of error events of the form

$$\begin{aligned} \mathbf{U}_j &= \mathbf{U}_t \oplus \begin{bmatrix} \underline{e}_u \\ \underline{e}_u \end{bmatrix} \\ \Rightarrow \mathbf{X}_j &= \mathbf{X}_t \oplus \begin{bmatrix} \underline{e}_v \\ \underline{e}_v \end{bmatrix}. \end{aligned} \tag{2.45}$$

For this class of error events, the columns of the error matrix $\mathbf{X}_j \oplus \mathbf{X}_t$ must be in $\{[0\ 0]^T, [1\ 1]^T\}$. *This is the key difference between our proof and the proofs for the general multiple access channel or for the point-to-point channel with multilevel coding using independent encoders for each level.*

We can move forward by splitting the sum in (2.42) into different events for which $\mathbf{X}_j$ and $\mathbf{X}_t$ are conditionally independent. Define a set of $p \leq \ell$ disjoint subsets $\mathcal{S}_1, ..., \mathcal{S}_p \subseteq \{1, ..., \ell\}$. Let τ_i be the smallest element of $\mathcal{S}_i,\ i \in \{1, ..., p\}$, and define the sets $\mathcal{T} = \{\tau_1, ..., \tau_p\}$, $\mathcal{S} = \mathcal{S}_1 \cup ... \cup \mathcal{S}_p$, and $\overline{\mathcal{S}} = \{1, ..., \ell\} \setminus \mathcal{S}$. For each set

of subsets, define an index set $\mathcal{J}_{\mathcal{S}_1,...,\mathcal{S}_p}$ given by

$$\mathcal{J}_{\mathcal{S}_1,...,\mathcal{S}_p} = \left\{ j \left| \underline{u}_j^k = \left\{ \begin{array}{ll} \underline{u}_t^k \oplus \underline{e}_{u,1} & , k \in \mathcal{S}_1 \\ & \vdots \\ \underline{u}_t^k \oplus \underline{e}_{u,p} & , k \in \mathcal{S}_p \\ \underline{u}_t^k & , k \in \overline{\mathcal{S}} \end{array} \right. \right\}. \tag{2.46}$$

Here each message error vector, $\underline{e}_{u,i} \in \mathbb{F}_2^K \setminus \{\underline{0}\} \ \forall \ i \in \{1, ..., p\}$ satisfies $\underline{e}_{u,i} \neq \underline{e}_{u,i'} \ \forall \ i \neq i'$. For clarification, a subset $\mathcal{S}_i$ gives the levels of an incorrect message $\mathbf{U}_j$ which differ from $\mathbf{U}_t$ with the an equal message error vector $\underline{e}_{u,i}$. Thus the disjoint families of subsets $\{\mathcal{S}_1, ..., \mathcal{S}_p\} \in \mathcal{Z}_\ell$ allow us to describe all classes of error events similar to (2.45) for general ℓ. Recall that $\mathcal{Z}_\ell$ is the set of disjoint families of subsets of $\{1, ..., \ell\}$. Each index set $\mathcal{J}_{\mathcal{S}_1,...,\mathcal{S}_p} \subseteq \{0, ..., 2^{K\ell} - 1\}$ contains the integer indices of all incorrect messages $\mathbf{U}_j$ which differ from $\mathbf{U}_t$ as indicated by the subsets $\mathcal{S}_1, ..., \mathcal{S}_p$. This notation may seem redundant, however, it allows us to simplify many expressions through the remainder of the proof. For the sake of clarity, we will complete the analysis of error probability for the case where $\ell = 2$, and then extend the results to a general ℓ.

Case $\{\ell = 2\}$*:* If $\ell = 2$, the subsets in (2.46) can be written as

$$\mathcal{J}_{\{1\}} = \left\{ j \,\middle|\, \mathbf{U}_j = \mathbf{U}_t \oplus \begin{bmatrix} \underline{e}_{u,1} \\ \underline{0} \end{bmatrix} \right\}$$

$$\mathcal{J}_{\{2\}} = \left\{ j \,\middle|\, \mathbf{U}_j = \mathbf{U}_t \oplus \begin{bmatrix} \underline{0} \\ \underline{e}_{u,1} \end{bmatrix} \right\}$$

$$\mathcal{J}_{\{1,2\}} = \left\{ j \,\middle|\, \mathbf{U}_j = \mathbf{U}_t \oplus \begin{bmatrix} \underline{e}_{u,1} \\ \underline{e}_{u,1} \end{bmatrix} \right\}$$

$$\mathcal{J}_{\{1\}\{2\}} = \left\{ j \,\middle|\, \mathbf{U}_j = \mathbf{U}_t \oplus \begin{bmatrix} \underline{e}_{u,1} \\ \underline{e}_{u,2} \end{bmatrix} \right\}.$$

These subsets are disjoint and describe each class of error events, so

$$\mathcal{J}_{\{1\}} \cup \mathcal{J}_{\{2\}} \cup \mathcal{J}_{\{1,2\}} \cup \mathcal{J}_{\{1\}\{2\}} = \{0, ..., 2^{2K} - 1\} \setminus \{t\}.$$

Therefore, the union bound on the probability of error for $\ell = 2$ can be written as

$$\begin{aligned} P(Err|t) \leq & P(\overline{E}_t|t) + \sum_{j \in \mathcal{J}_{\{1\}}} P(E_j|t, j \in \mathcal{J}_{\{1\}}) \\ & + \sum_{j \in \mathcal{J}_{\{2\}}} P(E_j|t, j \in \mathcal{J}_{\{2\}}) \\ & + \sum_{j \in \mathcal{J}_{\{1,2\}}} P(E_j|t, j \in \mathcal{J}_{\{1,2\}}) \\ & + \sum_{j \in \mathcal{J}_{\{1\}\{2\}}} P(E_j|t, j \in \mathcal{J}_{\{1\}\{2\}}). \end{aligned} \tag{2.47}$$

We define $\underline{e}_{v,i} = \underline{e}_{u,i}\mathbf{G},\ i \in \{1, 2\}$ as the codeword error vector associated with subset $\mathcal{S}_i$. The subscript u or v is used to differentiate between the message error vector and codeword error vector respectively. Over the ensemble of codes, each $\underline{e}_{v,i}$ is uniformly distributed in $\mathbb{F}_2^N$, and codeword error vectors $\underline{e}_{v,i}, \underline{e}_{v,j},\ i \neq j$ are

independent. These facts can be shown using steps similar to the proofs of Lemmas II.2 and II.3.

If $\{j \in \mathcal{J}_{\{1\}}\}$ is given, then we know that

$$\mathbf{X}_j = \mathbf{X}_t \oplus \begin{bmatrix} \underline{e}_{v,1} \\ \underline{0} \end{bmatrix}.$$

By Lemmas II.2 and II.3, $\underline{v}_t^1$ and $\underline{v}_j^1$ are independent and uniformly distributed on $\mathbb{F}_2^N$. Lemma II.2 also tells us that $\underline{v}_t^2$ and $\underline{v}_j^2$ are equal and uniformly distributed on $\mathbb{F}_2^N$.

Define $\underline{v}^2$ as the common value taken by $\underline{v}_j^2 = \underline{v}_t^2$. The joint AEP provides an asymptotically tight upper bound to each $P(E_j|t, j \in \mathcal{J}_{\{1\}})$ if we can show that

$$P(\mathbf{X}_j, \underline{Y}|t, j \in \mathcal{J}_{\{1\}}) = P(\mathbf{X}_j|t, j \in \mathcal{J}_{\{1\}})P(\underline{Y}|t, j \in \mathcal{J}_{\{1\}}). \tag{2.48}$$

This is equivalent to showing that

$$P(\mathbf{X}_j, \underline{Y}|\underline{v}^2, j \in \mathcal{J}_{\{1\}}) = P(\mathbf{X}_j|\underline{v}^2, j \in \mathcal{J}_{\{1\}})P(\underline{Y}|\underline{v}^2, j \in \mathcal{J}_{\{1\}}). \tag{2.49}$$

for each value of $\underline{v}^2$. Therefore, consider some arbitrary fixed $\underline{v}^2$. We can use the definition of conditional probability to get

$$\begin{aligned} &P(\mathbf{X}_j, \underline{Y}|\underline{v}^2, j \in \mathcal{J}_{\{1\}}) \\ &\quad = P(\mathbf{X}_j|\underline{v}^2, j \in \mathcal{J}_{\{1\}})P(\underline{Y}|\mathbf{X}_j, \underline{v}^2, j \in \mathcal{J}_{\{1\}}) \\ &\quad = P(\mathbf{X}_j|\underline{v}^2, j \in \mathcal{J}_{\{1\}})P(\underline{Y}|\underline{v}_j^1, \underline{v}^2, j \in \mathcal{J}_{\{1\}}). \end{aligned} \tag{2.50}$$

Since $\underline{y} \sim P(\underline{Y}|\mathbf{X}_t) = P(\underline{Y}|\underline{v}_t^1, \underline{v}^2)$, we can see that $\underline{y}$ is a random function of $\underline{v}_t^1$ conditioned on $\underline{v}^2$. This is expressed as

$$\underline{y} = g(\underline{v}_t^1; \underline{v}^2). \tag{2.51}$$

Since $j \in \mathcal{J}_{\{1\}}$, $\underline{v}_t^1$ and $\underline{v}_j^1$ are independent according to Lemma II.3. Therefore, we have

$$P(\underline{Y}|\underline{v}_j^1, \underline{v}^2, j \in \mathcal{J}_{\{1\}}) = P(\underline{Y}|\underline{v}^2, j \in \mathcal{J}_{\{1\}}). \tag{2.52}$$

This allows us to conclude that (2.49) holds so we can use [13, Theorem 15.2.3] to get the following bound

$$P(E_j|t, j \in \mathcal{J}_{\{1\}}) < 2^{-N(I(Y;X^1|X^2)-3\epsilon)}. \tag{2.53}$$

Similar steps can be used for the case when $j \in \mathcal{J}_{\{2\}}$ to get

$$P(E_j|t, j \in \mathcal{J}_{\{2\}}) < 2^{-N(I(Y;X^2|X^1)-3\epsilon)}. \tag{2.54}$$

For the case when $j \in \mathcal{J}_{\{1,2\}}$, we have

$$\mathbf{X}_j = \mathbf{X}_t \oplus \begin{bmatrix} \underline{e}_{v,1} \\ \underline{e}_{v,1} \end{bmatrix}.$$

The most direct way to find a bound for this case is to reassign the address vectors so that this case is similar to the case when $j \in \mathcal{J}_{\{1\}}$. Define a binary matrix $\mathbf{\Delta}_{\{1,2\}}$ given by

$$\mathbf{\Delta}_{\{1,2\}} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}. \tag{2.55}$$

Define effective codeword matrices $\widetilde{\mathbf{X}}_t$ and $\widetilde{\mathbf{X}}_j$ by

$$\widetilde{\mathbf{X}}_t = \mathbf{\Delta}_{\{1,2\}}\mathbf{X}_t$$

$$\widetilde{\mathbf{X}}_j = \mathbf{\Delta}_{\{1,2\}}\mathbf{X}_j = \widetilde{\mathbf{X}}_t \oplus \begin{bmatrix} \underline{e}_{v,1} \\ \underline{0} \end{bmatrix}.$$

This is the same as the case where $j \in \mathcal{J}_{\{1\}}$ if the relay observes the $\underline{\widetilde{y}}$ corresponding to codeword matrix, $\widetilde{\mathbf{X}}_t$ through the effective channel $P(\underline{\widetilde{Y}}|\mathbf{\Delta}_{\{1,2\}}\underline{X}_t)$. Therefore for

the case where $j \in \mathcal{J}_{\{1,2\}}$, we have the bound

$$P(E_j|t, j \in \mathcal{J}_{\{1,2\}}) < 2^{-N(I(Y;\widetilde{X}^1|\widetilde{X}^2)-3\epsilon)} \tag{2.56}$$

which can be expressed in terms of the original address variables as

$$P(E_j|t, j \in \mathcal{J}_{\{1,2\}}) < 2^{-N(I(Y;X^1|X^1\oplus X^2)-3\epsilon)}. \tag{2.57}$$

By the definition of mutual information, we have

$$\begin{aligned} &I(Y; X^1|X^1 \oplus X^2) \\ &\quad = H(Y|X^1 \oplus X^2) - H(Y|X^1, X^1 \oplus X^2) \\ &\quad = H(Y|X^1 \oplus X^2) - H(Y|X^1, X^2) \\ &\quad = I(Y; X^1, X^2|X^1 \oplus X^2). \end{aligned}$$

Therefore the bound is equivalent to

$$P(E_j|t, j \in \mathcal{J}_{\{1,2\}}) < 2^{-N(I(Y;X^1,X^2|X^1\oplus X^2)-3\epsilon)}. \tag{2.58}$$

Lastly, for the case when $j \in \mathcal{J}_{\{1\}\{2\}}$, $\mathbf{X}_t$ and $\mathbf{X}_j$ are i.i.d. by Lemmas II.2 and II.3. We can therefore use joint AEP directly to get the bound

$$P(E_j|t, j \in \mathcal{J}_{\{1\}\{2\}}) < 2^{-N(I(Y;X^1,X^2)-3\epsilon)}. \tag{2.59}$$

Applying the upper bounds for each index set to (2.47), we get the following

bound

$$\begin{aligned} P(Err|t) < \epsilon &+ \sum_{j\in\mathcal{J}_{\{1\}}} 2^{-N(I(Y;X^1|X^2)-3\epsilon)} \\ &+ \sum_{j\in\mathcal{J}_{\{2\}}} 2^{-N(I(Y;X^2|X^1)-3\epsilon)} \\ &+ \sum_{j\in\mathcal{J}_{\{1,2\}}} 2^{-N(I(Y;X^1,X^2|X^1\oplus X^2)-3\epsilon)} \\ &+ \sum_{j\in\mathcal{J}_{\{1\}\{2\}}} 2^{-N(I(Y;X^1,X^2)-3\epsilon)}. \end{aligned} \tag{2.60}$$

There are $2^{N\mathcal{R}}-1$ elements in the sets $\mathcal{J}_{\{1\}}, \mathcal{J}_{\{2\}}$, and $\mathcal{J}_{\{1,2\}}$, and there are fewer than $2^{2N\mathcal{R}}$ elements in the last set $\mathcal{J}_{\{1\}\{2\}}$. Thus the upper bound on the probability of error for this code ensemble can be expressed

$$\begin{aligned} P(Err|t) <\epsilon &+ 2^{N(\mathcal{R}-I(Y;X^1|X^2)+3\epsilon)} \\ &+ 2^{N(\mathcal{R}-I(Y;X^2|X^1)+3\epsilon)} \\ &+ 2^{N(\mathcal{R}-I(Y;X^1,X^2|X^1\oplus X^2)+3\epsilon)} \\ &+ 2^{N(2\mathcal{R}-I(Y;X^1,X^2)+3\epsilon)}. \end{aligned} \tag{2.61}$$

Each of these terms can be made arbitrarily close to zero by increasing N as long as $\mathcal{R}$ satisfies

$$\mathcal{R} < \max(I(Y;X^1|X^2), I(Y;X^2|X^1), I(Y;X^1,X^2|X^1\oplus X^2), \frac{1}{2}I(Y;X^1,X^2)). \tag{2.62}$$

Note that this proof holds for an arbitrary t which means that the bound holds independent of the transmitted message.

Case $\{\ell \geq 2\}$*:* For a general ℓ, the proof is very similar. We split (2.42) into the

disjoint classes of error events in (2.46) to get

$$P(Err|t) \leq P(\overline{E}_t|t) + \sum_{p=1}^{\ell} \sum_{\{\mathcal{S}_1,...,\mathcal{S}_p\}\mathcal{Z}_\ell} \sum_{j \in \mathcal{J}_{\mathcal{S}_1,...,\mathcal{S}_p}} P(E_j|t, j \in \mathcal{J}_{\mathcal{S}_1,...,\mathcal{S}_p}). \tag{2.63}$$

Then we find upper bounds on the probability of error for different classes of error events.

First, we consider the case where each $\mathcal{S}_i,\ i \in \{1, ..., p\}$ contains only its smallest element τ_m. This is analogous to the case where $\{j \in \mathcal{J}_{\{1\}}\}$ for the proof when $\ell = 2$. By Lemmas II.2 and II.3, we have

$$P(\underline{V}_j^k = \underline{v}_1, \underline{V}_t^k = \underline{v}_2 | j \in \mathcal{J}_{\mathcal{S}_1,...,\mathcal{S}_p}) = \begin{cases} 2^{-2N} & ,\ k \in \mathcal{S} \\ 2^{-N} & ,\ \underline{v}_1 = \underline{v}_2 \text{ and } k \notin \mathcal{S} \\ 0 & ,\ \underline{v}_1 \neq \underline{v}_2 \text{ and } k \notin \mathcal{S}. \end{cases} \tag{2.64}$$

That is if $k \in \mathcal{S}$ then $\underline{v}_j^k$ and $\underline{v}_t^k$ are independent and uniformly distributed. If $k \notin \mathcal{S}$ they are equal and uniformly distributed. Let $\underline{v}^k,\ k \notin \mathcal{S}$ be the common value taken by the k_{th} row of $\mathbf{X}_j$ and $\mathbf{X}_t$.

The joint AEP gives an asymptotically tight upper bound to $P(E_j|t, j \in \mathcal{J}_{\mathcal{S}_1,...,\mathcal{S}_p})$ if we can show that

$$P(\mathbf{X}_j, \underline{Y}|t, j \in \mathcal{J}_{\mathcal{S}_1,...,\mathcal{S}_p}) = P(\mathbf{X}_j|t, j \in \mathcal{J}_{\mathcal{S}_1,...,\mathcal{S}_p}) P(\underline{Y}|t, j \in \mathcal{J}_{\mathcal{S}_1,...,\mathcal{S}_p}). \tag{2.65}$$

This is equivalent to showing that

$$\begin{aligned} &P(\mathbf{X}_j, \underline{Y}|\{\underline{v}^k, k \notin \mathcal{S}\}, j \in \mathcal{J}_{\mathcal{S}_1,...,\mathcal{S}_p}) \\ &\quad = P(\mathbf{X}_j|\{\underline{v}^k, k \notin \mathcal{S}\}, j \in \mathcal{J}_{\mathcal{S}_1,...,\mathcal{S}_p}) P(\underline{Y}|\{\underline{v}^k, k \notin \mathcal{S}\}, j \in \mathcal{J}_{\mathcal{S}_1,...,\mathcal{S}_p}) \end{aligned} \tag{2.66}$$

for each possible set of values $\{\underline{v}^k, k \notin \mathcal{S}\}$. This is possible using the same arguments as for the $\ell = 2$ case. Particularly, if we condition on $\{\underline{v}^k, k \notin \mathcal{S}\}, j \in \mathcal{J}_{\mathcal{S}_1,...,\mathcal{S}_p}$, then

$\underline{y} \sim P(\underline{Y}|\mathbf{X}_t) = P(\underline{Y}|\{\underline{v}_t^k, k \in \mathcal{S}\}, \{\underline{v}^k, k \notin \mathcal{S}\})$. Since $j \in \mathcal{J}_{\mathcal{S}_1,...,\mathcal{S}_p}$, we have that $\underline{v}_j^k$ and $\underline{v}_t^k$ are independent for every $k \in \mathcal{S}$. Therefore, $\underline{y}$ is conditionally a random function of $\{\underline{v}_t^k, k \in \mathcal{S}\}$ and we can conclude that

$$\begin{aligned}
&P(\mathbf{X}_j, \underline{Y}|\{\underline{v}^k\ , k \notin \mathcal{S}\}, j \in \mathcal{J}_{\mathcal{S}_1,...,\mathcal{S}_p}) \\
&\quad = P(\mathbf{X}_j|\{\underline{v}^k\ , k \notin \mathcal{S}\}, j \in \mathcal{J}_{\mathcal{S}_1,...,\mathcal{S}_p}) P(\underline{Y}|\mathbf{X}_j, \{\underline{v}^k\ , k \notin \mathcal{S}\}, j \in \mathcal{J}_{\mathcal{S}_1,...,\mathcal{S}_p}) \\
&\quad = P(\mathbf{X}_j|\{\underline{v}^k\ , k \notin \mathcal{S}\}, j \in \mathcal{J}_{\mathcal{S}_1,...,\mathcal{S}_p}) P(\underline{Y}|\{\underline{v}_j^k, k \in \mathcal{S}\}, \{\underline{v}^k\ , k \notin \mathcal{S}\}, j \in \mathcal{J}_{\mathcal{S}_1,...,\mathcal{S}_p}) \\
&\quad = P(\mathbf{X}_j|\{\underline{v}^k\ , k \notin \mathcal{S}\}, j \in \mathcal{J}_{\mathcal{S}_1,...,\mathcal{S}_p}) P(\underline{Y}|\{\underline{v}^k\ , k \notin \mathcal{S}\}, j \in \mathcal{J}_{\mathcal{S}_1,...,\mathcal{S}_p}) \qquad (2.67)
\end{aligned}$$

Therefore (2.65) holds which allows us to apply joint AEP to get the upper bound

$$P(E_j|t, j \in \mathcal{J}_{\mathcal{S}_1,...,\mathcal{S}_p}) \leq 2^{-N(I(Y;\{X^k, k\in\mathcal{S}\}|\{X^k, k\notin\mathcal{S}\})-3\epsilon)}. \qquad (2.68)$$

To extend this result to the general case where each $\mathcal{S}_1, ..., \mathcal{S}_p$ can contain multiple elements, we make this problem look like the first case. Define a matrix $\mathbf{\Delta}_{\mathcal{S}_1,...,\mathcal{S}_p}$ whose m_{th} column $\underline{d}_m$ is given by

$$\underline{d}_{\tau_k}[n] = \begin{cases} 1\ ,\ n \in \mathcal{S}_k \\ 0\ ,\ n \notin \mathcal{S}_k \end{cases} \quad \forall\ k = 1, ..., p$$

$$\underline{d}_m[n] = \begin{cases} 1\ ,\ n = m \\ 0\ ,\ n \neq m \end{cases} \quad \forall\ m \notin \mathcal{T}. \qquad (2.69)$$

For example, if $\ell = 6$, $\mathcal{S}_1 = \{2, 4, 5\}$, and $\mathcal{S}_2 = \{3, 6\}$ we have

$$\mathbf{D}_{\{2,4,5\},\{3,6\}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}.$$

Define effective codeword matrices $\widetilde{\mathbf{X}}_j$ and $\widetilde{\mathbf{X}}_t$ by

$$\begin{aligned} \widetilde{\mathbf{X}}_j &= \mathbf{D}_{\mathcal{S}_1,...,\mathcal{S}_p} \mathbf{X}_j \\ \widetilde{\mathbf{X}}_t &= \mathbf{D}_{\mathcal{S}_1,...,\mathcal{S}_p} \mathbf{X}_t. \end{aligned} \tag{2.70}$$

Then the k_{th} row $\underline{\widetilde{v}}_j^k$ of $\widetilde{\mathbf{X}}_j$ is given by

$$\underline{\widetilde{v}}_j^k = \begin{cases} \underline{\widetilde{v}}_t^k & , \ k \notin \mathcal{T} \\ \underline{\widetilde{v}}_t^k \oplus \underline{e}_{v,m} & , \ k = \tau_m, \ m = 1, ..., p \end{cases} \tag{2.71}$$

for some pairwise independent set of error vectors $\underline{e}_{v,1}, ..., \underline{e}_{v,p} \in \mathbb{F}_2^N \setminus \{\underline{0}\}$.

