Capacity, Error Exponent, and Structural Results for Communication Networks

by

Mohsen Heidari Khoozani

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical Engineering and Computer Science)
in The University of Michigan
2019

Doctoral Committee:

Professor S. Sandeep Pradhan, Chair
Associate Professor Achilleas Anastasopoulos
Professor David L. Neuhoff
Professor Martin J. Strauss
Professor Wojciech Szpankowski, Purdue University

Mohsen Heidari-Khoozani

mohsenhd@umich.edu

ORCID iD: 0000-0002-0012-2900

To Maman and Baba with love.

ACKNOWLEDGEMENTS

I would like to express my high gratitude to my advisor Professor Sandeep Pradhan. I am truly thankful of him for his constant support, encouragement and expert guidance over the past five years. I have been very fortunate to work with an advisor who gave me the freedom to explore on my own, while being deeply involved in my research. Sandeep's high standards on the quality of research as well as ethical conducts are aspects I hope to emulate in my future career.

It has been a pleasure to have Professor Achilleas Anastopolos, Professor David Neuhoff, Professor Martin Strauss, and Professor Wojciech Szpankowski in my dissertation committee. I am especially indebted to Professor Anastopolos for being my collaborator and for many extensive and illuminating discussions we had over the years. I am grateful to Professor Szpankowski and Professor Neuhoff for taking the time to provide me with their invaluable advice. I wish to thank Professors Neuhoff and Professor Anastasopoulous for excellent courses in Source Coding theory and Channel Coding theory. I am thankful to Professor Strauss for providing a complementary perspective on my research.

I would like to thank the entire faculty of the Department of Electrical Engineering for creating a great learning atmosphere. I would like to express my special gratitude to Professor Demos Teneketzis for teaching me probability and random processes and for many interesting discussions over the past five years. It was a great pleasure and a learning experience to have been a GSI for my advisor Professor Pradhan and Professor Teneketzis. I am also grateful to the faculty of the Department of

Mathematics for inspiration and intellectual incentives.

Ann Arbor has been a second home for me with wonderful memories, thanks to amazing friends I have had here over the years. My special thanks go to Hamidreza Aghasi and Farhad Shirani for being incredible friends and colleagues. I extend my special thanks to my uncle Hossein and his wife Fatima for their selfless support since day one of my PhD life. I also cannot emphasize enough the importance of all the fantastic friends that I was lucky to meet during these years: Parisa Ghaderi, Mehrzad Samadi, Armin Jam, Avish Kosari, Mina Jafari, Mehrdad Moharami, Salimeh Yasayee Sekeh, Ali Mostajeran, Mahmoud Barangi, Hamidreza Tavafoghi, Parinaz Naghizadeh, Payam Mirshams, Nina Zabihi, Azadeh Ansari, Morteza Noushad, Nyousha Navidi, Armin Sarabi and Mohammad Masoud among them. I am grateful to my colleagues Aria Sahebi, Arun Padakandla, Touheed Atif and Deepanshu Vasal.

Lastly, it is my greatest wish to thank my mother, my father and my little brothers. I deeply indebted to them for their support, encouragements and especially many sacrifices they made for my sake throughout of my life. No amount of thanks could repay them for their kindness and love.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

LIST OF ABBREVIATIONS

PtP point-to-point

VLC variable-length code

IID independent identically distributed

MAC multiple-access channel

BC broadcast channel

IC Interference channel

DMC discrete memoryless channel

QSC quasi-structured code

QGC Quasi Group Codes

MAC-FB MAC with feedback

AWGN additive white Gaussian noise

MD multiple descriptions

CES Cover-El Gamal-Salehi

CL Cover-Leung

ABSTRACT

In various multi-terminal communication scenarios, contrary to point-to-point communication, characterization of fundamental limits such as capacity and error exponent is still an open problem. We study such fundamental limits and the structure of optimality achieving codes. This thesis consists of two parts: in the first part, we investigate the role of algebraic structures in multi-terminal communications. We show the necessity of various types of algebraic structure in capacity achieving codes and argue that the lack of such structures in the conventional random codes leads to their sub-optimality. We develop a new class of partially structured codes called quasi-structured code (QSC). Such codes span the spectrum from completely structured to completely unstructured codes. It is shown that the application of QSCs leads to improvements over the current coding strategies for many problems including distributed source coding and multiple-access channel (MAC) with feedback.