For the $\ell = 6$ example, this means that

$$\widetilde{\mathbf{X}}_j = \widetilde{\mathbf{X}}_t \oplus \begin{bmatrix} \underline{0} \\ \underline{e}_{1,v} \\ \underline{e}_{2,v} \\ \underline{0} \\ \underline{0} \\ \underline{0} \end{bmatrix}.$$

This is the same as the case where each $\mathcal{S}_1, ..., \mathcal{S}_p$ contains only one element. Thus,

we can apply the bound in (2.68) to get

$$P(E_j|t, j \in \mathcal{J}_{\mathcal{S}_1,...,\mathcal{S}_p}) \leq 2^{-N(I(Y;\{\widetilde{X}^k,\ k\in\mathcal{T}\}|\{\widetilde{X}^k,\ k\notin\mathcal{T}\})-3\epsilon)}. \tag{2.72}$$

The only step that remains is to show that the mutual information in (2.72) can be expressed as

$$\begin{aligned} &I(Y;\{\widetilde{X}^k,\ k\in\mathcal{T}\}|\{\widetilde{X}^k,\ k\notin\mathcal{T}\}) \\ &\quad = I(Y;\{X^k|k\in\mathcal{S}\}|\{X^k|k\notin\mathcal{S}\},\{X^k\oplus Z_i|k\in\mathcal{S}_i\}\ \forall\ i=1,...,p) \end{aligned} \tag{2.73}$$

where $\mathcal{S} = \bigcup_{i=1}^{p}\mathcal{S}_i$, and each Z_i is an auxiliary Bernoulli random variable with parameter $\frac{1}{2}$. By (2.70), we have

$$\widetilde{x}^k = \begin{cases} x^k & ,\ k\in\overline{\mathcal{S}}\cup\mathcal{T} \\ x^k\oplus x^{\tau_m} & ,\ k\in\mathcal{S}_m\setminus\{\tau_m\}. \end{cases} \tag{2.74}$$

We therefore have

$$\begin{aligned} &\{\widetilde{x}^k|k\in\mathcal{T}\} \Leftrightarrow \{x^k|k\in\mathcal{T}\} \\ &\{\widetilde{x}^k|k\notin\mathcal{T}\} \Leftrightarrow \{x^k|k\notin\mathcal{S}\}\cup\bigcup_{m=1}^{p}\{x^k\oplus x^{\tau_m}|k\in\mathcal{S}_m\setminus\{\tau_m\}\}. \end{aligned}$$

The mutual information can therefore be expressed

$$\begin{aligned} &I(Y;\{\widetilde{X}^k|k\in\mathcal{T}\}|\{\widetilde{X}^k|k\notin\mathcal{T}\}) \\ &\quad = I(Y;\{X^k|k\in\mathcal{T}\}|\{X^k|k\notin\mathcal{S}\}\cup\bigcup_{m=1}^{p}\{X^k\oplus X^{\tau_m}|k\in\mathcal{S}_m\setminus\{\tau_m\}\}) \\ &\quad = H(Y|\{X^k|k\notin\mathcal{S}\}\cup\bigcup_{m=1}^{p}\{X^k\oplus X^{\tau_m}|k\in\mathcal{S}_m\setminus\{\tau_m\}\}) - H(Y|X^1,...,X^\ell). \end{aligned}$$

The last equality follows because if we know x^{τ_m} and $x^k\oplus x^{\tau_m}$ then we know both x^{τ_m} and x^k. Which tells us that knowing $\{x^k|k\in\mathcal{T}\}\cup\bigcup_{m=1}^{p}\{x^k\oplus x^{\tau_m}|k\in\mathcal{S}_m\setminus\{\tau_m\}\}$

is equivalent to knowing $\{x^k | k \in \mathcal{S}\}$.

It can be shown for each $\mathcal{S}_m,\ m \in \{1, ..., p\}$ that

$$\{x^k \oplus x^{\tau_m} | k \in \mathcal{S}_m \setminus \{\tau_m\}\} \Leftrightarrow \{x^k \oplus z_m | k \in \mathcal{S}_m\}. \tag{2.75}$$

For example, if we consider our $\ell = 6$ case, we have $\mathcal{S}_1 = \{2, 4, 5\}$. If we know that

$$(x^4 \oplus x^2, x^5 \oplus x^2) = (a, b)\ ,\ a, b \in \{0, 1\}$$

then we have

$$(x^2, x^4, x^5) \in \{(0, a, b), (1, \overline{a}, \overline{b})\}$$

which is equivalent to knowing

$$(x^2 \oplus z_1, x^4 \oplus z_1, x^5 \oplus z_1).$$

We therefore have

$$\begin{aligned}
&I(Y; \{\widetilde{X}^k | k \in \mathcal{T}\} | \{\widetilde{X}^k | k \notin \mathcal{T}\}) \\
&\quad = H(Y | \{X^k | k \notin \mathcal{S}\} \cup \bigcup_{m=1}^{p} \{X^k \oplus X^{\tau_m} | k \in \mathcal{S}_m \setminus \{\tau_m\}\}) - H(Y | X^1, ..., X^\ell) \\
&\quad = H(Y | \{X^k | k \notin \mathcal{S}\} \cup \bigcup_{m=1}^{p} \{X^k \oplus Z_m | k \in \mathcal{S}_m\}) - H(Y | X^1, ..., X^\ell) \\
&\quad = I(Y; \{X^k | k \in \mathcal{S}\} | \{X^k | k \notin \mathcal{S}\}, \{X^k \oplus Z_i | k \in \mathcal{S}_i\}\ \forall\ i = 1, ..., p).
\end{aligned} \tag{2.76}$$

This is the same as (2.73), which allows us to restate the bound in (2.72) as

$$\begin{aligned}
&P(E_j | t, j \in \mathcal{J}_{\mathcal{S}_1, ..., \mathcal{S}_p}) \\
&\quad \leq 2^{-N(I(Y; \{X^k | k \in \mathcal{S}\} | \{X^k | k \notin \mathcal{S}\}, \{X^k \oplus Z_i | k \in \mathcal{S}_i\}\ \forall\ i = 1, ..., p) - 3\epsilon)} \\
&\quad \triangleq 2^{-N(I(Y; \mathcal{S}_1, ..., \mathcal{S}_p) - 3\epsilon)}.
\end{aligned} \tag{2.77}$$

The last step defines a mutual information $I(Y; \mathcal{S}_1, ..., \mathcal{S}_p)$. This slight abuse of

notation simplifies the last few steps of the proof.

Plugging this into (2.63), we have

$$P(Err|t) \leq P(\overline{E}_t|t) + \sum_{p=1}^{\ell} \sum_{\mathcal{S}_1,...,\mathcal{S}_p} \sum_{j \in \mathcal{J}_{\mathcal{S}_1,...,\mathcal{S}_p}} 2^{-N(I(Y;\mathcal{S}_1,...,\mathcal{S}_p)-3\epsilon)}. \tag{2.78}$$

For each possible $\mathcal{S}_1, ..., \mathcal{S}_p$ we have

$$|\mathcal{J}_{\mathcal{S}_1,...,\mathcal{S}_p}| = (2^{N\mathcal{R}} - 1)(2^{N\mathcal{R}} - 2)...(2^{N\mathcal{R}} - p) < 2^{N\mathcal{R}p}.$$

Therefore we have

$$\begin{aligned} P(Err|t) &\leq \epsilon + \sum_{p=1}^{\ell} \sum_{\mathcal{S}_1,...,\mathcal{S}_p} 2^{N\mathcal{R}p} 2^{-N(I(Y;\mathcal{S}_1,...,\mathcal{S}_p)-3\epsilon)} \\ &\leq \epsilon + \sum_{p=1}^{\ell} \sum_{\mathcal{S}_1,...,\mathcal{S}_p} 2^{N(\mathcal{R}p - I(Y;\mathcal{S}_1,...,\mathcal{S}_p)+3\epsilon)}. \end{aligned}$$

This bound approaches zero as long as

$$\mathcal{R} < \min_{\mathcal{S},\overline{\mathcal{S}},\mathcal{S}_1,...,\mathcal{S}_p} \frac{1}{p} I(Y; \{X^k | k \in \mathcal{S}\} | \{X^k | k \in \overline{\mathcal{S}}\}, \{X^k \oplus Z_i | k \in \mathcal{S}_i\} \ \forall \ i \in \{1, ..., p\}).$$

This completes the proof.

CHAPTER III

JOINT-COMPUTE-AND-FORWARD FOR BINARY MEMORYLESS MULTIPLE ACCESS CHANNELS

In the previous chapter, our proposed multilevel coding scheme with identical linear codes for each bit level practically facilitates generous computation flexibility with reliable decoding at the relay. Our analysis had shown that it is beneficial to allow the relay to attempt both CF and DF decoding depending on the channel parameters. Indeed, either decoding method considered separately is suboptimal for computation with general channel parameters. Therefore a joint decoding paradigm for computation has been proposed in a message passing framework to address this suboptimality [17], [16], [18], [19]. We call the decoding paradigm considered in these papers joint-compute-and-forward (JCF). These authors provide simulation results, based on JCF message passing, which show that JCF can strictly outperform CF or DF decoding for certain channel conditions. In fact, it has been conjectured in [17], that JCF decoding achieve strictly better computation rates than CF and DF for the two way relaying problem.

The main result in this chapter is that, while JCF decoding is rate optimal for computing the finite field sum of transmitted codewords, JCF cannot achieve better computation rates than the better of CF and DF decoding. This surprising negative result verifies that CF and DF provide a fundamental limit for linear computation. This result is proved for the very general class of binary-input memoryless multiple access channels. We obtain the converse by restricting that nodes A and B use the ensemble of independent cosets of an identical binary linear codebook generated uniformly at random. A key implication of this analysis is that, if it possible to achieve better computation rates suitable for transmission over a two way relay

network, it will require clever joint design of the encoders at nodes A and B. Indeed, the conjecture in [17] is requires the design of channel codes such that a computable network function is perfectly matched to the channel parameters. The details of the information-theoretic analysis in this paper should be useful for such an investigation.

Note about notation: In Chapter II, we used two different symbols $\underline{v}$ and $\underline{x}$ to refer to a binary codeword or address vector respectively. This unusual notation enabled us to shift our discussion between the adaptive network coding and reliable channel coding components efficiently. This is not necessary in this chapter, so we use $\underline{x} \in \mathcal{C}$ to refer to a codeword and refer to arbitrary or indexed elements of $\underline{x}$ by x or $x[n]$ respectively. Throughout the proofs in this chapter, we use $\delta(\epsilon)$ to refer to a function of $\epsilon > 0$ for which $\delta(\epsilon) \to 0$ as $\epsilon \to 0$. Also, $\epsilon_N \geq 0$ is used as shorthand for sequences which approach 0 as $N \to \infty$.

III.1. Three Decoding Paradigms

For the system model in this chapter, nodes A and B encode their binary message sequences $\underline{u}_A, \underline{u}_B \in \{0,1\}^K$ into length N codewords $\underline{x}_A \in \mathcal{C}_A$ and $\underline{x}_B \in \mathcal{C}_B$. The relay observes the output of a memoryless multiple access channel $P(Y_R|X_A, X_B)$. The objective of the relay is to reliably decode the message $\underline{u}_R = \underline{u}_A \oplus \underline{u}_B$ which is encoded and broadcast to nodes A and B. Again, the achievable computation rate $\mathcal{R}$ is the largest $\frac{K}{N}$ such that $\underline{u}_R$ can be decoded reliably in the usual information-theoretic sense. We primarily focus on the case where nodes A and B select independent cosets of identical linear codebooks generated uniformly at random for encoding. This ensemble provides the structure necessary for CF decoding, and exhibits the penalty constraints discovered in Chapter II if the relay uses DF decoding.

Decode-and-Forward: In the traditional DF scheme, nodes A and B use

independent codebooks, so the blockwise optimal DF decoder computes

$$\begin{aligned}(\hat{\underline{x}}_A, \hat{\underline{x}}_B)(\underline{y}_R) &= \underset{(\underline{x}_A, \underline{x}_B) \in \mathcal{C}_A \times \mathcal{C}_B}{\arg\max} P(\underline{y}_R | \underline{x}_A, \underline{x}_B) \\ \hat{\underline{x}}_A \to \hat{\underline{u}}_A \quad \hat{\underline{x}}_B \to \hat{\underline{u}}_B \\ \Rightarrow \hat{\underline{u}}_{R,DF} &= \hat{\underline{u}}_A \oplus \hat{\underline{u}}_B. \end{aligned} \tag{3.1}$$

DF can achieve all equal exchange rates subject to [13]

$$\mathcal{R}_{DF} < \min \left\{ \frac{1}{2} I(Y_R; X_A, X_B), I(Y_R; X_A | X_B), I(Y_R; X_B | X_A) \right\}. \tag{3.2}$$

Define $\mathcal{R}'_{DF}$ as the achievable rate for decoding $(\underline{x}_A, \underline{x}_B)$ if nodes A and B use an identical linear codebook $\mathcal{C}$. In Chapter II, we showed that if nodes A and B use the same linear code, then a rate penalty must be satisfied to recover $(\underline{x}_A, \underline{x}_B)$. To understand this, note that recovering $(\underline{x}_A, \underline{x}_B)$ means that the codeword $\underline{x}_R = \underline{x}_A \oplus \underline{x}_B \in \mathcal{C}$ has also been decoded. Therefore the multiple access channel constraints for the codeword pair $(\underline{x}_A, \underline{x}_R)$, given by

$$\mathcal{R} < \min \left\{ \frac{1}{2} I(Y_R; X_A, X_R), I(Y_R; X_A | X_R), I(Y_R; X_R | X_A) \right\}. \tag{3.3}$$

must be satisfied in addition to those required to recover $(\underline{x}_A, \underline{x}_B)$ with independent codebooks. The achievable rates for DF decoding with identical linear codebooks $\mathcal{R}'_{DF}$ is the union of the constraints in (3.2) and (3.3). Such a scheme can achieve all rates subject to

$$\mathcal{R}'_{DF} < \min \left\{ \mathcal{R}_{DF}, I(Y_R; X_A, X_B | X_R) \right\}. \tag{3.4}$$

Compute-and-Forward: The CF scheme requires node A and B to use the

same linear codebook. The blockwise optimal CF decoder computes

$$\begin{aligned}\hat{\underline{x}}_{R,CF}(\underline{y}_R) &= \arg\max_{\underline{x}_R \in \mathcal{C}} \sum_{\{\underline{x}_A, \underline{x}_B \in \{0,1\}^N | \underline{x}_R = \underline{x}_A \oplus \underline{x}_B\}} P(\underline{y}_R | \underline{x}_A, \underline{x}_B) \\ &= \arg\max_{\underline{x}_R \in \mathcal{C}} \prod_{n=1}^{N} P(y_R[n] | x_R[n]). \end{aligned} \tag{3.5}$$

CF can achieve all rates subject to

$$\mathcal{R}_{CF} < I(Y_R; X_R). \tag{3.6}$$

Joint-Compute-and-Forward: The JCF scheme also requires the use of identical linear codebooks. The blockwise optimal JCF decoder computes

$$\hat{\underline{x}}_{R,JCF}(\underline{y}_R) = \arg\max_{\underline{x}_R \in \mathcal{C}} \sum_{\{\underline{x}_A, \underline{x}_B \in \mathcal{C} | \underline{x}_R = \underline{x}_A \oplus \underline{x}_B\}} P(\underline{y}_R | \underline{x}_A, \underline{x}_B). \tag{3.7}$$

Note that the summation is taken over the set of codewords $\mathcal{C}$ instead of the set $\{0,1\}^N$ as in (3.5). Comparing (3.5) and (3.7), it is clear that the JCF decoder is better than the CF decoder because it only considers sequences which could possibly be transmitted from nodes A and B.

Notice that with either CF or JCF decoding, $\underline{u}_R = \underline{u}_A \oplus \underline{u}_B$ can be computed from $\underline{x}_R = \underline{x}_A \oplus \underline{x}_B$ because of the structure of the identical linear codebook $\mathcal{C}$.

III.2. Simultaneous Non-unique Decoding

A direct analysis of the probability of error for the maximum likelihood JCF decoder (3.7) is very difficult. However, we can move forward by adapting an analysis of the achievable rate region for interference networks presented in [46]. They consider a discrete memoryless multiple access channel $P(Y_R|X_A, X_B)$ in which the receiver is interested in decoding only one of the codewords $\underline{x}_A$ and does not need $\underline{x}_B$. The

simultaneous non-unique decoder from [46] works by making a list of all codeword pairs $(\underline{x}_A, \underline{x}_B)$ which are jointly typical with $\underline{y}_R$. If every pair in the list contains a unique $\hat{\underline{x}}_A$, the simultaneous non-unique decoder declares $\hat{\underline{x}}_A$ to be the desired codeword. If the list of typical codeword pairs is empty or contains multiple $\underline{x}_A$'s, the decoder declares an error.

For this problem the authors in [46] prove two key results. First, this simultaneous non-unique decoder can achieve any rate which is achievable by attempting to decode the exact pair $(\underline{x}_A, \underline{x}_B)$ or by treating the interfering codeword $\underline{x}_B$ as noise with a known distribution. Second, they conversely show that it is impossible to recover $\underline{x}_A$ at any higher rates than those achievable with the better of these two schemes. Key to their analysis is a restriction to code ensembles "with superposition coding and time sharing of independent and identically distributed (i.i.d.) codewords." For our purposes, it is enough to say that their results hold if $\mathcal{C}_A$ and $\mathcal{C}_B$ are generated independently at random.

In what follows, we perform an analysis of maximum likelihood JCF decoding for computation similar to the analysis in [46]. First, we define an ensemble of codeword triplets $(\underline{x}_A, \underline{x}_B, \underline{x}_R)$ about which we prove some properties which will be important for our proofs. Then we define a JCF typicality decoder which is analogous to the simultaneous non-unique typicality decoder in [46]. We derive the achievable computation rates for JCF typicality decoding with our ensemble of codeword triplets. Then, we conversely show that reliable computation of $\underline{x}_R$ is impossible at code rates higher than those achievable with JCF typicality decoding. Combining these results, we can conclude that maximum likelihood JCF decoding is rate-optimal for computation over binary-input discrete memoryless MACs if nodes A and B use independent cosets of a uniformly generated identical linear codebook.

III.3. Ensemble of Coset Codeword Triplets

The problem in [46] is very similar to our JCF problem, because the relay only needs to decode $\underline{x}_R$ and does not need to know the transmitted pair $(\underline{x}_A, \underline{x}_B)$. In order to perform a similar analysis for a JCF typicality decoder, we need to understand the ensemble of codeword triplets $(\underline{x}_A, \underline{x}_B, \underline{x}_R)$ which is obtained if nodes A and B use independent cosets of an identical linear code $\mathcal{C}$.

Specifically, let the elements of the generator matrix $\mathbf{G} \in \{0,1\}^{K \times N}$ and coset vectors $\underline{\lambda}_A, \underline{\lambda}_B \in \{0,1\}^N$ be i.i.d. bernoulli random variables with parameter $\frac{1}{2}$. Let m_A, m_B, and m_R denote the message indices whose binary expansions are $\underline{b}_A$, $\underline{b}_B$, and $\underline{b}_R$ respectively. Then let $m_R = m_A \oplus m_B$ refer to the fact that $\underline{b}_R = \underline{b}_A \oplus \underline{b}_B$. Then node A transmits

$$\underline{x}_A(m_A) = \underline{b}_A \mathbf{G} \oplus \underline{\lambda}_A, \tag{3.8}$$

and $\underline{x}_B(m_B)$ is encoded similarly. We use $\delta(\epsilon)$ to refer to a function of $\epsilon > 0$ for which $\delta(\epsilon) \to 0$ as $\epsilon \to 0$. Also, $\epsilon_N \geq 0$ is used as shorthand for sequences which approach 0 as $N \to \infty$.

A particular $\{\mathbf{G}, \underline{\lambda}_A, \underline{\lambda}_B\}$ from our code ensemble defines a set of $2^{2N\mathcal{R}}$ triplets $(\underline{x}_A, \underline{x}_B, \underline{x}_R)$ of length N coset codewords. Let $\mathcal{C}_N$ refer to the ensemble of such sets of triplets along with the mappings $(m_A, m_B, m_R) \to (\underline{x}_A, \underline{x}_B, \underline{x}_R)$. We recall from [44] that for the ensemble of random coset codes, the codeword associated with a fixed message is uniformly distributed on $\{0,1\}^N$, and codewords with unequal input messages are pairwise independent. This allows the use of the joint AEP to obtain upper bounds on the probability of error. The following lemmas about $\mathcal{C}_N$ are useful for our analysis and follow from the uniformity and pairwise independence of the ensemble of random linear coset codes.

Lemma III.1. *For a fixed* $(m_A, m_B, m_R) \in \mathcal{C}_N$ *the conditional distribution* $P(\underline{X}_A, \underline{X}_B | \underline{X}_R)$

on the ensemble $\mathcal{C}_N$, is

$$P(\underline{X}_A, \underline{X}_B|\underline{X}_R) = \begin{cases} \frac{1}{2^N} & , \quad \underline{x}_R = \underline{x}_A \oplus \underline{x}_B \\ 0 & , \qquad otherwise \end{cases} \tag{3.9}$$

Proof. First, note the joint distribution

$$\begin{aligned} P(\underline{X}_A, \underline{X}_R) &= P(\underline{X}_A)P(\underline{X}_B \oplus \underline{X}_A|\underline{X}_A) \\ &= P(\underline{X}_A)P(\underline{X}_B|\underline{X}_A) \\ &= P(\underline{X}_A)P(\underline{X}_B) = \frac{1}{2^{2N}}. \end{aligned}$$

Thus $\underline{X}_A$ and $\underline{X}_R$ are independent and uniformly distributed on $\{0,1\}^N$. Therefore, we obtain

$$\begin{aligned} P(\underline{X}_A, \underline{X}_B|\underline{X}_R) &= P(\underline{X}_A|\underline{X}_R)P(\underline{X}_B|\underline{X}_A, \underline{X}_R) \\ &= \frac{1}{2^N} 1_{\{\underline{x}_R = \underline{x}_A \oplus \underline{x}_B\}}. \end{aligned} \tag{3.10}$$

□

Lemma III.2. *For a fixed $(m_A, m_B, m_R) \in \mathcal{C}_N$, the joint distribution $P(\underline{X}_A, \underline{X}_B, \underline{X}_R)$ on the code ensemble $\mathcal{C}_N$, is*

$$P(\underline{X}_A, \underline{X}_B, \underline{X}_R) = \begin{cases} \frac{1}{2^{2N}} & , \quad \underline{x}_R = \underline{x}_A \oplus \underline{x}_B \\ 0 & , \qquad otherwise \end{cases}. \tag{3.11}$$

Proof. Since $\underline{X}_A$ and $\underline{X}_B$ are independent and uniformly distributed on $\{0,1\}^N$, we have

$$\begin{aligned} P(\underline{X}_A, \underline{X}_B, \underline{X}_R) &= P(\underline{X}_A, \underline{X}_B)P(\underline{X}_R|\underline{X}_A, \underline{X}_B) \\ &= \frac{1}{2^{2N}} 1_{\{\underline{x}_R = \underline{x}_A \oplus \underline{x}_B\}}. \end{aligned} \tag{3.12}$$

□

III.4. Achievable Rates for JCF Typicality Decoding

JCF Typicality Decoder: The JCF typicality decoder declares that $\hat{m}_R$ is the message index for the desired $\underline{x}_R$ if it is the unique $\hat{m}_R$ which satisfies

$$(\underline{x}_A(m_A), \underline{x}_B(m_B), \underline{x}_R(\hat{m}_R), \underline{y}_R) \in \mathcal{A}_\epsilon^N \text{ for some } m_A, m_B. \tag{3.13}$$

If there is no such $\hat{m}_R$ (i.e. none of the codeword triplets are jointly typical with $\underline{y}_R$) or if there are multiple such $\hat{m}_R$, then the decoder declares an error. Note that it does not matter which (m_A, m_B) was transmitted as long as a unique $\hat{m}_R$ is represented in the set of jointly typical triplets.

In (3.13), $\mathcal{A}_\epsilon^N$ is the set of jointly typical sequences of the random variables (X_A, X_B, X_R, Y_R) as defined in [13]. We include the definition for completeness. Define a subset of these four random variables by $\mathcal{S} \subseteq \{X_A, X_B, X_R, Y_R\}$. Then let $\underline{s}$ refer to an outcome of taking N independent observations of the random variables $\mathcal{S}$ according to the marginal distribution $P(\mathcal{S})$. Then the jointly typical sequences $\mathcal{A}_\epsilon^N$ are those which satisfy

$$\mathcal{A}_\epsilon^N = \left\{ (\underline{x}_A, \underline{x}_B, \underline{x}_R, \underline{y}_R) \;\middle|\; \left| \frac{1}{N} \log(P(\underline{s})) - H(\mathcal{S}) \right| < \epsilon \;\forall\; \mathcal{S} \subseteq \{X_A, X_B, X_R, Y_R\} \right\}. \tag{3.14}$$

This decoder satisfies our notion of JCF because it considers only the sequence pairs $(\underline{x}_A, \underline{x}_B)$ which satisfy the code constraints, yet the decoder will facilitate decoding even if some incorrect pairs $(\underline{x}'_A, \underline{x}'_B)$ are typical with $\underline{y}_R$ as long as $\underline{x}_R = \underline{x}'_A \oplus \underline{x}'_B$. In the following we prove that $\mathcal{R} < \mathcal{R}_{JCF} = \max\{\mathcal{R}_{CF}, \mathcal{R}_{DF}\}$ is achievable with a JCF typicality decoder and the ensemble $\mathcal{C}_N$. Then we conversely show that no decoder can reliably recover $\underline{x}_R$ from $\underline{y}_R$ with $\mathcal{C}_N$ unless $\mathcal{R} < \max\{\mathcal{R}_{CF}, \mathcal{R}_{DF}\}$. We emphasize that all of these results indicate the performance averaged over the ensemble $\mathcal{C}_N$.

Theorem III.1. *The JCF typicality decoder can reliably compute* $\underline{x}_R \in \mathcal{C}$ *at all computation rates subject to*

$$\mathcal{R}_{JCF} < \max(\mathcal{R}_{CF}, \mathcal{R}_{DF}) \tag{3.15}$$

if nodes A and B use a random code from the ensemble $\mathcal{C}_N$.

Proof. Since the channel is symmetrized by the random cosets, we can assume without loss of generality that the transmitted messages are $(m_A, m_B, m_R) = (0, 0, 0)$. An error occurs for the JCF typicality decoder only if one of the following events occur

$$\begin{aligned}
\mathcal{E}_1 &= \{(\underline{x}_A(0), \underline{x}_B(0), \underline{x}_R(0), \underline{y}_R) \notin \mathcal{A}_\epsilon^N\} \\
\mathcal{E}_2 &= \{(\underline{x}_A(m_A), \underline{x}_B(m_B), \underline{x}_R(m_R), \underline{y}_R) \in \mathcal{A}_\epsilon^N \text{ for some } m_R \neq 0 \text{ and some } m_A, m_B\}.
\end{aligned} \tag{3.16}$$

We have $P(\mathcal{E}_1) < \epsilon$ by the joint AEP [13].

By the definition of $\mathcal{A}_\epsilon^N$, we also have that

$$\{(\underline{x}_A(m_A), \underline{x}_B(m_B), \underline{x}_R(m_R), \underline{y}_R) \in \mathcal{A}_\epsilon^N\} \;\Rightarrow\; \{(\underline{x}_R(m_R), \underline{y}_R) \in \mathcal{A}_\epsilon^N\}. \tag{3.17}$$

Therefore,

$$\mathcal{E}_2 \subseteq \{(\underline{x}_R(m_R), \underline{y}_R) \in \mathcal{A}_\epsilon^N \text{ for some } m_R \neq 0\} = \mathcal{E}_2'. \tag{3.18}$$

By the joint AEP, we have

$$P(\mathcal{E}_2) \leq P(\mathcal{E}_2') \leq 2^{-N(\mathcal{R} - I(Y_R; X_R) - \delta(\epsilon))}. \tag{3.19}$$

We can also partition $\mathcal{E}_2$ into the disjoint events

$$\mathcal{E}_{21} = \{(\underline{x}_A(0), \underline{x}_B(m_B), \underline{x}_R(m_R), \underline{y}_R) \in \mathcal{A}_\epsilon^N \text{ for some } m_R = m_B \neq 0\}$$
$$\mathcal{E}_{22} = \{(\underline{x}_A(m_A), \underline{x}_B(0), \underline{x}_R(m_R), \underline{y}_R) \in \mathcal{A}_\epsilon^N \text{ for some } m_R = m_A \neq 0\}$$
$$\mathcal{E}_{23} = \{(\underline{x}_A(m_A), \underline{x}_B(m_B), \underline{x}_R(m_R), \underline{y}_R) \in \mathcal{A}_\epsilon^N \text{ for some } m_R \neq 0, \text{ and } 0 \neq m_B \neq m_A \neq 0\}.$$

The condition $\{m_R = m_B \neq 0\}$, recalls the fact that $\underline{b}_R = \underline{b}_A \oplus \underline{b}_B$. Therefore if $\underline{b}_A = \underline{0}$ which corresponds to node A's transmitted message, then $\underline{b}_R = \underline{b}_B \neq \underline{0}$ forms a class of $|\mathcal{E}_{21}| = 2^{N\mathcal{R}}$ error events. A similar argument can be made for the $\{m_R = m_A \neq 0\}$ condition. Finally, the condition $\{m_R \neq 0, \text{ and } 0 \neq m_B \neq m_A \neq 0\} \Leftrightarrow \{\underline{0} \neq \underline{b}_A \neq \underline{b}_B \neq \underline{0}\}$. Thus $\underline{b}_A, \underline{b}_B$ are both unequal to the correct messages, and the error patterns differ in at least one bit position. This forms a class of $|\mathcal{E}_{23}| = (2^{N\mathcal{R}} - 1)(2^{N\mathcal{R}} - 2) \leq 2^{2N\mathcal{R}}$ error events (i.e. there are $(2^{N\mathcal{R}} - 1)$ incorrect messages, and we must chose two without replacement).

By the joint AEP, we obtain the following bounds

$$\begin{aligned} P(\mathcal{E}_{21}) &< 2^{-N(\mathcal{R} - I(Y_R; X_R, X_A | X_B) - \delta(\epsilon))} \\ P(\mathcal{E}_{22}) &< 2^{-N(\mathcal{R} - I(Y_R; X_R, X_B | X_A) - \delta(\epsilon))} \\ P(\mathcal{E}_{23}) &< 2^{-N(2\mathcal{R} - I(Y_R; X_R, X_A, X_B) - \delta(\epsilon))}. \end{aligned} \tag{3.20}$$

Since $I(Y_R; X_R | X_A, X_B) = 0$, we use the chain rule to get

$$\begin{aligned} I(Y_R; X_R, X_A | X_B) &= I(Y_R; X_A | X_B) \\ I(Y_R; X_R, X_B | X_A) &= I(Y_R; X_B | X_A) \\ I(Y_R; X_R, X_A, X_B) &= I(Y_R; X_A, X_B). \end{aligned} \tag{3.21}$$

Combining (3.20) and (3.21) we have

$$\begin{aligned} P(\mathcal{E}_{21}) &< 2^{-N(\mathcal{R}-I(Y_R;X_A|X_B)-\delta(\epsilon))} \\ P(\mathcal{E}_{22}) &< 2^{-N(\mathcal{R}-I(Y_R;X_B|X_A)-\delta(\epsilon))} \\ P(\mathcal{E}_{23}) &< 2^{-N(2\mathcal{R}-I(Y_R;X_A,X_B)-\delta(\epsilon))}. \end{aligned} \tag{3.22}$$

Since $\mathcal{E}_2 = \mathcal{E}_{21} \cup \mathcal{E}_{22} \cup \mathcal{E}_{23}$, we have

$$P(\mathcal{E}_2) \leq \min\left\{P(\mathcal{E}_{21}), P(\mathcal{E}_{22}), P(\mathcal{E}_{23})\right\}. \tag{3.23}$$

Combining (3.19) and (3.23), we have $P(\mathcal{E}_2) \to 0$ as $N \to \infty$ if

$$\mathcal{R} \leq \max(\mathcal{R}_{CF}, \mathcal{R}_{DF}). \tag{3.24}$$

□

Notice that the penalty constraint $I(Y_R; X_A, X_B|X_R)$ from (3.4) does not appear in the achievability proof. The penalty constraint can be used with the joint AEP to upper bound the probability of the event

$$\mathcal{E}_3 = \{(\underline{x}_A(m_A), \underline{x}_B(m_B), \underline{x}_R(m_R), \underline{y}_R) \in \mathcal{A}_\epsilon^N \text{ for } m_R = 0, \text{ and } (m_A, m_B) \neq (0,0)\}. \tag{3.25}$$

Since we only want to recover the correct $m_R = 0$, we do not care if this event occurs.

III.5. Converse for Binary Linear Codebooks

Interestingly, the penalty constraint and the corresponding bound obtained from the joint AEP expresses itself in the converse proof for the achievable computation rates on $P(Y_R|X_A, X_B)$ averaged over the ensemble $\mathcal{C}_N$. To obtain the converse, we will require the following lemma which is adapted from [46].

Lemma III.3.

$$\lim_{N\to\infty}\frac{1}{N}H(\underline{Y}_R|\underline{X}_R,\mathcal{C}_N) = H(Y_R|X_A,X_B) + \min\{\mathcal{R}, I(Y_R;X_A,X_B|X_R)\}. \tag{3.26}$$

Proof. We first obtain the upper bounds

$$H(\underline{Y}_R|\underline{X}_R,\mathcal{C}_N) \le NH(Y_R|X_R) \tag{3.27}$$

and

$$\begin{aligned} H(\underline{Y}_R|\underline{X}_R,\mathcal{C}_N) &\le H(\underline{Y}_R,M_A,M_B|\underline{X}_R,\mathcal{C}_N) \\ &= N\mathcal{R} + H(\underline{Y}_R|\underline{X}_A,\underline{X}_B) \\ &\le N(\mathcal{R} + H(Y_R|X_A,X_B)). \end{aligned} \tag{3.28}$$

These are obtained because the channel is memoryless and by Lemmas III.1 and III.2.

We continue our assumption that the true messages are $(m_A,m_B,m_R)=(0,0,0)$ to simplify notation. We use the identity $H(Y|Z) = H(Y,X|Z) - H(X|Y,Z) = H(Y|X,Z)+H(X|Z)-H(X|Y,Z)$ with the substitutions $Y\to\underline{Y}_R$, $X\to(M_A,M_B)$, and $Z\to(\mathcal{C}_N,\underline{X}_R)$ to get

$$\begin{aligned} H(\underline{Y}_R|\underline{X}_R,\mathcal{C}_N) = H(\underline{Y}_R|\underline{X}_R,\mathcal{C}_N,M_A,M_B) + H(M_A,M_B|\underline{X}_R,\mathcal{C}_N) \\ - H(M_A,M_B|\underline{X}_R,\mathcal{C}_N,\underline{Y}_R). \end{aligned} \tag{3.29}$$

The first two terms on the RHS can be simplified to

$$\begin{aligned} H(\underline{Y}_R|\underline{X}_R,\mathcal{C}_N,M_A,M_B) &= NH(Y_R|X_R,X_A,X_B) \\ &= NH(Y_R|X_A,X_B) \end{aligned} \tag{3.30}$$

$$H(M_A,M_B|\underline{X}_R,\mathcal{C}_N) = N\mathcal{R} \tag{3.31}$$

which follow by the channel structure and Lemma III.1.

We upper bound the term $H(M_A, M_B|\underline{X}_R, \mathcal{C}_N, \underline{Y}_R)$ to obtain a lower bound on $H(\underline{Y}_R|\underline{X}_R, \mathcal{C}_N)$. Construct a list $\mathcal{L} \subseteq \{0, ..., 2^{N\mathcal{R}} - 1\}^2$ of pairs (m_A, m_B) which satisfy

$$\mathcal{L} = \{(m_A, m_B) \mid (\underline{X}_A(m_A), \underline{X}_B(m_B), \underline{X}_R, \underline{Y}_R) \in \mathcal{A}_\epsilon^N\}.$$

There are only $2^{N\mathcal{R}}$ pairs (m_A, m_B) which correspond to the given $\underline{x}_R(0)$. This includes the transmitted pair $(m_A, m_B) = (0, 0)$ and $2^{N\mathcal{R}} - 1$ incorrect message pairs. The probability that any of the incorrect message pairs is jointly typical with $(\underline{Y}_R, \underline{X}_R)$ is upper bounded by

$$\begin{aligned}
&P((\underline{X}_A(m_A), \underline{X}_B(m_B), \underline{X}_R(0), \underline{Y}_R) \in \mathcal{A}_\epsilon^N) \\
&= \sum_{(\underline{x}_A, \underline{x}_B, \underline{x}_R, \underline{y}_R) \in \mathcal{A}_\epsilon^N} P(\underline{X}_A, \underline{X}_B|\underline{X}_R) P(\underline{X}_R, \underline{Y}_R) \\
&= |\mathcal{A}_\epsilon^N| P(\underline{X}_A, \underline{X}_B|\underline{X}_R) P(\underline{X}_R, \underline{Y}_R) \\
&\leq 2^{N(H(X_A, X_B, X_R, Y_R) + \delta(\epsilon))} 2^{-N(H(X_A, X_B|X_R) - \delta(\epsilon))} 2^{-N(H(X_R, Y_R) - \delta(\epsilon))} \\
&= 2^{-N(H(X_A, X_B|X_R) + H(X_R, Y_R) - H(X_A, X_B, X_R, Y_R) - \delta(\epsilon))} \\
&= 2^{-N(H(X_A, X_B|X_R) - H(X_A, X_B|X_R, Y_R) - \delta(\epsilon))} \\
&= 2^{-N(I(Y_R; X_A, X_B|X_R) - \delta(\epsilon))}.
\end{aligned} \tag{3.32}$$

The term $P(\underline{X}_A, \underline{X}_B|\underline{X}_R)$ is used in the first step because we consider incorrect message pairs for which $\underline{x}_R(0)$ is correct.

We obtain our desired upper bound through an argument on the cardinality of $\mathcal{L}$. Particularly, we have $|\mathcal{L}| \leq 1 + T$ where T is a binomial random variable with $2^{N\mathcal{R}} - 1$ trials and a maximum probability of success $2^{-N(I(Y_R; X_A, X_B|X_R) - \delta(\epsilon))}$. Therefore

$$E[|\mathcal{L}|] \leq 1 + 2^{N(\mathcal{R} - I(X_A, X_B; Y_R|X_R) + \delta(\epsilon))}. \tag{3.33}$$

The correct pair $(m_A, m_B) = (0, 0)$ is likely to be in $\mathcal{L}$. Therefore, define an indicator random variable $E = 1_{\{(0,0)\in\mathcal{L}\}}$. Following the logic from [46], we have

$$\begin{aligned}
H(M_A, M_B|\underline{X}_R, \underline{Y}_R, \mathcal{C}_N) &= I(M_A, M_B; E|\underline{X}_R, \underline{Y}_R, \mathcal{C}_N) + H(M_A, M_B|\underline{X}_R, \underline{Y}_R, \mathcal{C}_N, E) \\
&\leq H(E|\underline{X}_R, \underline{Y}_R, \mathcal{C}_N) + H(M_A, M_B|\underline{X}_R, \underline{Y}_R, \mathcal{C}_N, E) \\
&\overset{(a)}{\leq} 1 + H(M_A, M_B|\underline{X}_R, \underline{Y}_R, \mathcal{C}_N, E) \\
&= 1 + P(E = 0)H(M_A, M_B|\underline{X}_R, \underline{Y}_R, \mathcal{C}_N, E = 0) \\
&\quad + P(E = 1)H(M_A, M_B|\underline{X}_R, \underline{Y}_R, \mathcal{C}_N, E = 1) \\
&\leq 1 + P(E = 0)N\mathcal{R} + P(E = 1)H(M_A, M_B|\underline{X}_R, \underline{Y}_R, \mathcal{C}_N, E = 1) \qquad (3.34)
\end{aligned}$$

where (a) holds because E is a binary random variable. The last term can be upper bounded by arguing that if $E = 1$, then the transmitted (m_A, m_B) must be one of the $|\mathcal{L}|$ pairs in $\mathcal{L}$. Thus the conditional entropy can be no more than $\log(|\mathcal{L}|)$. Therefore

$$\begin{aligned}
H(M_A, M_B|\underline{X}_R, \underline{Y}_R, \mathcal{C}_N, E = 1) &\overset{(a)}{=} H(M_A, M_B|\underline{X}_R, \underline{Y}_R, \mathcal{C}_N, \mathcal{L}, |\mathcal{L}|, E = 1) \\
&\leq H(M_A, M_B|\mathcal{L}, |\mathcal{L}|, E = 1) \\
&= \sum_{\ell=0}^{2^{N\mathcal{R}-1}} P(|\mathcal{L}| = \ell)H(M_A, M_B|\mathcal{L}, |\mathcal{L}| = \ell, E = 1) \\
&\leq \sum_{\ell=0}^{2^{N\mathcal{R}-1}} P(|\mathcal{L}| = \ell)\log(\ell) \\
&= E[\log(|\mathcal{L}|)] \\
&\overset{(b)}{\leq} \log(E[|\mathcal{L}|]) \\
&\leq \log(1 + 2^{N(\mathcal{R}-I(X_A,X_B;Y_R|X_R)+\delta(\epsilon))}) \qquad (3.35)
\end{aligned}$$

where (a) is true because $\mathcal{L}$ and $|\mathcal{L}|$ are functions of $\underline{X}_R, \underline{Y}_R, \mathcal{C}_N$, and (b) follows from Jensen's inequality [13]. We obtain an upper bound on the last step of (3.35)

in two ways. Particularly, note that for any real constant α,

$$\begin{aligned}
&\log(1+2^{\alpha}) \leq \log(1+1) = 1 \quad \forall\ \alpha \leq 0 \\
&\log(1+2^{\alpha}) \leq \log(2^{\alpha}+2^{\alpha}) = 1+\alpha \quad \forall\ \alpha \geq 0.
\end{aligned} \tag{3.36}$$

Substituting $\alpha = N(\mathcal{R} - I(X_A, X_B; Y_R|X_R) + \delta(\epsilon))$ for the last step of (3.35), we get

$$H(M_A, M_B|\underline{X}_R, \underline{Y}_R, \mathcal{C}_N, E=1) \leq 1 + \max\{0, N(\mathcal{R} - I(X_A, X_B; Y_R|X_R) + \delta(\epsilon))\} \tag{3.37}$$

Combining (3.37) and (3.34) we obtain

$$\begin{aligned}
&H(M_A, M_B|\underline{X}_R, \underline{Y}_R, \mathcal{C}_N) \\
&\quad \leq 2 + N\mathcal{R}P(E=0) + \max\{0, N(\mathcal{R} - I(Y_R; X_A, X_B|X_R)) + \delta(\epsilon)\}.
\end{aligned} \tag{3.38}$$

Substituting back into (3.29) and applying (3.30) and (3.31), we get

$$\begin{aligned}
&H(\underline{Y}_R|\underline{X}_R, \mathcal{C}_N) \\
&\quad \geq NH(Y_R|X_A, X_B) + N\mathcal{R} - 2 - N\mathcal{R}P(E=0) \\
&\quad - \max\{0, N(\mathcal{R} - I(Y_R; X_A, X_B|X_R)) + \delta(\epsilon)\}.
\end{aligned} \tag{3.39}$$

Dividing by N and simplifying, we obtain

$$\begin{aligned}
&\frac{1}{N}H(\underline{Y}_R|\underline{X}_R, \mathcal{C}_N) \\
&\quad \geq H(Y_R|X_A, X_B) + \min\{\mathcal{R}, I(Y_R; X_A, X_B|X_R)) + \delta(\epsilon)\} - \frac{2}{N} - \mathcal{R}P(E=0).
\end{aligned} \tag{3.40}$$

Recall that we take the limit as $N \to \infty$ so that the last two terms approach zero. Also, $\delta(\epsilon)$ can be made arbitrarily small by our selection of ϵ. The result follows. $\square$

Theorem III.2. *If* $P_e^N(\mathcal{C}_N) \to 0$ *as* $N \to \infty$ *then*

$$\mathcal{R} < \max\{\mathcal{R}_{CF}, \mathcal{R}_{DF}\}. \tag{3.41}$$

Proof. We use the weak version of Fano's inequality

$$H(M_R|\underline{Y}_R, \mathcal{C}_N) \leq 1 + N\mathcal{R}P_e^N(\mathcal{C}_N) \leq N\epsilon_N. \tag{3.42}$$

where $\epsilon_N \to 0$ as $N \to \infty$ since $P_e^N(\mathcal{C}_N) \to 0$.

First, $\mathcal{R} < I(Y_R; X_R|X_A)$ must always be satisfied.

$$\begin{aligned}
N(\mathcal{R} - \epsilon_N) &= H(M_R|\mathcal{C}_N) - N\epsilon_N \\
&= I(\underline{Y}_R; M_R|\mathcal{C}_N) + H(M_R|\underline{Y}_R, \mathcal{C}_N) - N\epsilon_N \\
&\overset{(a)}{\leq} I(\underline{Y}_R; M_R|\mathcal{C}_N) + N\epsilon_N - N\epsilon_N \\
&\leq I(\underline{Y}_R; \underline{X}_R|\mathcal{C}_N) \\
&= H(\underline{X}_R|\mathcal{C}_N) - H(\underline{X}_R|\underline{Y}_R, \mathcal{C}_N) \\
&\overset{(b)}{=} H(\underline{X}_R|\underline{X}_A, \mathcal{C}_N) - H(\underline{X}_R|\underline{Y}_R, \mathcal{C}_N) \\
&\leq H(\underline{X}_R|\underline{X}_A, \mathcal{C}_N) - H(\underline{X}_R|\underline{Y}_R, \underline{X}_A, \mathcal{C}_N) \\
&= I(\underline{Y}_R; \underline{X}_R|\underline{X}_A, \mathcal{C}_N) \\
&= H(\underline{Y}_R|\underline{X}_A, \mathcal{C}_N) - H(\underline{Y}_R|\underline{X}_R, \underline{X}_A, \mathcal{C}_N) \\
&\overset{(c)}{=} H(\underline{Y}_R|\underline{X}_A, \mathcal{C}_N) - NH(Y_R|X_R, X_A) \\
&\overset{(d)}{\leq} NH(Y_R|X_A) - NH(Y_R|X_R, X_A) \\
&= NI(Y_R; X_R|X_A).
\end{aligned} \tag{3.43}$$

Here (a) holds by Fano's inequality, (b) holds by Lemma III.2, (c) holds because the channel is memoryless, so each Y_R depends only on $\{X_R, X_A\} \Leftrightarrow \{X_A, X_B\}$ which are i.i.d. by Lemma III.2, (d) holds by the chain rule and by removing conditioning.

Therefore $\mathcal{R} < I(Y_R; X_R|X_A)$ if $P_e(\mathcal{C}_N) \to 0$. Similarly, $\mathcal{R} < I(Y_R; X_R|X_B)$. Note that $I(Y_R; X_R|X_A) = I(Y_R; X_B|X_A)$ and $I(Y_R; X_R|X_B) = I(Y_R; X_A|X_B)$.