In the second part of the thesis, we study the optimal error exponent in various multi-terminal communication scenarios. We derive a lower and upper bound on the error exponent of discrete memoryless MAC with noiseless feedback and variable-length codes (VLCs). The bounds increase linearly with respect to a specific Euclidean distance measure defined between the transmission rate pair and the capacity boundary. The bounds are shown to be tight for specific classes of MACs.

CHAPTER I

Introduction

1.1 Point-to-point Communications

Information theory is a "mathematical theory of communications" [1], providing an abstract model to analyze communication systems. Based on this model, a communication system consists of the following essential features:

- An information source, producing an *a priori* unknown message or a sequence of messages, modeled as a random variable (or a random sequence) taking values from a set $\mathcal{M}$.
- A channel, representing the medium over which the communication takes place. The channel's input and output alphabets are denoted by $\mathcal{X}$ and $\mathcal{Y}$, respectively. The effect of the channel on the input is modeled by a random mapping from $\mathcal{X}$ to $\mathcal{Y}$[1].
- A (block) encoder that maps the observed message (sequence) to a channel input sequence of length n. Such sequence is called a *codeword* and n is called the *blocklength*.
- A decoder that observes the channel output sequences and outputs an estimate of the original message.

[1] In this work, we restrict ourselves to stationary and memoryless channels.

Typically, the message is selected with uniform distribution from $\mathcal{M}$. Given a fidelity measure, such as the probability of decoding the correct message, one can determine whether a target threshold is met. Often, probability of decoding a wrong message, known as *error probability*, is considered as a measure of the *reliability* of the communication system. The ratio

$$R \triangleq \frac{\log_2 |\mathcal{M}|}{n}, \tag{1.1}$$

known as the *transmission rate*, gives the amount of transmitted information and is measured in bits per channel use. The design objective for a communication system is to find a pair of encoder - decoder satisfying an error probability ϵ with the highest possible transmission rate. A pair encoder - decoder is often referred to as a *coding strategy* or *coding scheme*.

1.1.1 The Capacity

The *capacity* is defined as the maximum transmission rate for which a communication with vanishing probability of error, $\epsilon \to 0$, is possible. More precisely, the capacity is expressed as

$$C \triangleq \lim_{\epsilon \to 0} \lim_{n \to \infty} \frac{\log_2 |M(n, \epsilon)|}{n}, \tag{1.2}$$

where $M(n, \epsilon)$ is the maximum possible message size M for any code with blocklength n and error probability lower than ϵ. Two imperative results in information theory are 1) recognizing that the capacity of a communication system is fundamental to its performance limits, and 2) characterizing the capacity of a channel in terms of commutable quantities called *mutual information*.

1.1.2 Random Codebooks

In search of codes for reliable communications in information theory, one considers a method involving so-called independent identically distributed (IID) random codebooks[2] [2] that is proved to be capacity achieving [1]. In this method, for each possible realization of the message an IID random sequence X^n is generated according to an appropriately predefined probability distribution P_X. Such a code possesses only single-letter empirical properties. This enables one to derive performance limits, in terms of achievable rates, as a functional of the underling probability distribution P_X. In this context, the capacity of any stationary and memoryless channel is achievable using unstructured random codes and is expressed as

$$C = \max_{P_X} I(X;Y),$$

where $I(X;Y)$ is the *mutual information* [1–3].

1.1.3 Error Exponent

The work on characterizing the channel capacity indicates that communications with arbitrary small probability of error is possible if and only if $R < C$ [1, 2]. This result, as in equation (1.2), is a characterization for asymptotically large blocklength n. However, due limitations on the delay of the communication in practical applications, the blocklength is finite. Moreover, the communication often takes place with non-zero error probability. For these applications, it is required to specify the rate of the decay of the error probability as a function of rate and blocklength.