Now, we want to show

$$\mathcal{R} < \max\left\{ I(Y_R; X_R), \frac{1}{2} I(Y_R; X_A, X_B) \right\}.$$

Using Fano's inequality and Lemma III.3, we get

$$\begin{aligned}
N(\mathcal{R} - \epsilon_N) &\leq I(\underline{Y}_R; \underline{X}_R|\mathcal{C}_N) \\
&= H(\underline{Y}_R|\mathcal{C}_N) - H(\underline{Y}_R|\underline{X}_R, \mathcal{C}_N) \\
&\leq NH(Y_R) - H(\underline{Y}_R|\underline{X}_R, \mathcal{C}_N) \\
&\leq NH(Y_R) - NH(Y_R|X_A, X_B) - N(\min\{\mathcal{R}, I(Y_R; X_A, X_B|X_R)\} + N\epsilon_N \\
&= N\left(I(Y_R; X_A, X_B) - \min\{\mathcal{R}, I(Y_R; X_A, X_B|X_R)\} + \epsilon_N \right).
\end{aligned}$$

Dividing by N and omitting ϵ_N, we get

$$\begin{aligned}
\mathcal{R} &\leq I(Y_R; X_A, X_B) - \min\{\mathcal{R}, I(Y_R; X_A, X_B|X_R)\} \\
\mathcal{R} &\leq \max\{I(Y_R; X_A, X_B) - \mathcal{R}, I(Y_R; X_R)\} \\
\mathcal{R} &\leq \max\left\{ \frac{1}{2} I(Y_R; X_A, X_B), I(Y_R; X_R) \right\}. \qquad (3.44)
\end{aligned}$$

We conclude that

$$\mathcal{R} < \max\left\{ \mathcal{R}_{CF}, \mathcal{R}_{DF} \right\} \qquad (3.45)$$

□

We have shown that JCF performs as well as the better of CF or DF for the binary computation problem for a large class of channels for the ensemble of identical linear coset codes. We have further shown that no decoder can reliably recover $\underline{x}_R$ if identical linear codes are used with rates larger than $\max\{\mathcal{R}_{CF}, \mathcal{R}_{DF}\}$. We note

that the penalty constraint $I(Y_R; X_A, X_B|X_R)$ in (3.4) must be satisfied if we want to recover the entire pair $(\underline{x}_A, \underline{x}_B)$. It is shown in Theorem III.3, that this constraint is never dominant if $\mathcal{R}_{DF} > \mathcal{R}_{CF}$.

Theorem III.3. *If nodes A and B use the ensemble $\mathcal{C}_N$. Then the penalty constraint $I(Y_R; X_A, X_B|X_R)$ is never dominant for any binary-input multiple access channel if $\mathcal{R}_{DF} > \mathcal{R}_{CF}$.*

Proof. The active constraint for $\mathcal{R}_{DF}$ is the smallest of $I(Y_R; X_A|X_B)$, $I(Y_R; X_B|X_A)$, and $\frac{1}{2}I(Y_R; X_A, X_B)$. Suppose, $\mathcal{R}_{DF} = I(Y_R; X_A|X_B)$. Then since $\mathcal{R}_{CF} = I(Y_R; X_R) \leq I(Y_R; X_A|X_B) = \mathcal{R}_{DF}$, we have

$$\begin{aligned} I(Y_R; X_A|X_B) &\leq \frac{1}{2}I(Y_R; X_A, X_B) = \frac{1}{2}(I(Y_R; X_R) + I(Y_R; X_A, X_B|X_R)) \\ I(Y_R; X_A|X_B) &\leq \frac{1}{2}(I(Y_R; X_A|X_B) + I(Y_R; X_A, X_B|X_R)). \end{aligned} \tag{3.46}$$

Simplifying we get

$$I(Y_R; X_A|X_B) \leq I(Y_R; X_A, X_B|X_R) \tag{3.47}$$

so the penalty constraint is not dominant. The same steps work for $\mathcal{R}_{DF} = I(Y_R; X_B|X_A)$.

Suppose $\mathcal{R}_{DF} = \frac{1}{2}I(Y_R; X_A, X_B)$. Then, applying (3.2) and (3.6) we have

$$\begin{aligned} I(Y_R; X_R) &\leq \frac{1}{2}I(Y_R; X_A, X_B) \\ &= \frac{1}{2}(I(Y_R; X_R) + I(Y_R; X_A, X_B|X_R)) \\ \Rightarrow I(Y_R; X_R) &\leq I(Y_R; X_A, X_B|X_R). \end{aligned} \tag{3.48}$$

Therefore we have

$$\begin{aligned} \frac{I(Y_R; X_A, X_B|X_R) + I(Y_R; X_R)}{2} &\leq \frac{I(Y_R; X_A, X_B|X_R) + I(Y_R; X_A, X_B|X_R)}{2} \\ \frac{1}{2}I(Y_R; X_A, X_B) &\leq I(Y_R; X_A, X_B|X_R). \end{aligned} \tag{3.49}$$

The penalty constraint is not dominant for each case. □

From Theorem III.3, we can conclude that the computation rates achieved by JCF decoding can be achieved by using the better of DF or CF without making any changes to the encoder. The discussion of the penalty constraint shifts our focus from the computation problem, but helps clarify the limitations of the ensemble $\mathcal{C}_N$. In the proof of Theorem III.2, if the error event in Fano's inequality represents $\hat{\underline{x}}_A \neq \underline{x}_A$, then swapping all the terms $X_R \leftrightarrow X_A$ in (3.42) and (3.43) shows that $\mathcal{R} < I(Y_R; X_A|X_R)$ to reliably recover $\underline{x}_A$. Repeat this analysis for $\underline{x}_B$ and note that $I(Y_R; X_A|X_R) = I(Y_R; X_A, X_B|X_R) = I(Y_R; X_B|X_R)$. Then the penalty constraint must be satisfied to recover either $\underline{x}_A$ or $\underline{x}_B$ with $\mathcal{C}_N$.

We remark that for ensembles similar to $\mathcal{C}_N$ which allow for the recovery of more than one type of function (e.g. multilevel codes or coding over larger finite fields discussed in Chapter II and [26]), there may be multiple penalty constraints and some may be dominant for certain channel parameters. The JCF decoding paradigm is practically suited to adaptive computation because it naturally uses the best decoding function or recovers a suitable subset of messages with a single decoder. In the next chapter we focus on practical implementation of JCF decoding and design codes which facilitate decoding at information rates near the theoretical limit for identical linear codebooks.

III.6. Concluding Remarks

In this chapter, we have analyzed the performance of joint-compute-and-forward decoding for reliable PLNC with identical binary linear codebooks. JCF achieves computation rates equal to the best of the compute-and-forward or decode-and-forward schemes for binary-input memoryless multiple access channels. Further, a

converse is provided for the binary case which shows that innovations to the decoder cannot improve achievable computation rates. Therefore, the best of CF and DF provides a fundamental limit for computation rates with the studied code ensemble. If it is possible to improve the computation rates for the binary-input memoryless MAC, it will require clever joint design of the code ensembles used by the transmitting nodes. Next in Chapter IV, we consider how to practically achieve the theoretical performance for identical linear codebooks with JCF message passing decoding.

CHAPTER IV

JOINT-COMPUTE-AND-FORWARD MESSAGE PASSING FOR TWO-WAY ERASURE MULTIPLE ACCESS CHANNELS*

In the previous chapter, we performed an information-theoretic analysis of DF, CF, and JCF decoding at the relay node. We found that JCF decoding essentially performs as well as the better of CF and DF decoding depending on the channel conditions. In this chapter, we consider how to achieve this theoretical performance with JCF message passing decoding using practical LDPC ensembles This investigation seems natural because JCF was originally proposed in a message passing framework [17], [16], [18], [19]. To this end, we introduce a simple class of erasure multiple access channels and derive the processing rules and a density evolution like analysis. We then perform numerical experiments for several classes of LDPC ensembles. Our numerical results show that spatially coupled LDPC ensembles are capable of achieving the optimal computation rates for identical linear codebooks. We show that uniform puncturing of the spatially coupled ensembles can be used to obtain this performance for a range of channel parameters with a single encoder and decoder.

One might reason from our previous analysis in Chapter III that the relay should simply use either CF or DF decoding as we have Chapter II. In this chapter, we find that for certain channel parameters and code ensembles JCF message passing may achieve strictly better decoding thresholds than CF or DF based message passing. Admittedly, these examples are for regular LDPC ensembles for which message passing decoding is known to be suboptimal [38]. However, we also provide one simple

example with the proposed channel model where nodes A and B use a length $N = 5$ identical linear codebook and JCF decoding successfully decodes while both CF and DF decoding fail. This chapter furthers our investigation and provides some useful tools for further research.

Note about notation: In this chapter, we use $\underline{x} \in \mathcal{C}$ to refer to codewords just as in Chapter III. For details, see the *note about notation* near the beginning of Chapter III. Normally, we use subscripts to indicate whether a variable is associated with a specific node. In this chapter, we make one exception to this convention by dropping the subscript on $x_{A,B}$ which is referred to simply by x. This is mentioned in the text near (4.11) but is used throughout the chapter.

IV.1. TWEMAC Channel Model

In order to exactly characterize the performance of a JCF message passing decoder for an LDPC code or ensemble, we must restrict our attention to a simplified channel model. We introduce the two way erasure multiple access (TWEMAC) channel model for which this characterization is numerically efficient. We introduced a special case of this channel very recently in [47]. A TWEMAC is randomly in one of five states $\tau \in \mathcal{T} = \{1, ..., 5\}$ during each channel use. The value of y_R is a deterministic function of x_A, x_B and the state τ. In this chapter, the probability P_{ch}^{τ} that the channel is in state τ is parameterized by an erasure probability $\epsilon \in [0, 1]$. The subscript ch associates this type distribution with the channel output. Thus the

channel is defined

$$Y_R = \begin{cases} (E, E) & \text{with probability} \quad P_{ch}^1(\epsilon) \\ (X_A, E) & \text{with probability} \quad P_{ch}^2(\epsilon) \\ (E, X_B) & \text{with probability} \quad P_{ch}^3(\epsilon) \\ (X_A \oplus X_B) & \text{with probability} \quad P_{ch}^4(\epsilon) \\ (X_A, X_B) & \text{with probability} \quad P_{ch}^5(\epsilon) \end{cases} \tag{4.1}$$

where E denotes an erasure and $\sum_{i=1}^{5} P_{ch}^i(\epsilon) = 1 \; \forall \; \epsilon \in [0, 1]$. The state of the channel is assumed to be known to the relay but unknown at nodes A and B.

Since we are interested in message passing decoding, it is useful to refer to the messages from the channel as having a message type from the set $\mathcal{T}$. Then messages passed along the edges in the decoder will also have a type from the set $\mathcal{T}$. We define the type distribution vector for a given channel according to

$$\underline{P}_{ch}(\epsilon) = \begin{bmatrix} P_{ch}^1(\epsilon) \\ P_{ch}^2(\epsilon) \\ P_{ch}^3(\epsilon) \\ P_{ch}^4(\epsilon) \\ P_{ch}^5(\epsilon) \end{bmatrix}. \tag{4.2}$$

Throughout this chapter, we will use the underlined $\underline{P}_{ch}$ to denote a 5-ary type distribution, and we use P_{ch}^i to denote the probability of a message of type $i \in \mathcal{T}$.

The class of TWEMACs has the advantage that the probabilities $P_{ch}^\tau, \tau \in \mathcal{T}$ can be chosen in order to mimic or isolate many characteristics of wireless channels. For example, the authors in [47] study capacity for recovering $\underline{b}_R = \underline{b}_A \oplus \underline{b}_B$ at a relay

with a channel equivalent to

$$\underline{P}_{ch} = \begin{bmatrix} 0 \\ \frac{\epsilon}{2} \\ \frac{\epsilon}{2} \\ 1-\epsilon \\ 0 \end{bmatrix}.$$

In this channel, the quality of the matching between $P(Y_R|X_A, X_B)$ and $X_R = X_A \oplus X_B$ is ideal when $\epsilon = 0$ and continuously degrades until it is unmatched when $\epsilon = 1$. This allowed the authors to analyze how decoding strategies should change when the quality of the matching changes. They also find that $\max\{\mathcal{R}_{CF}, \mathcal{R}_{DF})$ can be beaten if the transmitters are provided with strictly causal feedback of the state of the channel. This does not conflict with our converse result because they do not employ identical linear codebooks generated uniformly at random for this performance. The codebooks are adaptively (and jointly) designed based on knowledge of the relays observations.

The authors in [40] study the achievable rate region for decoding $\underline{x}_A$ and $\underline{x}_B$ at the relay with independent spatially coupled codes using a channel equivalent to

$$\underline{P}_{ch} = \begin{bmatrix} \epsilon \\ 0 \\ 0 \\ \frac{1-\epsilon}{2} \\ \frac{1-\epsilon}{2} \end{bmatrix}.$$

This parametrization mimics the effects of interference and noise for a wireless MAC channel, while the quality of the matching between $P(Y_R|X_A, X_B)$ and X_R remains constant with ϵ (i.e. half the non-erased symbols interfere). This makes the density

evolution analysis similar to that of the point-to-point erasure channel where the erasure probability emulates noise.

Our TWEMAC was recently generalized to an erasure multi-way relay channel (EMWRC) in [48]. They consider the problem in which all source nodes want to fully exchange their information with the use of a central relay. Their EMWRC is parameterized by independent erasure probabilities for each link to mimic packet erasures in a fading environment with wireless superposition. The simplicity of the channel is leveraged to design and analyze packet mixing strategies based on fountain codes for the multiple access and broadcast phases. A relay network consisting of a class of erasure multiple access channels was also studied in [49]. The simplicity of the erasure structure was used for a network information-theoretic analysis of the unicast problem. Their model similarly mimics a fading environment with wireless superposition.

For this chapter, we will mostly present numerical results for the TWEMAC parameterized by

$$\underline{P}_{ch}(\epsilon) = \begin{bmatrix} \epsilon^2 \\ (1-\epsilon)\epsilon \\ \epsilon(1-\epsilon) \\ (1-\epsilon)^2 \\ 0 \end{bmatrix}. \tag{4.3}$$

This channel is generated if we assume that x_A and x_B are erased with probability ϵ independently. When neither symbol is erased, the relay observes the type 4 message which gives $x_R = x_A \oplus x_B$. When x_A (x_B) is erased and x_B (x_A) is not, the relay receives a message of type 3 (type 2). For this channel, the erasure parameter ϵ mimics the signal to noise ratio for an AWGN channel. When both channels are strong, the matching improves until it is ideal at $\epsilon = 1$.

IV.2. Example: JCF beats CF and DF

Before we dive into the details of the message processing rules and performance analysis, a brief example utilizing the TWEMAC channel model is useful to demonstrate the potential benefits of JCF decoding. Suppose that nodes A and B each use an identical linear codebook corresponding to generator matrix

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix}. \tag{4.4}$$

This generates the codebook

$$\mathcal{C} = \left\{ \begin{array}{ccccc} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 \end{array} \right\}. \tag{4.5}$$

Then, suppose that the relay observes the following output from a TWEMAC

$$\underline{y}_R = [(1, E), (E, 1), (1), (1, 1), (1)] \tag{4.6}$$

which corresponds to the sequence of message types $[2, 3, 4, 5, 4]$.

Upon receiving this $\underline{y}_R$, the maximum likelihood CF decoder attempts to find a unique codeword $\underline{x}_R \in \mathcal{C}$ such that the last three bits are 1,0,1. There are two

codewords which satisfy this

$$\underline{x}_{R,CF} \in \left\{ \begin{array}{ccccc} 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 \end{array} \right\}. \tag{4.7}$$

Therefore, the maximum likelihood CF decoder fails.

The DF decoder attempts to find a unique pair of codewords $(\underline{x}_A, \underline{x}_B)$ which could correspond to the received sequence. If we first consider only the received signals of type 2, 3, and 5, then the possible codeword pairs include

$$(\underline{x}_A, \underline{x}_B) \in \left\{ \begin{array}{ccccc|c|ccccc} 1 & 0 & 0 & 1 & 1 & , & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & , & 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & , & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 & , & 0 & 1 & 1 & 1 & 0 \end{array} \right\}. \tag{4.8}$$

The set of possible codeword pairs can be further reduced by considering the messages of type 4 in the third and fifth bit position. Then the possible codeword pairs with DF decoding are

$$(\underline{x}_A, \underline{x}_B)_{DF} \in \left\{ \begin{array}{ccccc|c|ccccc} 1 & 0 & 0 & 1 & 1 & , & 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & , & 0 & 1 & 0 & 1 & 1 \end{array} \right\}. \tag{4.9}$$

Since $\underline{x}_A$ and $\underline{x}_B$ cannot be computed exactly, the maximum likelihood DF decoder fails.

The JCF decoder considers the possible set of codeword triplets $(\underline{x}_A, \underline{x}_B, \underline{x}_R)$ and succeeds if all possible triplets correspond to the same $\underline{x}_R$. Following the same

steps as for the DF decoder we have the following triplets

$$(\underline{x}_A, \underline{x}_B, \underline{x}_R) \in \left\{ \begin{array}{ccccccccccccccc} 1 & 0 & 0 & 1 & 1 & , & 0 & 1 & 1 & 1 & 0 & , & 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & , & 0 & 1 & 0 & 1 & 1 & , & 1 & 1 & 1 & 0 & 1 \end{array} \right\}.$$

$$\Rightarrow \underline{x}_R = \left[\begin{array}{ccccc} 1 & 1 & 1 & 0 & 1 \end{array} \right] \tag{4.10}$$

Since $\underline{x}_R$ is represented uniquely, the maximum likelihood JCF decoder succeeds.

Notice that in this example JCF decoding succeeds while CF and DF both fail. Since nodes A and B use an identical linear codebook and the TWEMAC is a binary-input memoryless channel, this example may seem to contradict our results from Chapter III where we have shown that JCF does not improve the computation rates for identical linear codebooks generated uniformly at random. This example suggests that codes or code ensembles may exist for which JCF can provide substantial gains. Note that in this example, if we were to permute the order of the channel states (e.g. $[2, 3, 4, 5, 4] \rightarrow [4, 5, 4, 2, 3]$) then CF, DF, and JCF may all succeed. For the remainder of this chapter we focus our analysis on JCF message passing decoding.

IV.3. Message Passing Framework

A JCF decoder utilizes the parity check constraints for the three codewords $\underline{x}_A, \underline{x}_B, \underline{x}_R \in \mathcal{C}$ for decoding. All three of these constraint sets can be jointly expressed by an Extended Tanner Graph (ETG) as shown in Fig. 19. In the following, we derive the check and variable node processing rules for JCF message passing decoding for the TWEMAC. Particularly, we show that the performance of a message passing decoder can be evaluated by restricting our attention to the 5 message types which define the channel.

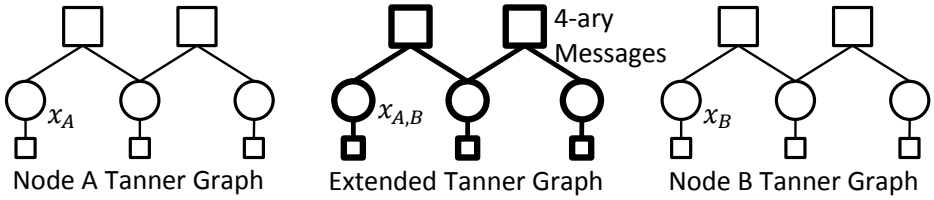


Fig. 19.: Extended Tanner graph when nodes A and B use a length 3 repetition code.

In the ETG, each variable node

$$x_{A,B}[n] = \begin{bmatrix} x_A[n] \\ x_B[n] \end{bmatrix}, \; n \in \{1, ..., N\} \tag{4.11}$$

takes a value from $\{0,1\}^2$. We will drop the subscript and use the shorthand $x[n] = x_{A,B}[n]$ to simplify many of the following expressions. The edges carry 4-ary messages which are local estimates of $x[n]$ of the form

$$\mu_n = \begin{bmatrix} P(X[n] = [00]^T | \underline{Y}_R) \\ P(X[n] = [01]^T | \underline{Y}_R) \\ P(X[n] = [10]^T | \underline{Y}_R) \\ P(X[n] = [11]^T | \underline{Y}_R) \end{bmatrix}. \tag{4.12}$$

By the standard message passing rules in [38], the initial message from the channel is defined by the 4-ary vector

$$\mu_{ch,n} = \begin{bmatrix} P(Y_R[n] | X[n] = [00]^T) \\ P(Y_R[n] | X[n] = [01]^T) \\ P(Y_R[n] | X[n] = [10]^T) \\ P(Y_R[n] | X[n] = [11]^T) \end{bmatrix}. \tag{4.13}$$

We will refer to elements of a message μ_n according to the associated value of $x[n] \in$

$\{0,1\}^2$. For example,

$$\mu_{ch,n}(01) = P(Y_R[n]|X[n] = [01]^T). \tag{4.14}$$

According to (4.1) and (4.13) for any TWEMAC, we can say that the initial message from the channel must be in the set

$$\mu_{ch} \in \left\{ \begin{array}{ccc} \left[\frac{P_{ch}^1}{4}, \frac{P_{ch}^1}{4}, \frac{P_{ch}^1}{4}, \frac{P_{ch}^1}{4}\right]^T & \left[\frac{P_{ch}^2}{2}, \frac{P_{ch}^2}{2}, 0, 0\right]^T & \left[0, 0, \frac{P_{ch}^2}{2}, \frac{P_{ch}^2}{2}\right]^T \\ \left[\frac{P_{ch}^3}{2}, 0, \frac{P_{ch}^3}{2}, 0\right]^T & \left[0, \frac{P_{ch}^3}{2}, 0, \frac{P_{ch}^3}{2}\right]^T & \left[\frac{P_{ch}^4}{2}, 0, 0, \frac{P_{ch}^4}{2}\right]^T \quad \left[0, \frac{P_{ch}^4}{2}, \frac{P_{ch}^4}{2}, 0\right]^T \\ \left[\frac{P_{ch}^5}{4}, 0, 0, 0\right]^T & \left[0, \frac{P_{ch}^5}{4}, 0, 0\right]^T & \left[0, 0, \frac{P_{ch}^5}{4}, 0\right]^T \quad \left[0, 0, 0, \frac{P_{ch}^5}{4}\right]^T \end{array} \right\}$$

These correspond to the five message types $\mathcal{T}$ and the associated parametrization $P_{ch}^i(\epsilon), i \in \mathcal{T}$. Recall from [38] that the normalization of a message provides no information about the value of any $x[n]$. Thus we can consider a simplified message alphabet $\mathcal{U}$ in which each component $\mu(x),\ x \in \{0,1\}^2$ is either 0 or 1, and there are 1, 2, or 4 non-zero entries in each μ. Matching the messages in the set $\mathcal{U}$ with their appropriate types from the set $\mathcal{T}$, we see that

$$\begin{aligned} \tau = 1 &\Leftrightarrow \mu \in \{[1111]^T\} \\ \tau = 2 &\Leftrightarrow \mu \in \{[1100]^T, [0011]^T\} \\ \tau = 3 &\Leftrightarrow \mu \in \{[1010]^T, [0101]^T\} \\ \tau = 4 &\Leftrightarrow \mu \in \{[1001]^T, [0110]^T\} \\ \tau = 5 &\Leftrightarrow \mu \in \{[1000]^T, [0100]^T, [0010]^T, [0001]^T\} \end{aligned} \tag{4.15}$$

Note that each component $\mu_n(x),\ x \in \{0,1\}^2$ no longer refers to a probability in the strict sense. Rather, each $\mu_n(x)$ is a binary indicator of whether it is possible that

$X[n] = x$ given that the sequence $\underline{Y}_R = \underline{y}_R$ was received at the relay.

We want to develop the notion that to predict the performance of an LDPC code or ensemble for the class of TWEMACs, it is enough to track the types $\tau \in \mathcal{T}$ of the messages sent by the variable and check nodes. This requires that we derive the variable and check node operations $VAR(\cdot)$ and $CHK(\cdot)$ for a variable and check node respectively. We use this notation $VAR(\cdot)$ or $CHK(\cdot)$ to describe the variable or check node operations on the sets $\mathcal{U}$, $\mathcal{T}$, and later on the set of type distribution vectors. We first need to verify that these operations are closed on the set $\mathcal{U}$.

IV.3.1. Variable Node Processing

For variable nodes, the message processing rules from [38] state that the outgoing message along each edge is the componentwise product of the incoming messages. Since the channel never makes an error, it is impossible for the outgoing message to be the all zero vector (i.e. the incident information to a variable node must agree). For example, if the correct value of a variable node $x[n] = [10]^T$, all of the incoming messages must be from the set $\{[1111]^T, [0011]^T, [1010]^T, [0110]^T, [0010]^T\} \subset \mathcal{U}$. Notice that there is one message associated with each $\tau \in \mathcal{T}$ and that the outgoing message must also come from this subset of $\mathcal{U}$. We conclude that $VAR(\cdot)$ is closed on the set of messages $\mathcal{U}$ and is subsequently closed on the set of types $\mathcal{T}$.

A degree d_v variable node is connected to d_v check nodes and one function node associated with the channel observation. Thus, each outgoing message from a degree d_v variable node is a function of d_v input messages $\mu_1, ..., \mu_{d_v}$. By the associativity and commutativity of multiplication, the output of the variable node operation on $\mathcal{U}$ satisfies

$$\mu^{out} = VAR(\mu_1, VAR(\mu_2, ...VAR(\mu_{d_v-1}, \mu_{d_v}))). \tag{4.16}$$

It is therefore constructive to create an operator table for $VAR(\cdot)$ with two input types $\tau_1, \tau_2 \in \mathcal{T}$. This can be accomplished by exhaustively evaluating $VAR(\mu_1, \mu_2)$ for each $\mu_1, \mu_2 \in \mathcal{U}$ associated with message types $\tau_1, \tau_2 \in \mathcal{T}$. Since the channel never makes errors, each μ_1, μ_2 pair must have at least one 1 in common. The result is Table I.

VAR		τ_1				
		1	2	3	4	5
τ_2	1	1	2	3	4	5
	2	2	2	5	5	5
	3	3	5	3	5	5
	4	4	5	5	4	5
	5	5	5	5	5	5

Table I.: Variable node operator table with respect to types $\mathcal{T}$.

For a variable node of degree d_v, if the input messages have types $\tau_1, ..., \tau_{d_v}$, then the output message type, τ^{out} is found by recursively applying $VAR(\cdot)$ according to

$$\tau^{out} = VAR(\tau_1, VAR(\tau_2, ...VAR(\tau_{d_v-1}, \tau_{d_v}))). \tag{4.17}$$

The commutativity and associativity of $VAR(\cdot)$ is maintained on the set $\mathcal{T}$.

IV.3.2. Check Node Processing

Consider a check node of degree d_c connected to variable nodes $x[n],\ n \in \{1, ..., d_c\}$. In order for $\underline{x}_A, \underline{x}_B, \underline{x}_R \in \mathcal{C}$, these variable nodes must satisfy

$$\begin{aligned} x_A[1] \oplus ... \oplus x_A[d_c] &= 0 \\ x_B[1] \oplus ... \oplus x_B[d_c] &= 0 \\ x_R[1] \oplus ... \oplus x_R[d_c] &= 0. \end{aligned} \tag{4.18}$$

These constraints are equivalent to

$$x[1] \oplus ... \oplus x[d_c] = [00]^T \tag{4.19}$$

where the $\oplus$ operation is applied element-wise. Note that if the constraints for $\underline{x}_A, \underline{x}_B \in \mathcal{C}$ are satisfied for all the parity checks, then they are satisfied for $\underline{x}_R$. The identical linear code $\mathcal{C}$ ensures that the check constraints for $\underline{x}_A, \underline{x}_B$ can have identical edge connections.

The message processing rules from [38] state that the outgoing message on edge d_c is given by

$$\mu^{out}(x[d_c]) = \sum_{\sim x[d_c]} 1_{\{x[d_c] = \bigoplus_{i=1}^{d_c-1} x[i]\}} \prod_{i=1}^{d_c-1} \mu_i(x[i]). \tag{4.20}$$

The shorthand notation $\sim x[d_c]$ indicates that the summation is over the values that can be taken by $x[1], ..., x[d_c - 1] \in \{0, 1\}^2$. For a degree 3 check node, this is

$$\begin{aligned} \mu^{out}(x[3]) &= \sum_{\sim x[3]} 1_{\{x[3] = x[1] \oplus x[2]\}} \mu_1(x[1]) \mu_2(x[2]) \\ &= \sum_{x[1], x[2] \in \{0,1\}^2} \mu_1(x[1]) \mu_2(x[2] \oplus x[3]). \end{aligned} \tag{4.21}$$

This defines the operation $CHK(\cdot)$ on $\mathcal{U}$. We can verify that $CHK(\cdot)$ is closed on the set $\mathcal{U}$, by exhaustively evaluating $CHK(\mu_1, \mu_2)$ with all pairs $(\mu_1, \mu_2) \in \mathcal{U} \times \mathcal{U}$.

For a degree d_c check node, the output message μ^{out} is computed from the input messages $\mu_1, ..., \mu_{d_c-1}$ by recursively applying $CHK(\cdot)$ as

$$\mu^{out} = CHK(\mu_1, CHK(\mu_2, ...CHK(\mu_{d_c-2}, \mu_{d_c-1}))). \tag{4.22}$$

This follows from the commutativity and associativity of multiplication, addition, and $\oplus$. The degree d_c check node operation is also closed on $\mathcal{U}$ by induction.

Similar to the variable node operation, we define $CHK(\cdot)$ on the set $\mathcal{T}$ by exhaustively evaluating $CHK(\mu_1, \mu_2)$ for pairs $(\mu_1, \mu_2) \in \mathcal{U} \times \mathcal{U}$ and determining the input and output message types. The result is given in Table II.

CHK		τ_1^{in}				
		1	2	3	4	5
τ_2^{in}	1	1	1	1	1	1
	2	1	2	1	1	2
	3	1	1	3	1	3
	4	1	1	1	4	4
	5	1	2	3	4	5

Table II.: Check node operator table with respect to types $\mathcal{T}$.

For a check node of degree d_c, if the input messages have types $\tau_1, ..., \tau_{d_c-1}$, then the output message type, τ^{out} is found by recursively applying $CHK(\cdot)$ according to

$$\tau^{out} = CHK(\tau_1, CHK(\tau_2, ...CHK(\tau_{d_c-2}, \tau_{d_c-1}))). \tag{4.23}$$

The commutativity and associativity of $CHK(\cdot)$ is also maintained on the set $\mathcal{T}$.

IV.3.3. Type Distribution Processing Rules

We have defined the input output relationship for variable and check nodes on the sets $\mathcal{U}$ and $\mathcal{T}$. Here, these rules are extended to the set of 5-ary type distributions. Similar to (4.2), we define the type distribution of messages from the variable to check (check to variable) nodes during iteration ℓ as $\underline{P}_{vc}^{(\ell)}$ ($\underline{P}_{cv}^{(\ell)}$). We will index these pmfs according to

$$\underline{P}_{vc}^{(\ell)} = \begin{bmatrix} P_{vc}^{1,(\ell)} \\ P_{vc}^{2,(\ell)} \\ P_{vc}^{3,(\ell)} \\ P_{vc}^{4,(\ell)} \\ P_{vc}^{5,(\ell)} \end{bmatrix} \qquad \underline{P}_{cv}^{(\ell)} = \begin{bmatrix} P_{cv}^{1,(\ell)} \\ P_{cv}^{2,(\ell)} \\ P_{cv}^{3,(\ell)} \\ P_{cv}^{4,(\ell)} \\ P_{cv}^{5,(\ell)} \end{bmatrix} \tag{4.24}$$

which is consistent with our notation for $\underline{P}_{ch}$ in (4.2).

The probability that an output message from a variable or check node has type $i \in \mathcal{T}$ can be explicitly computed by evaluating the sum of the probabilities of input message types which result in an output message of type $i \in \mathcal{T}$. Consider a variable node of degree d_v whose messages have input type distributions $\underline{P}_{cv,1}^{(\ell)}, ..., \underline{P}_{cv,d_v-1}^{(\ell)}$ and $\underline{P}_{ch}$ from the check nodes and the channel respectively. Following the rules for evaluating output message types for variable node (4.17), construct the set

$$\mathcal{T}_{d_v,i} = \{(j_1, ..., j_{d_v}) \in \mathcal{T}^{d_v} | i = VAR(j_1, VAR(j_2, ...VAR(j_{d_v-1}, j_{d_v})))\}. \tag{4.25}$$

Then, the probability that the output message has type $i \in \mathcal{T}$ is expressed

$$P(\tau^{out} = i) = \sum_{(j_1,...,j_{d_v}) \in \mathcal{T}_{d_v,i}} P_{cv,1}^{j_1,(\ell)} ... P_{cv,d_v-1}^{j_{d_v-1},(\ell)} P_{ch}^{j_{d_v}}(\epsilon). \tag{4.26}$$

It is helpful to think of the variable node operation on the set of type distributions as a series of state transitions in a trellis. This is illustrated for a degree $d_v = 3$

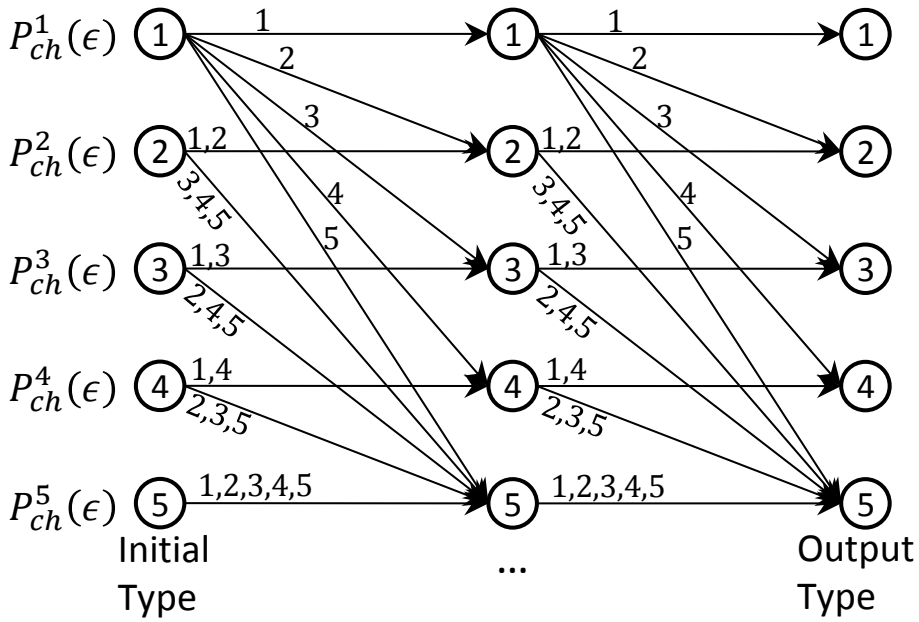


Fig. 20.: Trellis diagram for the type distribution update operation of a degree 3 variable node. Each edge in the trellis is labeled with the message types corresponding to the illustrated transition.

variable node in Fig. 20. The distribution for the initial state is given by the channel type distribution $\underline{P}_{ch}$. Then, there are two input messages from check nodes whose type distributions are applied in the form of transitions through the trellis. These transitions are labeled according to the message types $\tau \in \mathcal{T}$ which correspond to the labeled transition. The probability associated with each transition is therefore the sum of the probabilities of the corresponding types (e.g. the probability of transition from type 3 to 5 is $P_{cv}^{2,(\ell)} + P_{cv}^{4,(\ell)} + P_{cv}^{5,(\ell)}$).

With this trellis structure in mind, it is natural to express this type distribution update operation as a series of multiplications of state-transition matrices. For a degree 2 variable node, we have

$$\underline{P}_{vc}^{(\ell)} = \mathbf{P}_{cv}^{(\ell)} \underline{P}_{ch}. \tag{4.27}$$

The $(i,j)_{th}$ element of the matrix $\mathbf{P}_{cv}^{(\ell)}$ is defined

$$\left[\mathbf{P}_{cv}^{(\ell)}\right]_{i,j} = P(j \to i) = \sum_{\{k \in \mathcal{T} \mid i = VAR(k,j)\}} P_{cv}^{k,(\ell)} \tag{4.28}$$

where the elements of the summation follow the rules given in Table I. The notation $P(j \to i)$ refers to the probability of a transition from type j to type i as depicted in Fig. 20.

This is equivalent to

$$\mathbf{P}_{cv}^{(\ell)} = \begin{bmatrix} P_{cv}^{1,(\ell)} & 0 & 0 & 0 & 0 \\ P_{cv}^{2,(\ell)} & P_{cv}^{1+2,(\ell)} & 0 & 0 & 0 \\ P_{cv}^{3,(\ell)} & 0 & P_{cv}^{1+3,(\ell)} & 0 & 0 \\ P_{cv}^{4,(\ell)} & 0 & 0 & P_{cv}^{1+4,(\ell)} & 0 \\ P_{cv}^{5,(\ell)} & P_{cv}^{3+4+5,(\ell)} & P_{cv}^{2+4+5,(\ell)} & P_{cv}^{2+3+5,(\ell)} & 1 \end{bmatrix}$$

with the shorthand $P_{cv}^{i+j,(\ell)} = P_{cv}^{i,(\ell)} + P_{cv}^{j,(\ell)}$.

More generally, consider a variable node with degree $d_v \geq 2$ and input type distributions $\underline{P}_{cv,1}^{(\ell)}, ..., \underline{P}_{cv,d_v-1}^{(\ell)}$ from check nodes and type distribution $\underline{P}_{ch}$ from the channel. Then the type distribution for the output message is given by

$$\underline{P}_{vc}^{(\ell)} = \mathbf{P}_{cv,1}^{(\ell)}, ..., \mathbf{P}_{cv,d_v-1}^{(\ell)} \underline{P}_{ch}. \tag{4.29}$$

To see this, consider $d_v = 3$ with the two type update matrices $\mathbf{P}_{cv,1}^{(\ell)}$ and $\mathbf{P}_{cv,2}^{(\ell)}$.

Each element of the product of these matrices is given by

$$\begin{aligned}\left[\mathbf{P}_{cv,1}^{(\ell)}\mathbf{P}_{cv,2}^{(\ell)}\right]_{i,j} &= \sum_k [\mathbf{P}_{cv,1}^{(\ell)}]_{i,k}[\mathbf{P}_{cv,2}^{(\ell)}]_{k,j} \\ &= \sum_k \sum_{\{j_1\in\mathcal{T}|i=VAR(j_1,k)\}} P_{cv,1}^{j_1,(\ell)} \sum_{\{j_2\in\mathcal{T}|k=VAR(j_2,j)\}} P_{cv,2}^{j_2,(\ell)} \\ &= \sum_k \sum_{\{j_1,j_2\in\mathcal{T}|i=VAR(j_1,k),k=VAR(j_2,j)\}} P_{cv,1}^{j_1,(\ell)} P_{cv,2}^{j_2,(\ell)} \\ &= \sum_{\{j_1,j_2\in\mathcal{T}|i=VAR(j_1,VAR(j_2,j))\}} \underline{P}_{cv,1}^{j_1,(\ell)} P_{cv,2}^{j_2,(\ell)}. \end{aligned} \tag{4.30}$$

This is the probability that the output message transitions from a message of type j to i because of the input type distributions $\underline{P}_{cv,1}^{(\ell)}$ and $\underline{P}_{cv,2}^{(\ell)}$. Multiplying by $\underline{P}_{ch}$, we obtain

$$\begin{aligned}\left[\mathbf{P}_{cv,1}^{(\ell)}\mathbf{P}_{cv,2}^{(\ell)}\underline{P}_{ch}\right]_i &= \sum_{\{j_1,j_2,j\in\mathcal{T}|i=VAR(j_1,VAR(j_2,j))\}} P_{cv,1}^{j_1,(\ell)} P_{cv,2}^{j_2,(\ell)} P_{ch}^{j}(\epsilon) \\ &= \sum_{(j_1,j_2,j)\in\mathcal{T}_{3,i}} P_{cv,1}^{j_1,(\ell)} P_{cv,2}^{j_2,(\ell)} P_{ch}^{j}(\epsilon) \\ &= P(\tau^{out}=i). \end{aligned} \tag{4.31}$$

The equivalence of (4.26) and (4.31) for general d_v follows by induction.

Note that

$$\begin{aligned}\left[\mathbf{P}_{cv,1}^{(\ell)}\mathbf{P}_{cv,2}^{(\ell)}\right]_{i,j} &= \sum_{\{j_1,j_2\in\mathcal{T}|i=VAR(j_1,VAR(j_2,j))\}} P_{cv,1}^{j_1,(\ell)} P_{cv,2}^{j_2,(\ell)} \\ &= \sum_{\{j_1,j_2\in\mathcal{T}|i=VAR(j_2,VAR(j_1,j))\}} P_{cv,1}^{j_1,(\ell)} P_{cv,2}^{j_2,(\ell)} \\ &= \left[\mathbf{P}_{cv,2}^{(\ell)}\mathbf{P}_{cv,1}^{(\ell)}\right]_{i,j}. \end{aligned} \tag{4.32}$$

The commutativity and associativity of the state transition matrix associated with the variable node operation follows directly from the commutativity and associativity

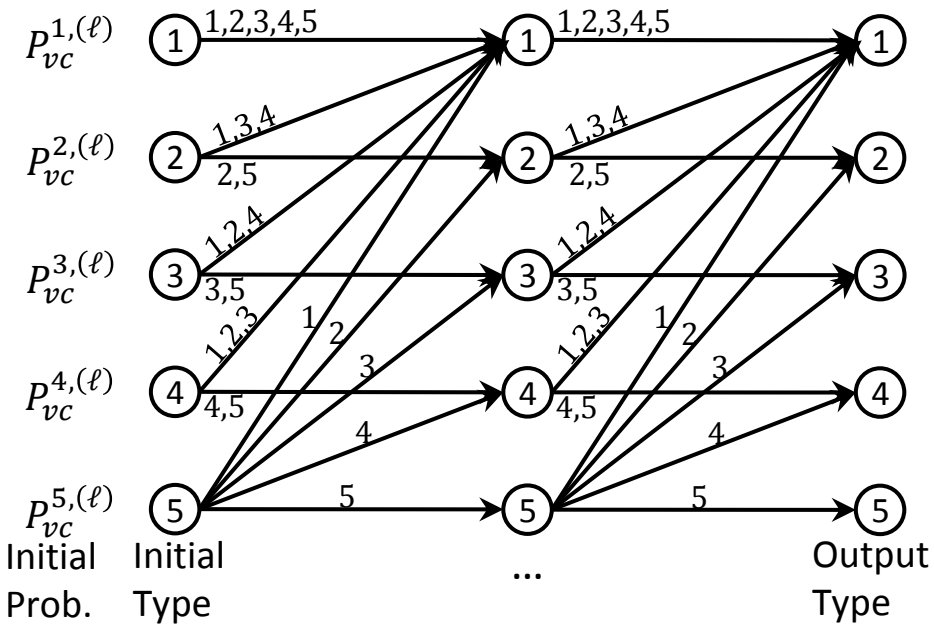


Fig. 21.: Trellis diagram type distribution update for a degree 4 check node. Each edge in the trellis is labeled with the message types corresponding to the illustrated transition.

of $VAR(\cdot)$ on the set $\mathcal{T}$. It can be shown by induction that this also holds for general d_v.

For a check node of degree d_c with input type distributions $\underline{P}_{vc,1}^{(\ell-1)}, ..., \underline{P}_{vc,d_c-1}^{(\ell-1)}$ we define the set

$$\mathcal{T}_{d_c,i} = \{(j_1, ..., j_{d_c-1}) \in \mathcal{T}^{d_c-1} | i = CHK(j_1, CHK(j_2, ...CHK(j_{d_c-2}, j_{d_c-1})))\}. \tag{4.33}$$

Then the probability that the output message from the check node has type $i \in \mathcal{T}$ is expressed

$$P(\tau^{out} = i) = \sum_{(j_1,...,j_{d_c-1}) \in \mathcal{T}_{d_c,i}} P_{vc,1}^{j_1,(\ell-1)} ... P_{vc,d_c-1}^{j_{d_c-1},(\ell-1)}. \tag{4.34}$$

The trellis structure for a degree 4 check node is depicted in Fig. 21 where the

transitions are arranged according to Table II.

The derivation of the matrix form for the check node operation on type distributions is similar to the variable node derivation. The state transition matrix $\mathbf{P}_{vc}^{(\ell)}$ is defined

$$\mathbf{P}_{vc}^{(\ell)} = \begin{bmatrix} 1 & P_{vc}^{1+3+4,(\ell)} & P_{vc}^{1+2+4,(\ell)} & P_{vc}^{1+2+3,(\ell)} & P_{vc}^{1,(\ell)} \\ 0 & P_{vc}^{2+5,(\ell)} & 0 & 0 & P_{vc}^{2,(\ell)} \\ 0 & 0 & P_{vc}^{3+5,(\ell)} & 0 & P_{vc}^{3,(\ell)} \\ 0 & 0 & 0 & P_{vc}^{4+5,(\ell)} & P_{vc}^{4,(\ell)} \\ 0 & 0 & 0 & 0 & P_{vc}^{5,(\ell)} \end{bmatrix}.$$

For a check node of degree d_c with input type distributions $\underline{P}_{vc,1}^{(\ell-1)}, ..., \underline{P}_{vc,d_c-1}^{(\ell-1)}$, the type distribution for the output message is given by

$$\underline{P}_{cv}^{(\ell)} = \mathbf{P}_{vc,1}^{(\ell-1)} ... \mathbf{P}_{vc,d_c-2}^{(\ell-1)} \underline{P}_{vc,d_c-1}^{(\ell-1)}. \tag{4.35}$$

As with the variable node case, the type distribution update operation for check nodes is invariant to the order of matrix multiplications. The vector $\underline{P}_{vc,d_c-1}^{(\ell-1)}$ provides an initial type distribution for the transition matrices associated with the other input type distributions.

IV.4. Type Distribution Evolution

In this section, we characterize the performance of several LDPC ensembles using JCF message passing decoding on a TWEMAC. For the normal binary erasure channel (BEC), the asymptotic (in length) performance of an LDPC ensemble is often analyzed using density evolution analysis [38]. The main idea in density evolution is to compute the probability that a message from a variable to check node (or check to variable node) is erased during each iteration of message passing. Analogous density evolution for the BEC we track the distribution of message types for each iteration.

Hence, we refer to this performance characterization as type distribution evolution.

To be consistent with much of the literature, each variable node is initialized (iteration 0) with the observation from the channel and forwards this message to all connected check nodes. The check nodes process these messages and forward their estimates back to the variable nodes. This process continues until the iterations reach a fixed point. As in [38], we work under the assumption that, for a finite number of iterations and for some N large enough, the local graph around any variable node is a tree with high probability. This allows us to assume the statistical independence of multiple input messages to a single variable or check node.

IV.4.1. Type Distribution Evolution for Regular LDPC Ensembles

For a (d_v, d_c) regular ensemble, the incoming messages to a variable or check node are distributed according to $\underline{P}_{cv}^{(\ell)}$ or $\underline{P}_{vc}^{(\ell)}$ respectively. Since the first message from the variable to check nodes is the message from the channel, the type distribution evolution is initialized by

$$\begin{aligned} \underline{P}_{vc}^{(0)} &= \underline{P}_{ch} \\ \underline{P}_{cv}^{(1)} &= \left(\mathbf{P}_{vc}^{(0)}\right)^{d_c-2} \underline{P}_{vc}^{(0)}. \end{aligned} \tag{4.36}$$

For every iteration $\ell \geq 2$, the type distribution evolution is fully characterized by

$$\underline{P}_{cv}^{(\ell)} = \left(\mathbf{P}_{vc}^{(\ell-1)}\right)^{d_c-2} \underbrace{\left(\mathbf{P}_{cv}^{(\ell-1)}\right)^{d_v-1} \underline{P}_{ch}}_{\underline{P}_{vc}^{(\ell-1)}}. \tag{4.37}$$

If we allow the decoder to complete ℓ_{max} iterations, the type distribution at the output of the decoder is

$$\underline{P}_{out}^{(\ell_{max})} = \left(\mathbf{P}_{cv}^{(\ell_{max})}\right)^{d_v} \underline{P}_{ch}. \tag{4.38}$$

The objective of the relay is to recover the codeword $\underline{x}_R$. Therefore, the bitwise

probability of successful computation after ℓ_{max} iterations is defined by

$$P_{dec} = P_{out}^{4,(\ell_{max})} + P_{out}^{5,(\ell_{max})}. \tag{4.39}$$

This is because the value of each $x_R[n]$ is known if the decoder output is either type 4 or 5. We define ϵ_{thesh} as the largest ϵ such that reliable decoding is possible for an LDPC ensemble (i.e. $P_{dec} \xrightarrow{\ell_{max}\to\infty} 1$).

We have proven analytically that optimal JCF decoding cannot outperform optimal CF and DF decoding for random linear codebooks. However, the ϵ_{thresh} obtained from type distribution evolution analysis for regular ensembles may be different for DF, CF, or JCF message passing. The type distribution evolution for CF is performed with the exact same processing rules, however, we make a change to the channel model as

$$\underline{P}_{ch,CF} = \begin{bmatrix} P_{ch}^1(\epsilon) + P_{ch}^2(\epsilon) + P_{ch}^3(\epsilon) \\ 0 \\ 0 \\ P_{ch}^4(\epsilon) + P_{ch}^5(\epsilon) \\ 0 \end{bmatrix}. \tag{4.40}$$

Therefore, the message passing decoder operates on estimates $P(X_R|Y_R)$ as discussed previously. The type distribution evolution becomes exactly equivalent to density evolution for a BEC in this case.

The type distribution evolution for DF is also performed with the same processing rules, however we define

$$P_{dec,DF} = P_{out}^{5,(\ell_{max})} \tag{4.41}$$

as the probability for successful decoding.

IV.4.2. Type Distribution Evolution for Spatially Coupled Ensembles

Here we characterize the performance of the (d_v, d_c, L, w) spatially coupled LDPC ensemble which is defined rigorously in [40]. Since node A and B each use the same spatially coupled LDPC code, the ETG used for decoding at the relay is also spatially coupled. The common method to achieve the capacity of a given channel is to find some distribution on the variable and check node degrees such that a message passing decoder can reliably decode at rates near capacity. It has recently been discovered that for many spatially coupled LDPC ensembles a message passing decoder can nearly achieve the performance of an optimal decoder [39], [40]. This holds for a large class of binary-input channels [42], [41].

In the (d_v, d_c, L, w) ensemble, M variable nodes are placed in each position $i \in \{-L, ..., L\}$ so that the codeword length is $N = (2L+1)M$. Each variable and check node has degree d_v and d_c respectively. The ensemble is constructed by uniformly and independently connecting the d_v (d_c) edges from a variable (check) node at position i to check (variable) nodes at positions $\{i, ..., i+w-1\}$ ($\{i-w+1, ..., i\}$). There are Md_v edges in each position which requires that $\frac{d_v}{d_c}M$ check nodes be placed at positions $i \in \{-L, ..., L+w-1\}$. All check node connections to variable nodes outside of positions $\{-L, ..., L\}$ are connected to pseudo variable nodes whose value is fixed to $x_{A,B} = [00]^T$. This decreases the effective degree of these check nodes.

The nominal rate for a (d_v, d_c, L, w) ensemble is defined

$$\mathcal{R}(d_v, d_c, L, w) = \left(1 - \frac{d_v}{d_c}\right) - \frac{d_v}{d_c}\frac{w + 1 - 2\sum_{i=0}^{w}(\frac{i}{w})^{d_c}}{2L+1}. \tag{4.42}$$

This is a lower bound on the actual rate, however, it approaches the rate of the regular ensemble $\mathcal{R}(d_v, d_c) = 1 - \frac{d_v}{d_c}$ as $L \to \infty$ for a fixed w. This ensemble is primarily useful for information-theoretic analysis.

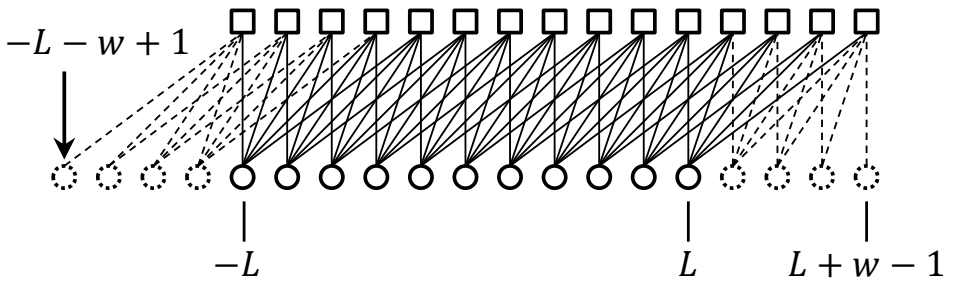


Fig. 22.: Protograph diagram for a $(d_v, d_c, L = 5, w = 5)$ code ensemble. d_v and d_c remain indeterminant because this diagram is valid for many values for d_v and d_c. All protograph edge connections represent potential edge connections for a random code in this ensemble.

A protograph diagram for a $(d_v, d_c, L = 5, w = 5)$ ensemble is shown in Fig. 22 with variable and check nodes depicted by circles and squares respectively. The dashed edges depict potential connections to pseudo variable nodes which will always give a message of type 5. Note that each protograph edge in Fig. 22 represents a potential edge connection for the connected protograph nodes. Particularly, for a fixed code in this ensemble, a specific check node at position $-L$ may only have connections to pseudo variable nodes.

To characterize the average performance of the (d_v, d_c, L, w) ensemble, we need to track the type distributions for the variable and check nodes in each position separately. Also, the uniformly distributed edge spreading means that the effective input type distribution to each variable (check) node is the average of the outgoing type distributions from each potentially connected check (variable) node. More rigorously, we define the type distribution for outgoing messages from a variable node at position i as $\underline{P}_{vc,i}^{(\ell)}$, and we define the corresponding *effective* input type distribution to a check node at position i as $\underline{\widetilde{P}}_{vc,i}^{(\ell)}$. We similarly define $\underline{P}_{cv,i}^{(\ell)}$ and $\underline{\widetilde{P}}_{cv,i}^{(\ell)}$. The

effective input type distributions are

$$\widetilde{\underline{P}}_{cv,i}^{(\ell)} = \frac{1}{w} \sum_{j=i}^{i+w-1} \underline{P}_{cv,j}^{(\ell)} \ \forall \ i \in \{-L, ..., L\}$$

$$\widetilde{\underline{P}}_{vc,i}^{(\ell)} = \frac{1}{w} \sum_{j=i-w+1}^{i} \underline{P}_{vc,j}^{(\ell)} \ \forall \ i \in \{-L, ..., L+w-1\}. \tag{4.43}$$

The type distribution evolution for the (d_v, d_c, L, w) ensemble is initialized by

$$\underline{P}_{vc,i}^{(0)} = \begin{cases} \underline{P}_{ch} & \forall \ i \in \{-L, ..., L\} \\ [00001]^T & \forall \ i \notin \{-L, ..., L\} \end{cases}. \tag{4.44}$$

Then, the type distribution evolution for this ensemble is fully characterized by

$$\begin{aligned} &\text{FOR } \ell = 1, ..., \ell_{max} \\ &\underline{P}_{cv,i}^{(\ell)} = \left(\widetilde{\mathbf{P}}_{vc,i}^{(\ell-1)}\right)^{d_c-2} \widetilde{\underline{P}}_{vc,i}^{(\ell-1)} \ , \ i \in \{-L, ..., L+w-1\} \\ &\underline{P}_{vc,i}^{(\ell)} = (\widetilde{\mathbf{P}}_{cv,i}^{(\ell)})^{d_v-1} \underline{P}_{ch} \qquad , \ i \in \{-L, ..., L\} \end{aligned} \tag{4.45}$$

where (4.43) is used to determine $\widetilde{\underline{P}}_{cv,i}^{(\ell)}$ and $\widetilde{\underline{P}}_{vc,i}^{(\ell)}$ as needed.

Similar to (4.38), if we allow ℓ_{max} iterations, the output type distribution for a variable node at position i for this ensemble is

$$\underline{P}_{out,i}^{(\ell_{max})} = \left(\widetilde{\mathbf{P}}_{cv,i}^{(\ell_{max})}\right)^{d_v} \underline{P}_{ch}. \tag{4.46}$$

Similar to (4.39), the bitwise probability of successful computation for a variable at position i is

$$P_{dec,i} = P_{out,i}^{4,(\ell_{max})} + P_{out,i}^{5,(\ell_{max})}. \tag{4.47}$$

We define ϵ_{thresh} as the largest ϵ such that reliable decoding is possible for this LDPC ensemble (i.e. $\min_{i \in \{-L,...,L\}} P_{dec,i} \xrightarrow{\ell_{max} \to \infty} 1$). Notice, that we are already considering the case that M is very large. This type distribution analysis defines the average

performance of the ensemble defined by the parameters (d_v, d_c, L, w).

IV.4.3. Type Distribution Evolution for Spatially Coupled Protograph LDPC Ensembles

The previous ensemble is quite interesting for information-theoretic analysis because it can be shown to achieve the optimal decoding performance of asymptotically regular LDPC codes for many types of channels. Unfortunately it does not provide much assistance for the design or analysis of practical spatially coupled ensembles. The check terminated protographs which use the edge spreading techniques proposed in [50] provide a large class of practical spatially coupled ensembles to investigate. We refer to [50], [51] for a thorough discussion of spatially coupled protograph ensembles. However we include a brief description which is intentionally similar to [51] to obtain some necessary notation. Then, we derive the type distribution evolution equations for spatially coupled protograph ensembles on the TWEMAC. The density evolution equations derived in this section can also be applied directly for uncoupled protograph ensembles.

A binary protograph matrix is used to organize the edge connections of a spatially coupled protograph ensemble. A common and convenient way to construct a check terminated protograph matrix is to design w sub-matrices $\mathbf{B}_1, ..., \mathbf{B}_w$ of size

(b_c, b_v). Then, a protograph matrix can be constructed according to

$$\mathbf{B}_{[1,\ldots,L]} = \begin{bmatrix} \mathbf{B}_1 & & \\ \vdots & \ddots & \\ \mathbf{B}_w & & \\ & & \mathbf{B}_1 \\ & \ddots & \vdots \\ & & \mathbf{B}_w \end{bmatrix}_{(L+w-1)b_c \times Lb_v} . \tag{4.48}$$

The parameter L represents the number of times the protograph structure is repeated, and w represents the memory of the protograph ensemble. Thus L and w perform similar functions for spatially coupled protograph ensembles as for the (d_v, d_c, L, w) ensemble. However, there are $(L + w - 1)b_c$ check node positions and Lb_v variable node positions, each with M check or variable nodes respectively. If we choose $b_c = d_v$ and $b_v = d_c$, the nominal rate for a code with a parity check matrix of the form (4.48) is

$$\begin{aligned} \mathcal{R}_{\mathbf{B}_{[1,\ldots L]}} &= \left(1 - \frac{b_c}{b_v}\right) - \frac{w-1}{L}\frac{b_c}{b_v} \\ &= \left(1 - \frac{d_v}{d_c}\right) - \frac{w-1}{L}\frac{d_v}{d_c}. \end{aligned} \tag{4.49}$$

The rate asymptotically approaches

$$\lim_{L\to\infty} \mathcal{R}_{\mathbf{B}_{[1,\ldots,L]}} = \left(1 - \frac{b_c}{b_v}\right) = \left(1 - \frac{d_v}{d_c}\right). \tag{4.50}$$

A parity check matrix for a spatially coupled protograph code is formed from $\mathbf{B}_{[1,\ldots,L]}$ in (4.48) by replacing each 1 with an $M \times M$ permutation matrix and each 0 by an $M \times M$ all zero matrix. We will limit our focus to protographs which can be expressed as in (4.48).

When we discuss type distribution evolution analysis for a protograph ensemble,

we are referring to the expected asymptotic performance of codes in the $\mathbf{B}_{[1,...,L]}$ ensemble as $M \to \infty$. The type distribution of the input (or output) messages to (or from) a variable or check node may not be the same for each edge. It is necessary to keep track of the type distributions for each edge individually.

Therefore, we define $\mathcal{N}_v = \{1, ..., Lb_v\}$ and $\mathcal{N}_c = \{1, ..., (L + w - 1)b_c\}$ as the sets of protograph variable and check node positions respectively. We denote the protograph edge $[i, j]$ which connects a check node at position $i \in \mathcal{N}_c$ to a variable node at position $j \in \mathcal{N}_v$. Let $\mathcal{E}$ be the set of all such edges. Define the type distribution for messages from a variable node to a check node along edge $[i, j]$ and during iteration ℓ as $\underline{P}^{(\ell)}_{vc,j\to i}$. Similarly define the check to variable node message returning on the same edge as $\underline{P}^{(\ell)}_{cv,i\to j}$. Then define $\mathcal{E}_{c,i}$ as the set of all edges connected to the protograph check node at position $i \in \mathcal{N}_c$. Similarly, $\mathcal{E}_{v,j}$ is the set of all edges connected to a protograph variable node at position $j \in \mathcal{N}_v$.

Consider the protograph check node at position $i \in \mathcal{N}_c$. Select two distinct edges $[i, j_1], [i, j_2] \in \mathcal{E}_{c,i}$. The type distribution for the output message from i to the variable node at position $j_1 \in \mathcal{N}_v$ is given by

$$\underline{P}_{cv,i\to j_1} = \left(\prod_{[i,j]\in\mathcal{E}_{c,i}\setminus\{[i,j_1],[i,j_2]\}} \mathbf{P}_{vc,j\to i} \right) \underline{P}_{vc,j_2\to i} \tag{4.51}$$

where the product term is the matrix product. Recall that the order of matrix multiplication and the selection of $[i, j_1], [i, j_2] \in \mathcal{E}_{c,i}$ may be arbitrary. Use the notation

$$\underline{P}_{cv,i\to j} = CHK_{[i,j']\in\mathcal{E}_{c,i}\setminus\{[i,j]\}}(\underline{P}_{vc,j'\to i}) \tag{4.52}$$

to denote the operation defined in (4.51).

For a protograph variable node at position $j \in \mathcal{N}_v$, the output on an edge

$[i,j] \in \mathcal{E}_{v,j}$ is given by

$$\underline{P}_{vc,j\to i} = \left(\prod_{[i',j]\in\mathcal{E}_{v,j}\setminus\{[i,j]\}} \mathbf{P}_{cv,i'\to j} \right) \underline{P}_{ch} \tag{4.53}$$

where the product term is the matrix product. For notational consistency, use

$$\underline{P}_{vc,j\to i} = VAR_{[i',j]\in\mathcal{E}_{v,j}\setminus\{[i,j]\}}(\underline{P}_{cv,i'\to j}, \underline{P}_{ch}) \tag{4.54}$$

to denote the operation defined in (4.53).

The type distribution evolution for the protograph ensemble is initialized by

$$\underline{P}^{(0)}_{vc,j\to i} = \underline{P}_{ch} \ \forall \ [i,j] \in \mathcal{E}. \tag{4.55}$$

Then, the type distribution evolution for the protograph ensemble is fully characterized by

$$\begin{aligned} &FOR\ \ell = 1, ..., \ell_{max} \\ &\quad \underline{P}^{(\ell)}_{cv,i\to j} = CHK_{[i,j']\in\mathcal{E}_{c,i}\setminus\{[i,j]\}}(\underline{P}^{(\ell-1)}_{vc,j'\to i}) \ \forall \ [i,j] \in \mathcal{E} \\ &\quad \underline{P}^{(\ell)}_{vc,j\to i} = VAR_{[i',j]\in\mathcal{E}_{v,j}\setminus\{[i,j]\}}(\underline{P}^{(\ell)}_{cv,i'\to j}, \underline{P}_{ch}) \ \forall \ [i,j] \in \mathcal{E}. \end{aligned} \tag{4.56}$$

If we allow ℓ_{max} iterations, the type distribution at the output of the decoder for variable node positions $j \in \mathcal{N}_v$ is

$$\underline{P}^{(\ell_{max})}_{out,j} = \left(\prod_{[i,j]\in\mathcal{E}_{v,j}} \mathbf{P}^{(\ell_{max})}_{cv,i\to j} \right) \underline{P}_{ch}. \tag{4.57}$$

The probability of successful computation for a variable node at position $j \in \mathcal{N}_v$ is then

$$P_{dec,j} = P^{4,(\ell_{max})}_{out,j} + P^{5,(\ell_{max})}_{out,j}. \tag{4.58}$$

Again, define ϵ_{thresh} as the largest ϵ such that reliable decoding is possible for this

protograph ensemble (i.e. $\min_{j\in\mathcal{N}_v} P_{dec,j} \xrightarrow{\ell_{max}\to\infty} 1$). Again, recall that we are already considering the performance for the case that $M \to \infty$.

Through some experimentation, we have found a family of spatially coupled protograph ensembles which perform efficiently. Our goal is to construct an asymptotically regular (d_v, d_c) protograph ensemble by proper choice of the sub-matrices $\mathbf{B}_1, ..., \mathbf{B}_w$. This means that each row of the concatenated matrix $[\mathbf{B}_1 ... \mathbf{B}_w]$ must have d_c 1's and each column of

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_1 \\ \vdots \\ \mathbf{B}_w \end{bmatrix}_{d_v w \times d_c} \tag{4.59}$$

must have d_v 1's. Our approach is to select $(b_c, b_v) = (d_v, d_c)$ and $w = \frac{d_c}{2}$. Then, we consecutively assign two 1's to each row of $\mathbf{B}$ which satisfies the row constraints. The 1's in each row are ordered in a way which satisfies the column constraints. Our First-Min spatially coupled protograph construction uses the following procedure.

In the following algorithm, let $F(\mathbf{B})$ denote the first column of $\mathbf{B}$ which has a column weight smaller than d_v. Then let $M(\mathbf{B})$ denote the first column of $\mathbf{B}$ which has a minimal column weight. Then the $d_v w$ rows of $\mathbf{B}$ from (4.59) are consecutively

designed according to

$$\mathbf{B} = \mathbf{0}_{d_v w \times d_c}$$

FOR $i = 1, ..., d_v w$

$j = F(\mathbf{B})$

$\mathbf{B}_{i,j} = 1$

$j = M(\mathbf{B})$

$\mathbf{B}_{i,j} = 1$

END FOR

For an asymptotically $(5, 6)$ code, this results in sub-matrices

$$\mathbf{B}_1 = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{B}_2 = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 \end{bmatrix}$$

$$\mathbf{B}_3 = \begin{bmatrix} 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}. \tag{4.60}$$

Notice that all possible rows with two 1's are represented without repetition.

For an asymptotically $(3, 6)$ code, this results in sub-matrices

$$\mathbf{B}_1 = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$
$$\mathbf{B}_2 = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 \end{bmatrix}$$
$$\mathbf{B}_3 = \begin{bmatrix} 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}. \tag{4.61}$$

An important feature of the First-Min construction is that for each repetition of $\mathbf{B}$ in $\mathbf{B}_{[1,...,L]}$, the variable nodes should be recovered in order from left to right. The first matrix $\mathbf{B}_1$ strongly protects the variable nodes in the first column with d_v consecutive 1's which are connected via the check constrains to d_v different variable positions. Thus the messages to and from the nodes in the first column of $\mathbf{B}$ should converge quickly. When these bits are recovered, the check nodes with connections to these bits behave as if their degree has been reduced by one. This pattern is continued for the remaining rows of $\mathbf{B}_2, ..., \mathbf{B}_w$. This successive recovery of variable nodes in each position is the key principle of spatially coupled ensembles. We have applied this principle in a nested way to design $\mathbf{B}_1, ..., \mathbf{B}_w$. A more rigorous analysis of this construction is in order, but that would diverge significantly from our analysis of the JCF decoder.

IV.5. Numerical Results

Here, we present ϵ_{thresh} for the three studied classes of LDPC ensembles. For each test, we plot the theoretically achievable rates for DF with and without the penalty constraint associated with identical linear codebooks from (3.2) and (3.4) respectively. We also plot the theoretically achievable rate for CF from (3.6). Recall that the JCF typicality decoder is able to achieve all rates subject to

$$\mathcal{R}_{JCF} < \max\{\mathcal{R}_{CF}, \mathcal{R}_{DF}\} \tag{4.62}$$

which is rate optimal for the ensemble of identical linear codes.

For the regular $(d_v, d_c) = (3, [5, ..., 11])$ LDPC ensembles, the values for ϵ_{thresh} from a type distribution analysis for DF, CF, and JCF decoding are plotted in Fig. 23. The channel parametrization for Fig. 23 is

$$\underline{P}_{ch} = \begin{bmatrix} \epsilon^2 \\ (1-\epsilon)\epsilon \\ \epsilon(1-\epsilon) \\ \frac{(1-\epsilon)^2}{2} \\ \frac{(1-\epsilon)^2}{2} \end{bmatrix}. \tag{4.63}$$

The threshold performance of JCF over CF is visually apparent for moderate rates with this parametrization. When DF achieves a better threshold than CF, however, we see that the JCF and DF thresholds are equal. It is interesting however, that JCF can strictly outperform DF and CF with message passing. This may have interesting implications for the design of uncoupled protograph ensembles which may not be optimized to the channel parameters.

JCF decoding is able to achieve the theoretically optimal performance for identical linear codebooks regardless of whether DF or CF are optimal for the channel

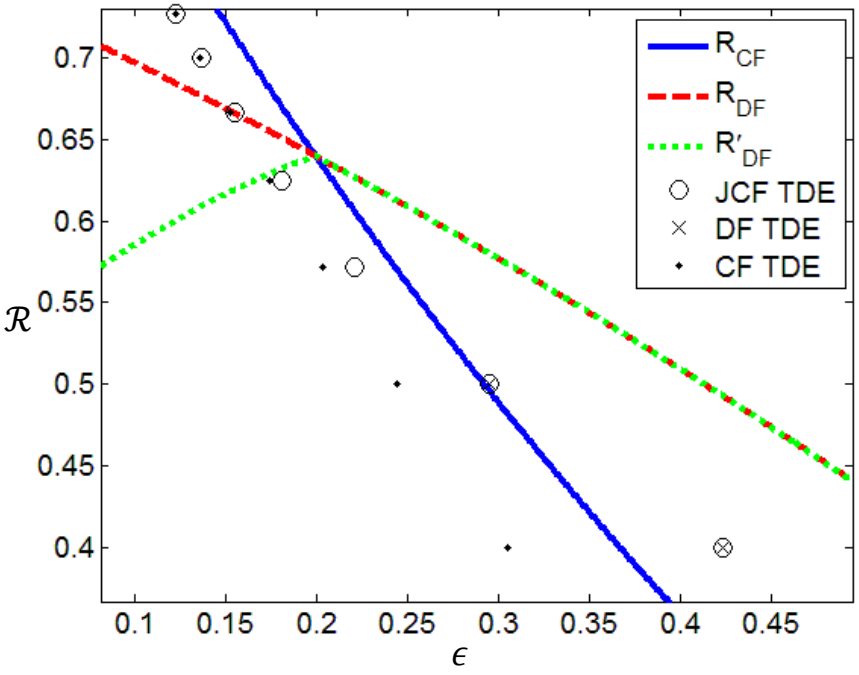


Fig. 23.: Numerically computed values for ϵ_{thresh} using CF, DF, or JCF type distribution evolution analysis for the channel parametrization in (4.63). JCF message passing strictly outperforms CF and DF message passing for some regular LDPC ensembles with this parametrization. Theoretically achievable rates $\mathcal{R}_{CF}$, $\mathcal{R}_{DF}$, and $\mathcal{R}'_{DF}$ vs ϵ are shown for comparison.

parameters. Therefore, we highlight a practical benefit associated with JCF message passing decoding. It is possible to achieve the theoretically optimal performance for a range of code rates with a single encoder and decoder with random puncturing. We apply an identical uniform puncturing sequence $\underline{\pi}$ at nodes A and B to adjust the rate of the transmitted codewords $\underline{x}_A$ and $\underline{x}_B$. Let $p_\pi = \frac{|\pi|}{N}$ be the probability that a given $x_{A,B}[n]$ is punctured. For the type distribution analysis, the punctured bits are treated as type 1 messages. Thus, the effective channel after puncturing is defined

$$\underline{P}_{ch,\pi} = \begin{bmatrix} P_{ch}^1(\epsilon)(1-p_\pi) + p_\pi \\ P_{ch}^2(\epsilon)(1-p_\pi) \\ P_{ch}^3(\epsilon)(1-p_\pi) \\ P_{ch}^4(\epsilon)(1-p_\pi) \\ P_{ch}^5(\epsilon)(1-p_\pi) \end{bmatrix}. \tag{4.64}$$

The rate of a channel code is defined as the number of message bits K divided by the codeword length N. Thus the rate of a code after puncturing is

$$\mathcal{R}_\pi = \frac{K}{N(1-p_\pi)} = \mathcal{R}\frac{1}{1-p_\pi}. \tag{4.65}$$

In Fig. 24, we plot the achievable rates for both un-punctured and punctured (d_v, d_c, L, w) ensembles as a function of ϵ_{thresh}. For the un-punctured test, we use the ensembles $(d_v, d_c, L, w) = ([3, ..., 9], 10, 4000, 100)$ with $\ell_{max} = 20,000$. We use the values of $d_v \in \{3, ..., 9\}$ to characterize the performance for a range of code rates without a puncturing sequence. For the punctured test, we use the $(d_v, d_c, L, w) = (9, 10, 10000, 100)$ ensemble with $\ell_{max} = 400,000$. We use a range of puncturing probabilities to adjust the code rate. Notice that the right most point for both the un-punctured and punctured tests corresponds to the thresholds for an asymptotically $(d_v, d_c) = (9, 10)$ code. With close inspection, we see that the

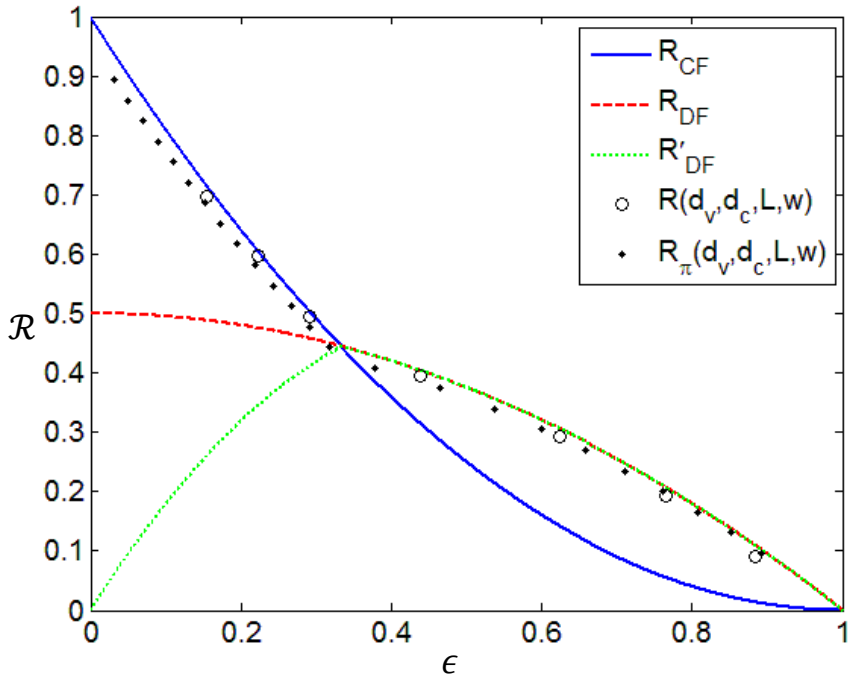


Fig. 24.: Numerically computed values for ϵ_{thresh} using type distribution evolution analysis for $(d_v, d_c, L, w) = ([3, ..., 9], 10, 4000, 100)$ and $(d_v, d_c, L, w) = (9, 10, 10000, 100)$ spatially coupled ensembles which are un-punctured and punctured respectively. The JCF message passing decoder approaches theoretical limits with spatially coupled ensembles. Theoretically achievable rates $\mathcal{R}_{CF}$, $\mathcal{R}_{DF}$, and $\mathcal{R}'_{DF}$ vs ϵ are shown for comparison.

choice of $L = 10,000$ in the punctured test results in decreased rate loss, and the $\ell_{max} = 400,000$ does improve the threshold performance over the un-punctured test with $\ell_{max} = 20,000$. Message passing decoding of these spatially coupled ensembles requires a large ℓ_{max} at thresholds near capacity because each of the variable node positions must be recovered sequentially. Puncturing of the low rate ensemble does appear to slightly decrease the threshold performance when the puncturing probability becomes large. This should be less visually apparent with a larger ℓ_{max}. As we have claimed, however, the spatially coupled ensembles achieve nearly the optimal performance for identical linear codebooks with and without puncturing.

We perform a similar puncturing analysis for the asymptotically $(7, 8)$ protograph ensemble whose sub-matrices $\mathbf{B}_1, ..., \mathbf{B}_4$ are designed according to the First-Min construction in Fig. 25. For the protograph puncturing test, however, we use a relatively small $L = 50$ and use a large $\ell_{max} = 2,000,000$ to characterize the theoretical performance of the protograph ensemble with a uniform puncturing sequence. The finite $L = 50$ results in a significant rate loss, however we plot the asymptotic rates as $L \to \infty$ against ϵ_{thresh} in Fig. 25. Notably, the threshold performance for the punctured First-Min protograph ensemble is extremely close to the theoretical limit. It is important to note, however, that this does not indicate that the protograph ensemble is capable of better threshold performance than a (d_v, d_c, L, w) ensemble in the limit as $M \to \infty$ and $L \to \infty$.

IV.6. JCF with Spatially Coupled Codes for the AWGN Channel

We present simulated results for JCF message passing decoding over the AWGN channel with BPSK signalling. The extended tanner graph is used to pass 4-ary messages of the form in (4.12). Each node maps its binary codewords to the BPSK

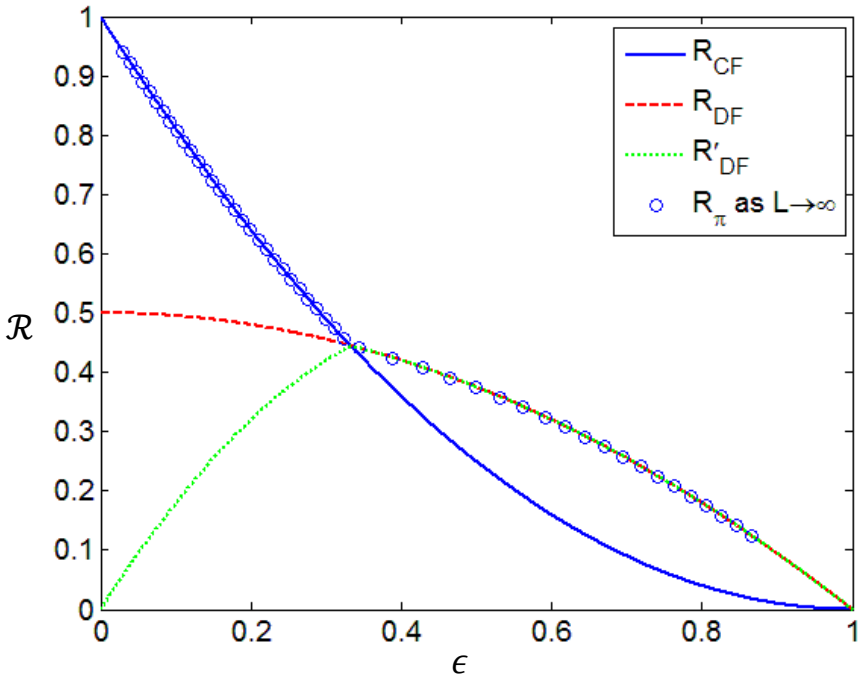


Fig. 25.: Numerically computed values for the punctured rate $\mathcal{R}_\pi$ as $L \to \infty$ vs. ϵ_{thresh} for the asymptotically $(7, 8)$ protograph ensemble designed by the First-Min construction. The JCF message passing decoder approaches the theoretical limits very tightly. Theoretically achievable rates $\mathcal{R}_{CF}$, $\mathcal{R}_{DF}$, and $\mathcal{R}'_{DF}$ vs ϵ are shown for comparison.

symbol sequences $\underline{s}_A, \underline{s}_B \in \{-1, 1\}^N$. Then the relay observes

$$\underline{y}_R = \underline{s}_A + \underline{s}_B + \underline{w}_R \tag{4.66}$$

where $\underline{w}_R$ a sequence of i.i.d. variables distributed as $N(0, \sigma^2)$. This channel is poorly matched to $\underline{x}_R$ at low SNR, but the matching is significantly improved as the SNR increases.

The variable nodes in the ETG are initialized with messages of the form

$$\mu_{ch,n} = \begin{bmatrix} P(X_{A,B}[n] = [00]^T | Y_R[n]) \\ P(X_{A,B}[n] = [01]^T | Y_R[n]) \\ P(X_{A,B}[n] = [10]^T | Y_R[n]) \\ P(X_{A,B}[n] = [11]^T | Y_R[n]) \end{bmatrix}. \tag{4.67}$$

The output messages from the variable nodes are the elementwise product of each input message. The output messages from the check nodes follow the rules in (4.21) and (4.22), however the message alphabet is the set of 4-ary pmfs rather than the restricted set $\mathcal{U}$.

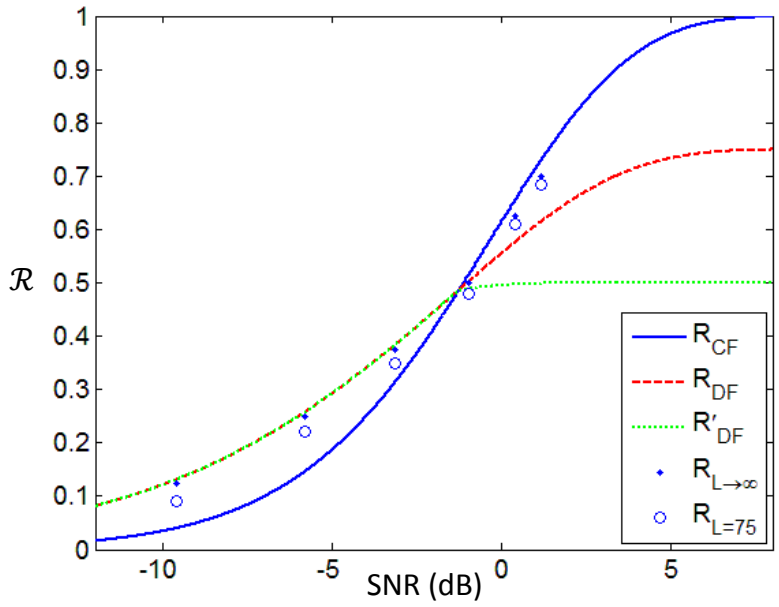
To reduce the computational complexity of the simulations in this section, we use a message passing schedule based on the windowed decoding for spatially coupled protograph ensembles developed in [52]. We leave a thorough discussion to [52] but provide the details of our simulation for completeness. Our schedule is defined by a sequence of L windows. The i_{th} window, $\mathcal{W}_i$ consists of a set of target variable positions $\{i, ..., i + b_v - 1\}$ and sets of active check and variable node positions which will process received messages. The target variable positions must be recovered before proceeding to the next window. The active protograph check positions in $\mathcal{W}_i$ are $\{i - wb_c, ..., i + 4wb_c\}$. The active protograph variable positions are those which have any edge connection to an active check position. Messages from inactive check

nodes to active variable nodes are either the last computed outgoing message (if the check node was previously active) or correspond to a uniform input message. The set of active check nodes is optionally expanded by up to $2wb_c$ on either side if the target variables are not decoded within a suitable number of iterations. Our windowed schedule reduces the computational complexity considerably for our simulations, but the effect on the decoding thresholds is minimal with these parameters.

We construct asymptotically $(d_v, d_c) = (\{3, ..., 7\}, 8), (3, 10)$ protograph codes using the First-Min construction with parameters $M = 5000$ and $L = 75$. The edge connections in each code are generated by random permutation, then all length four cycles are removed via randomized edge swapping within each permutation matrix. The terminated rate (4.49) and asymptotic rate (4.50) are plotted as a function of SNR in Fig. 26(a). The rate loss for this L is quite apparent, especially for the low rate codes. We perform the same test with parameters $M = 1250$ and $L = 200$ and give the results in Fig. 26(b). The rate loss is considerably smaller, and a slightly larger SNR is required for the communication. JCF decoding with these First-Min codes closely approaches the optimal thresholds for the AWGN MAC with identical linear codebooks and BPSK signalling.

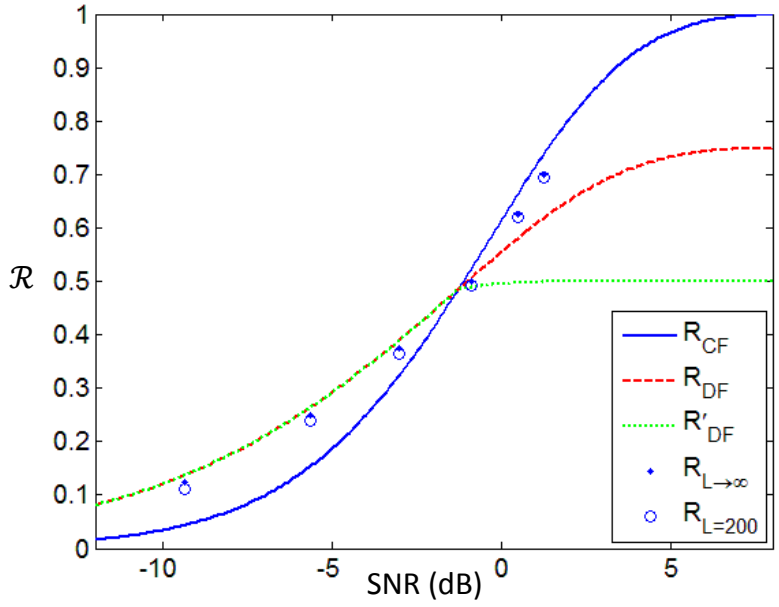
IV.7. Concluding Remarks

The proposed TWEMAC channel model is a useful tool for code design for many wireless communication problems. We derive and simplify the processing rules for JCF message passing decoding for this class of channels and derive a type distribution evolution analysis to simplify the design of LDPC ensembles for such practical decoding for computation. Our numerical results which show that spatially coupled LDPC ensembles can nearly achieve the theoretical performance limit for identical



(a) Asymptotically $(d_v, d_c) = (\{3, ..., 7\}, 8), (3, 10)$ First-Min codes with, $M = 5000$, and $L = 75$

Fig. 26.: Simulated performance of randomly generated First-Min protograph codes. The finite length and asymptotic (as $L \to \infty$) code rates are plotted as a function of the SNR (dB) for which 0 bit errors were observed with 5 simulated codeword transmissions. Theoretically achievable rates $\mathcal{R}_{CF}$, $\mathcal{R}_{DF}$, and $\mathcal{R}'_{DF}$ vs SNR (dB) are shown for comparison.



(b) Asymptotically $(d_v, d_c) = (\{3, ..., 7\}, 8), (3, 10)$ First-Min codes with, $M = 1250$, and $L = 200$

Fig. 26.: Fig. 26.: Continued.

linear codebooks. This is extended with simulation results for the AWGN multiple access channel with BPSK signalling. Generalizations of the TWEMAC channel model may help us develop practical codes suitable for computation for a wider range of practical problems. It would be of particular interest to consider computation over time varying channels with the parameters unknown to the transmitters.

Next in Chapter V, we conclude this thesis by discussing the way our results work together to shorten the gap between theory and practice for reliable PLNC. We also describe in detail the ways our results can be combined and extended with further research.

CHAPTER V

CONCLUSIONS AND FUTURE WORK

V.1. Summary of Findings

In the previous chapters, we have investigated novel coding schemes for reliable physical layer network coding over the two-way relay channel. Recent related research has shown that, with structured code ensembles, the wireless superposition may be used to compute a desired function of the transmitted messages. Our primary goal was to shorten the gap between theory and practice for reliable PLNC. We became particularly interested in the practical issue that nodes A and B may not know the structure of the wireless superposition, though it is reasonable to expect the relay to estimate the channel parameters from the observed signal [12]. Prior work has established the value adaptable network coding in such situations. Our primary contributions include a practical, reliable, and adaptable coding scheme, novel information-theoretic analysis of structured ensembles, and tools to design code ensembles which permit practical decoding for computation.

In Chapter II, we showed that significant decoding flexibility may be achieved using multilevel codes with identical linear codebooks over the binary field. We discovered that this coding structure induces penalty constraints to the achievable rates which are not present with independent codebooks. We provided a rigorous derivation of these constraints. Our numerical results indicated that decoding flexibility may be improved if the relay attempts both compute-and-forward and decode-and-forward based computation with the proposed structured encoder.

As with many recently studied problems in information and coding theory, these results emphasize the importance of reviewing the accuracy of our assumptions, par-

ticularly if they are based on established solutions to simpler problems. In classical information theory, random independent codebooks have been established as the go to solution to achieve the rate region for the multiple access channel [13]. More recently, the physical layer computation problem has revealed the value of structured code ensembles, such as lattice codes, for efficient use of linear wireless superposition. The compute-and-forward schemes based on nested lattice codes established the notion that it can be wasteful to decode all incoming codewords at a receiver [15], [20], [21]. In this thesis, we have shown that when the superposition is not known to the transmitters both methods of decoding are useful with structured ensembles.

In Chapter III, we realized that either DF or CF are independently suboptimal for reliable computation. Therefore, we studied an optimal decoding paradigm, joint-compute-and-forward, which was originally proposed in a message passing framework [17], [16], [18], [19]. We studied the simple yet general class of binary-input memoryless multiple access channels and considered the case where nodes A and B use independent cosets of identical linear codebooks generated uniformly at random. For this case, we showed that JCF essentially performs as well as the better of CF and DF depending on the channel parameters. Conversely, for the studied class of channels and code ensemble we showed that the linear combination of codewords cannot be reliably computed at rates greater than those achievable by the better of CF and DF. These results indicate that, if it is possible to improve the achievable rates for reliable computation, it will require clever design of joint encoders.

In Chapter IV, we proposed the TWEMAC channel model, which can be used to emulate various aspects of wireless superposition, and developed a framework for JCF message passing decoding on TWEMACs. This should be useful for the design of practical code ensembles which achieve our theoretical computation performance for many scenarios. We used our framework to show that spatially coupled LDPC

ensembles with JCF message passing decoding can approach the optimal computation rates achievable with identical linear codebooks. This is highlighted by our result that identical random puncturing of a spatially coupled ensemble can facilitate optimal performance for a large range of channel parameters with a single encoder and decoder. These results were supplemented with simulation results for the AWGN two-way relay channel with BPSK signaling.

Chapters III and IV reinforce our previous results and provide tools to develop practical coding implementations. Limiting our focus to the binary case and identical linear codebooks in these chapters allowed us to simplify our analysis and communicate key concepts effectively. We discovered that the additional computational complexity required for JCF decoding does not improve the computation rates over the better of DF and CF decoding; however, under certain assumptions such as finite length codes, JCF decoding may improve performance. The practical benefits and information-theoretic optimality of JCF suggest that JCF is the right way to think about decoding for computation. Throughout the thesis, we have focused on different aspects of the reliable computation problem for symmetric exchange over the two-way relay channel. Many of our results and analysis are applicable to more general networking scenarios or more general goals for the desired communication. Again, our narrow focus permitted an undistracted investigation of the ways wireless superposition may be practically used for reliable PLNC. For the remainder of this chapter, we discuss some interesting avenues for further research.

Note about notation: Throughout the thesis, we have attempted to keep notation uniform. However, we have made a few small changes to the notation, especially from Chapter II to Chapters III and IV. Important differences are mentioned in a *note about notation* near the beginning of each chapter. In the following, we combine the analysis for the different chapters and use the notation for the chapter most relevant

to the discussion. We attempt to limit the confusion by referencing equations or text in the previous chapters where relevant variables are defined. Note especially that, in sub-section V.2.V.2.1, $\underline{x}$ denotes the binary address vectors which are adaptively network coded as in Chapter II. In sub-sections V.2.V.2.3 and V.2.V.2.2, $\underline{x}$ denotes a binary codeword as in Chapters III and IV.

V.2. Future Work

During our investigation, we have developed tools and analysis which suggest further research for the reliable computation problem. In the following, we discuss specific ways in which our results may be combined or generalized and areas where further investigation should be especially informative.

V.2.1. Combining Our Results

In Chapter II, we provide numerical results for the case where the relay uses either CF or DF decoding. In Chapter III, we have studied an additional decoding paradigm called JCF for binary-input multiple access channels. JCF achieves all computation rates achievable by either CF or DF and cannot achieve better for binary memoryless MACs. However, this is not immediately clear for the multilevel case. For example, there may be certain channel parameters for which it is better to use DF decoding for one level and CF decoding for the other. For example, if nodes A and B use $\ell = 2$ levels, the relay could attempt to decode a function given by

$$f'(\underline{x}_A, \underline{x}_B) = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \underline{x}_A \\ \underline{x}_B \end{bmatrix}.$$

Then, the relay could use the values of $f'(\underline{x}_A, \underline{x}_B)$ to recover one of our originally proposed functions such as

$$f(\underline{x}_A, \underline{x}_B) = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \underline{x}_A \\ \underline{x}_B \end{bmatrix}$$

Our numerical results in Chapter II do not consider the computation of functions such as $f'(\underline{x}_A, \underline{x}_B)$. However, Theorem II.1 defines achievable rates for a MLC with identical linear codebooks for a general DMC and is therefore general enough to define achievable rates for such an expanded class of functions. Our theoretical notion of JCF is to recover as much information as possible about the observed codewords using a joint estimator, and then to combine these estimates into an efficient unambiguous network coding function for broadcast. For the MLC scheme with $\ell = 2$, the suitable functions are denoted by the set $\mathcal{F}$ defined in (2.9). We expect a JCF typicality decoder, generalized from the binary JCF typicality decoder used in Chapter III, can achieve the computation rates rigorously defined in the following conjecture

Conjecture V.1. *Let nodes A and B use the proposed MLC encoding scheme with independent cosets of an identical linear codebook generated uniformly at random. Suppose there exists a binary matrix* $\mathbf{D}_{k\times 2\ell}$, $\ell \leq k \leq 2\ell$, *such that*

$$f'(\underline{x}_A, \underline{x}_B) = \mathbf{D}_{k\times 2\ell} \begin{bmatrix} \underline{x}_A \\ \underline{x}_B \end{bmatrix} \tag{5.1}$$

is unambiguous and can be decoded according to Theorem II.1. Then there exists some $f \in \mathcal{F}$ *as defined* (2.9) *which can be reliably decoded at the relay from* $\underline{y}_R$.

This conjecture generalizes our theoretical notion of JCF to the multilevel decoder. It should be fairly straight forward to show that this result is achievable. Theorem II.1 is already applied in the conjecture to define the achievable compu-

tation rates for any unambiguous $\mathbf{D}$. Therefore, we would only need to show that decoding any unambiguous

$$f'(\underline{x}_A, \underline{x}_B) = \mathbf{D} \begin{bmatrix} \underline{x}_A \\ \underline{x}_B \end{bmatrix}$$

implies that some $f \in \mathcal{F}$ can be decoded. This is true if, for any unambiguous matrix $\mathbf{D}_{k\times 2\ell}$, there exists a full-rank matrix $\mathbf{E}_{\ell\times k}$ such that $\mathbf{ED} \in \mathcal{F}$. Then some $f \in \mathcal{F}$ is always a function of any unambiguous recoverable f'. A converse statement is more complicated to prove because we would need to extend our converse analysis to the multilevel case. However, it should be possible to prove the following conjecture with steps similar to the binary case in Chapter III.

Conjecture V.2. *Let nodes A and B use the proposed MLC encoding scheme with independent cosets of an identical linear codebook generated uniformly at random. Then define $P_{e,f}^N$ as the probability of error, averaged over the ensemble of length N codebooks, for decoding a function $f \in \mathcal{F}$. If $P_{e,f}^N \to 0$ for any $f \in \mathcal{F}$, then there exists a binary matrix $\mathbf{D}_{k\times\ell}$, $\ell \le k \le 2\ell$ such that the corresponding $f'(\underline{x}_A, \underline{x}_B)$ is unambiguous and can be decoded according to Theorem II.1.*

Conjecture V.2 is the converse to Conjecture V.1 because it states that if reliable computation is possible for any $f \in \mathcal{F}$, then the code rate $\mathcal{R}$ must satisfy the achievable rate constraints.

In Chapter IV, we developed a TWEMAC channel model for binary-input channels and showed that spatially coupled LDPC codes with JCF message passing decoding achieves the optimal computation rates for identical linear codebooks. It should be quite useful to extend the TWEMAC channel model, variable and check node processing rules, and type distribution evolution equations to the multilevel case. The analysis should not change significantly, but there should be a larger

number of message types as ℓ increases. The channel parameters for this extended erasure model could be selected to mimic the conditional entropies from some signalling constellation and channel gains $h_A, h_B \in \mathbb{C}$. Then this could be useful for designing practical codes for the multilevel coding scheme.

V.2.2. Relaxing Design Constraints

In order to communicate key concepts effectively, we have often proposed the use code ensembles with very specific structures. This is useful to simplify our analysis; however, greater design flexibility may be obtained by relaxing some of our design constraints. Here, we provide a few examples we find especially interesting.

In our proposed MLC scheme from Chapter II we use identical linear codebooks for each encoding level. Yet, a degree of decoding flexibility can be maintained if the ℓ levels are assigned to disjoint subsets, and each subset of levels uses identical linear codes which are independent from the other linear codes. For example, if $\ell = 4$, than we could use two linear codebooks $\mathcal{C}_\alpha$ and $\mathcal{C}_\beta$. Then we could choose $\mathcal{C}^1 = \mathcal{C}^2 = \mathcal{C}_\alpha$ and $\mathcal{C}^3 = \mathcal{C}^4 = \mathcal{C}_\beta$. Then any unambiguous function which linearly combines codewords from the same subset is acceptable for computation. This generalization facilitates a greater degree of design flexibility because $\mathcal{R}_\alpha$ need not be equal to $\mathcal{R}_\beta$. Also, the independence of $\mathcal{C}_\alpha$ and $\mathcal{C}_\beta$ means that fewer penalty constraints need to be satisfied at the decoder. Another useful generalization of the MLC scheme is to apply multiple labels to the each constellation point to provide a shaping gain. Finally, it should be interesting to study the decoding flexibility achieved with non-binary multilevel coding schemes with identical linear codebooks.

In Chapter III, we only consider the case where nodes A and B use identical linear codebooks generated uniformly at random. However, our analysis of the JCF typicality decoder can be applied to the case where nodes A and B use generator

matrices for which only a subset of the rows are identical and the rest are independent. This is not likely to improve the computation rate. However, if the channel is only partially matched to the linear function the relay may be able to decode some of the message bits from node A and/or node B in addition to the network coded message bits. This could be beneficial for some networks with different communication goals.

This can be better understood using a generalization of the notion of codeword triplets $(\underline{x}_A, \underline{x}_B, \underline{x}_R)$ developed in Chapter III. Particularly, consider an appended codeword $\underline{x}_{JCF} = [\underline{x}_A \underline{x}_B \underline{x}_R]$. If nodes A and B use an identical linear code $\mathcal{C}$ with generator matrix $\mathbf{G}$, then the codeword $\underline{x}_{JCF}$ is generated by an induced encoder according to

$$\underline{x}_{JCF} = [\underline{u}_A, \underline{u}_B] \begin{bmatrix} \mathbf{G} & \mathbf{0} & \mathbf{G} \\ \mathbf{0} & \mathbf{G} & \mathbf{G} \end{bmatrix} \tag{5.2}$$

or equivalently

$$\underline{x}_{JCF} = [\underline{u}_A, \underline{u}_B, \underline{u}_R] \begin{bmatrix} \mathbf{G} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{G} \end{bmatrix}. \tag{5.3}$$

More generally, suppose nodes A and B use linear codebooks $\mathcal{C}_A$ and $\mathcal{C}_B$ each of rate $\mathcal{R} = \frac{K}{N}$. Then let the first K' rows of $\mathbf{G}_A$ and $\mathbf{G}_B$ be identical and the remaining $K'' = K - K'$ rows be independently distributed. Define $\mathcal{R}' = \frac{K'}{N}$ and $\mathcal{R}'' = \frac{K''}{N}$, so that $\mathcal{R} = \mathcal{R}' + \mathcal{R}''$. The generator matrices are given by

$$\mathbf{G}_A = \begin{bmatrix} \mathbf{G}'_{K' \times N} \\ \mathbf{G}''_{K'' \times N} \end{bmatrix}$$

$$\mathbf{G}_B = \begin{bmatrix} \mathbf{G}'_{K' \times N} \\ \mathbf{G}'''_{K'' \times N} \end{bmatrix}$$

where the elements of $\mathbf{G}'$, $\mathbf{G}''$ and $\mathbf{G}'''$ are independently and uniformly distributed.

Then the induced encoder for $\underline{x}_{JCF}$ can be expressed

$$\underline{x}_{JCF} = [\underline{u}_A, \underline{u}_B] \begin{bmatrix} \mathbf{G}_A & \mathbf{0} & \mathbf{G}_A \\ \mathbf{0} & \mathbf{G}_B & \mathbf{G}_B \end{bmatrix} \tag{5.4}$$

or equivalently

$$\underline{x}_{JCF} = [\underline{u}_A, \underline{u}_B, \underline{u}_R] \left[\begin{array}{ccc} \mathbf{G}' & \mathbf{0} & \mathbf{0} \\ \mathbf{G}'' & \mathbf{0} & \mathbf{G}'' \\ \hline \mathbf{0} & \mathbf{G}' & \mathbf{0} \\ \mathbf{0} & \mathbf{G}''' & \mathbf{G}''' \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{G}' \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{array}\right]. \tag{5.5}$$

Notice that the last K'' rows of the induced generator matrix are all zeros. Therefore, the last K'' values of $\underline{u}_R$ must be recovered as in the DF paradigm. We believe our arguments in Chapter III can be extended for this generalized class of codebooks. A thorough investigation should reveal a non-trivial rate region for pairs $(\mathcal{R}', \mathcal{R}'')$. It should be especially interesting to see how the penalty constraints which we have discovered in this thesis should affect the analysis.

Finally, this generalized code ensemble may be applied to the multilevel coding scheme with interesting results. Even if $\ell = 2$ considerable design flexibility may be obtained by clever selection of the codebooks $\mathcal{C}_A^1, \mathcal{C}_A^2$, $\mathcal{C}_B^1$, and $\mathcal{C}_B^2$. An appended $\mathbf{X}_{JCF}$ codeword matrix may be formed according to

$$\mathbf{X}_{JCF} = \left[\underline{u}_A^1, \underline{u}_A^2, \underline{u}_B^1, \underline{u}_B^2\right] \begin{bmatrix} \mathbf{G}_A^1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{G}_A^1 & \mathbf{G}_A^1 & \mathbf{G}_A^1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{G}_A^1 & \mathbf{G}_A^1 & \mathbf{G}_A^1 & \mathbf{0} & \mathbf{G}_A^1 \\ \mathbf{0} & \mathbf{G}_A^2 & \mathbf{0} & \mathbf{0} & \mathbf{G}_A^2 & \mathbf{0} & \mathbf{0} & \mathbf{G}_A^2 & \mathbf{G}_A^2 & \mathbf{0} & \mathbf{G}_A^2 & \mathbf{G}_A^2 & \mathbf{0} & \mathbf{G}_A^2 & \mathbf{G}_A^2 \\ \mathbf{0} & \mathbf{0} & \mathbf{G}_B^1 & \mathbf{0} & \mathbf{0} & \mathbf{G}_B^1 & \mathbf{0} & \mathbf{G}_B^1 & \mathbf{0} & \mathbf{G}_B^1 & \mathbf{G}_B^1 & \mathbf{0} & \mathbf{G}_B^1 & \mathbf{G}_B^1 & \mathbf{G}_B^1 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{G}_B^2 & \mathbf{0} & \mathbf{0} & \mathbf{G}_B^2 & \mathbf{0} & \mathbf{G}_B^2 & \mathbf{G}_B^2 & \mathbf{0} & \mathbf{G}_B^2 & \mathbf{G}_B^2 & \mathbf{G}_B^2 & \mathbf{G}_B^2 \end{bmatrix}.$$

Then any combination of subsets of the rows from $\mathbf{G}_A^1$, $\mathbf{G}_A^2$, $\mathbf{G}_B^1$, and $\mathbf{G}_B^2$ may be identical or independent. Our analysis in Chapter III combined with the more general results in [46] should be helpful to such an investigation.

V.2.3. Extending Our Results

Code Ensembles for Computation: It would be most interesting to find a coding scheme which facilitates reliable computation at rates strictly better than CF or DF for the general binary-input memoryless multiple access channel. In Chapter III, we showed that with identical linear codebooks generated uniformly at random, it is not possible to reliably decode $\underline{x}_R = \underline{x}_A \oplus \underline{x}_B$ at rates higher than those achievable with CF or DF. We hope that a jointly designed code ensemble may facilitate reliable computation at higher rates.

Perhaps, polar codes as introduced in [53] offer such a promising method of joint design. For the binary-input two-user MAC channel, the authors in [54] show that a binary multiple access channel should polarize into one of five possible extremals. The channel is either useless, provides perfect information from one user and nothing about the other (one extremum for each user), or provides perfect information from both users. These first four channel extremum are analogous to the messages of types $1, 2, 3$, and 5 for our TWEMAC channel model. The last channel extremum found by [54] is a channel with a sum constraint of one bit per channel use which the authors claim can be used to transmit a bit from either user but not both. Perhaps this last extremum could emulate a message of type 4 in which both transmitters send a bit, but the receiver decodes the finite field sum. To the best of our knowledge, polar codes have not been studied for the computation problem. The potential benefit of a polar coding scheme for computation is that the transmitters know the positions of the bits which correspond to each channel extremum.

Computation for Time Dependent Channels: In Chapter IV, we use spatially coupled LDPC ensembles to achieve theoretical computation rates. Such codes may be too long to be practical for some wireless scenarios. In a time-varying channel

model (e.g. fast fading or block fading), it may be preferable to compute different functions of the received codewords at different time intervals. Our TWEMAC channel model with time-varying input type distributions should be useful for such code design.

Constellation Design for Computation: In Chapter II, we show numerical results for a few signalling constellations with our proposed MLC scheme. We stress the importance of considering the structure of the network/channel code and assumptions about the channel model for such constellation design. The denoise-and-forward adaptive network coding scheme for un-coded two-way relaying [28] [29] emphasizes the importance minimum distance between points in the induced constellation with un-equal function labels. Special attention is placed on avoiding singular fade states which are channel gains for which the minimum distance is zero for all available decoding functions. This design helps decrease the symbol error rate for un-coded PLNC. Very recently, some work for this un-coded PLNC problem on the two-way relay channel has shown that adaptive PLNC can benefit from the design of unconventional signalling sets [55].

For reliable PLNC schemes, such as our MLC scheme, the achievable computation rate is a more important performance metric. One reason we did not systematically approach the constellation design problem for adaptive PLNC is that we did not understand the fundamental limits for reliable decoding with structured ensembles. Even in this chapter, we have discussed some ways that greater design flexibility for the encoders may be obtained. Constellation design is a complicated optimization problem, and different design constraints and objective functions will lead to very different solutions or approaches. For example, some of our numerical results from Chapter II show that singular fade states and their respective distance shortening events may not be the dominant factor in system performance. See Fig.

10 and the relevant discussion for details. In order to design good signalling constellations for reliable PLNC, it will be important to carefully consider the goal of the desired communication, the adaptability of the network coding, and the rates which may be achieved with the channel coding structure. This thesis primarily addresses the network and channel coding part of this problem.

V.3. Conclusion

In this thesis, we have provided tools and analysis which should help close the gap between theory and practice for wireless physical layer network coding. Recently, we have realized that structured code ensembles have the potential to improve the achievable information rates for such communication. Our work highlights the substantial benefits and reveals certain limitations of structured codes for physical layer computation.

REFERENCES

[1] C. E. Shannon and W. Weaver, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 and 623–656, 1948.

[2] G. D. Forney and D. J. Costello, "Channel coding: The road to channel capacity," *Proceedings of the IEEE*, vol. 95, no. 6, pp. 1150–1177, 2007.

[3] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[4] S. L. Miller, "An adaptive direct-sequence code-division multiple-access receiver for multiuser interference rejection," *IEEE Trans. Commun.*, vol. 43, no. 234, pp. 1746–1755, 1995.

[5] P. Popovski and T. Koike-Akino, "Coded bidirectional relaying in wireless networks," *New Directions in Wireless Communications Research*, pp. 291–316, 2009.

[6] B. Nazer and M. Gastpar, "Reliable physical layer network coding," *Proceedings of the IEEE*, no. 99, pp. 1–23, 2011.

[7] R. Ahlswede, N. Cai, S.-Y. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1204–1216, 2000.

[8] S.-Y. Li, R. W. Yeung, and N. Cai, "Linear network coding," *IEEE Trans. Inf. Theory*, vol. 49, no. 2, pp. 371–381, 2003.

[9] R. Koetter and M. Médard, "An algebraic approach to network coding," *IEEE/ACM Trans. Netw.*, vol. 11, no. 5, pp. 782–795, 2003.

[10] S. Zhang, S. C. Liew, and P. P. Lam, "Hot topic: physical-layer network coding," in *Proc. 12th Annu. Int. Conf. on Mobile Computing and Networking*, (Los Angeles, CA), pp. 358–365, 2006.

[11] S. C. Liew, S. Zhang, and L. Lu, "Physical-layer network coding: Tutorial, survey, and beyond," *Physical Communication*, vol. 6, no. 0, pp. 4 – 42, 2013.

[12] M. Jain, S. L. Miller, and A. Sprintson, "Parameter estimation and tracking in physical layer network coding," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, (Houston, TX), pp. 1–6, 2011.

[13] T. Cover and J. Thomas, *Elements of Information Theory*. John Wiley and Sons, 2006.

[14] B. Nazer and M. Gastpar, "Computation over multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3498–3516, 2007.

[15] M. P. Wilson, K. Narayanan, H. D. Pfister, and A. Sprintson, "Joint physical layer coding and network coding for bidirectional relaying," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5641–5654, 2010.

[16] D. Wubben and Y. Lang, "Generalized sum-product algorithm for joint channel decoding and physical-layer network coding in two-way relay systems," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, (Miami, FL), pp. 1–5, 2010.

[17] S. Zhang and S. Liew, "Channel coding and decoding in a relay system operated with physical-layer network coding," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 5, pp. 788–796, 2009.

[18] L. Lu and S. Liew, "Asynchronous physical-layer network coding," *IEEE Trans. Wireless Commun.*, vol. 11, no. 2, pp. 819–831, 2012.

[19] L. Lu, S. Liew, and S. Zhang, "Optimal decoding algorithm for asynchronous physical-layer network coding," in *Proc. IEEE Int. Conf. Communications (ICC)*, (Kyoto), pp. 1–6, 2011.

[20] B. Nazer and M. Gastpar, "Lattice coding increases multicast rates for Gaussian multiple-access networks," in *Proc. 45th Annu. Allerton Conf. on Communication, Control, and Computing*, (Monticello, IL), 2007.

[21] W. Nam, S. Chung, and Y. Lee, "Capacity bounds for two-way relay channels," in *Proc. IEEE Int. Zurich Seminar on Communications*, pp. 144–147, 2008.

[22] U. Erez and R. Zamir, "Achieving 1/2 log (1+ snr) on the awgn channel with lattice encoding and decoding," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2293–2314, 2004.

[23] M. Wilson and K. Narayanan, "Power allocation strategies and lattice based coding schemes for bi-directional relaying," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, (Seoul), pp. 344–348, 2009.

[24] Y. Song and N. Devroye, "List decoding for nested lattices and applications to relay channels," in *Proc. 48th Annu. Allerton Conf. on Communication, Control, and Computing*, (Monticello, IL), pp. 1038–1045, 2010.

[25] J. Zhan, U. Erez, M. Gastpar, and B. Nazer, "Mimo compute-and-forward," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, (Seoul), pp. 2848–2852, 2009.

[26] B. Nazer and M. Gastpar, "Compute-and-forward: Harnessing interference through structured codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6463–6486, 2011.

[27] A. Avestimehr, S. Diggavi, and D. Tse, "Wireless network information flow: A deterministic approach," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 1872–1905, 2011.

[28] T. Koike-Akino, P. Popovski, and V. Tarokh, "Optimized constellations for two-way wireless relaying with physical network coding," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 5, pp. 773–787, 2009.

[29] T. Koike-Akino, P. Popovski, and V. Tarokh, "Denoising strategy for convolutionally-coded bidirectional relaying," in *Proc. IEEE Int. Conf. on Communications (ICC)*, (Germany), pp. 1–5, 2009.

[30] C. Feng, D. Silva, and F. R. Kschischang, "An algebraic approach to physical-layer network coding," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, (Austin, TX), pp. 1017–1021, 2010.

[31] G. Forney Jr and G. Ungerboeck, "Modulation and coding for linear Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2384–2415, 1998.

[32] A. Ozgur and S. Diggavi, "Approximately achieving Gaussian relay network capacity with lattice codes," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, (Austin, TX), pp. 669–673, 2010.

[33] S. Lim, Y. Kim, A. El Gamal, and S. Chung, "Noisy network coding," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 3132–3152, 2011.

[34] A. Sengupta, S. Brahma, A. Ozgur, C. Fragouli, and S. Diggavi, "Graph-based codes for quantize-map-and-forward relaying," in *Proc. Information Theory Workshop (ITW)*, (Campinas), pp. 140–144, 2011.

[35] A. Avestimehr, A. Sezgin, and D. Tse, "Capacity of the two-way relay channel within a constant gap," *European Transactions on Telecommunications*, vol. 21, no. 4, pp. 363–374, 2010.

[36] A. Sezgin, A. S. Avestimehr, M. A. Khajehnejad, and B. Hassibi, "Divide-and-conquer: Approaching the capacity of the two-pair bidirectional Gaussian relay network," *IEEE Trans. Inf. Theory*, vol. 58, no. 4, pp. 2434–2454, 2012.

[37] U. Wachsmann, R. Fischer, and J. Huber, "Multilevel codes: Theoretical concepts and practical design rules," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1361–1391, 1999.

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

[39] S. Kudekar, T. Richardson, and R. Urbanke, "Threshold saturation via spatial soupling: Why convolutional ldpc ensembles perform so well over the bec," *IEEE. Trans. Info. Theory*, vol. 57, no. 2, pp. 803–834, 2011.

[40] S. Kudekar and K. Kasai, "Spatially coupled codes over the multiple access channel," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, (St. Petersburg), pp. 2816–2820, 2011.

[41] S. Kudekar, T. Richardson, and R. Urbanke, "Spatially coupled ensembles universally achieve capacity under belief propagation," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, (Cambridge, MA), pp. 453–457, 2012.

[42] A. Yedla, P. S. Nguyen, H. D. Pfister, and K. R. Narayanan, "Universal codes for the Gaussian MAC via spatial coupling," in *Proc. 49th Annu. Allerton Conf. on Communication, Control, and Computing*, (Monticello, IL), pp. 1801–1808, 2011.

[43] A. Yedla, H. D. Pfister, and K. R. Narayanan, "Universality for the noisy slepian-wolf problem via spatial coupling," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, (St. Petersburg), pp. 2567–2571, 2011.

[44] R. Gallager, *Information Theory and Reliable Communication*. John Wiley and Sons, 1968.

[45] R. Gallager, "A perspective on multiaccess channels," *IEEE Trans. Inf. Theory*, vol. 31, no. 2, pp. 124–142, 1985.

[46] B. Bandemer, A. E. Gamal, and Y.-H. Kim, "Optimal achievable rates for interference networks with random codes," *submitted for publication in IEEE Trans. Inf. Theory*, vol. abs/1210.4596, 2012. available online at http://arxiv.org/abs/1210.4596.

[47] A. Khisti, B. Hern, and K. Narayanan, "On modulo-sum computation over an erasure multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 59, no. 7, pp. 4129–4138, 2013.

[48] M. Noori, H. Bagheri, and M. Ardakani, "Low-latency data sharing in erasure multi-way relay channels," *arXiv preprint arXiv:1211.1044*, 2012.

[49] B. Smith and S. Vishwanath, "Unicast transmission over multiple access erasure networks: Capacity and duality," in *Proc. Information Theory Workshop (ITW)*, (Lake Tahoe, CA), pp. 331–336, 2007.

[50] M. Lentmaier, G. Fettweis, K. Zigangirov, and D. Costello, "Approaching capacity with asymptotically regular ldpc codes," in *Proc. Information Theory and Applications Workshop (ITA)*, (La Jolla, CA), pp. 173–177, 2009.

[51] M. Lentmaier, D. Mitchell, G. Fettweis, and D. Costello, "Asymptotically regular ldpc codes with linear distance growth and thresholds close to capacity," in *Proc. Information Theory and Applications Workshop (ITA)*, (San Diego, CA), pp. 1–8, 2010.

[52] A. Iyengar, M. Papaleo, P. Siegel, J. Wolf, A. Vanelli-Coralli, and G. Corazza, "Windowed decoding of protograph-based ldpc convolutional codes over erasure channels," *IEEE Trans. Inf. Theory*, vol. 58, no. 4, pp. 2303–2320, 2012.

[53] E. Arikan, "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, 2009.

[54] E. Sasoglu, E. Telatar, and E. Yeh, "Polar codes for the two-user binary-input multiple-access channel," in *in Proc. IEEE Information Theory Workshop (ITW)*, (Dublin), pp. 1–5, 2010.

[55] K. Venugopal, V. Namboodiri, and B. S. Rajan, "On constellations for physical layer network coded two-way relaying," *arXiv preprint arXiv:1303.7296*, 2013.