The idea is to fix a rate R and study the function $P_e(n, R)$ which is the smallest error probability among codes with length n and rate R. The asymptotic behavior of this function for fixed rate is determined by the *reliability function* (also known as

[2]Such codes are sometimes referred to as unstructured random codes.

the *error exponent*) [4] which is defined as

$$E(R) \triangleq \limsup_{n\to\infty} -\frac{1}{n} \log P_e(n, R). \tag{1.3}$$

An implication of (1.3) is that if $R < C$, then the smallest possible error probability decreases exponentially as n increases. The exponent is determined by $E(R)$.

1.2 Multi-Terminal Communication

In the context of developing a mathematical formulation for communication among multiple transmitters and receivers, similar approaches, as in point-to-point (PtP) setting, are taken to model a communication network [5]. For that, many fundamental problems are identified such as multiple-access channel (MAC), broadcast channel (BC) and Interference channel (IC) [2, 5, 6]. These problems are viewed as building blocks; studying them gives insight into understanding larger networks. In this context, multiple transmission rates, one for each transmitter-receiver pair, are defined. The capacity region is, then, defined as the set of all rate-tuples for which communications with vanishing error probability is possible. Following the insights from PtP setting, the following fundamental problems need to be addressed:

(1) Computable characterization of the capacity region,

(2) The design of capacity achieving codes,

(3) Closed-form expression for the error exponent.

As for the first problem, followed by the work in [1], the capacity region is characterized in terms of "single-letter" information quantities for a few problems including MAC and *degraded* BC [2, 5] . However, characterizing the capacity region of several multi-terminal systems, such as IC, and BC, remains an open problem.

As for the second problem, based on the initial successes in PtP setting, it was widely believed that one can achieve the capacity of any network communication problem using IID codebooks. However, departing from traditional approaches, Körner and Marton [7] suggested a technique based on statistically correlated codebooks (identical random linear codes), referred to as (random) structured codes, that outperformed all techniques using random unstructured codes. This technique was proposed for compression of two correlated binary sources when the objective is to reconstruct the modulo-two sum of the sources. Also, recent results for the problem of IC [8] showed that the well-known Han-Kobayashi rate region [9] is strictly sub-optimal. These investigations, together with similar observations ([10–35]) indicate that coding strategies solely based on random unstructured codebooks may not be capacity achieving. This points out to the need for more investigations into the structure of the capacity achieving codes for multi-terminal communications.

1.2.1 On the Structure of Capacity Achieving Codes

In the context of PtP communications, if one constructs a random codebook simply by choosing the codewords using IID random variables, then, with high probability, the codebook is capacity achieving. However, this is not the case in multi-terminal communication systems. It appears that there is a trade-off between *cooperation* and *communication/compression* in networks. To see this, consider the following observations.

As depicted in Figure 1.1, suppose there are two correlated binary sources of information; each observed by one encoder. The objective of the encoders is to compress the sources in a distributed fashion such that a central decoder would be able to reconstruct the sources losslessly (Slepian-Wolf setting [36]). For this setup,

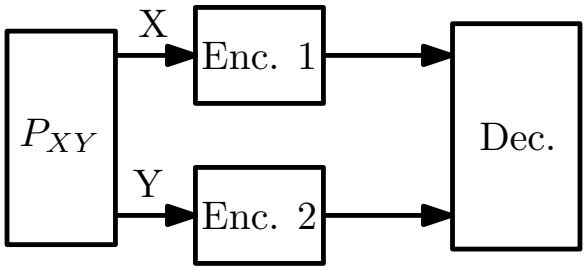

Figure 1.1: Distributed compression of two correlated binary source (X, Y). Each encoder observes one of the sources. The encoders communicate information to a central decoder. The decoder uses the received information to reconstruct (a function of) the sources. The design objective is to minimize the rate of transmissions.

the minimum required rate is

$$R_1 \geq H(X|Y), \quad R_2 \geq H(Y|X), \quad R_1 + R_2 \geq H(X, Y),$$

where $H(\cdot)$ is the entropy of the sources. In general, to achieve the Slepian-Wolf performance limit, one can use independent Shannon-style unstructured code ensembles [2]. However, if the objective is to reconstruct the modulo-two sum of the sources, then as Körner and Marton suggested, identical linear codes are needed. With this approach, when the joint distribution P_{XY} is symmetric, the minimum required rate is

$$R_1 = R_2 \geq H(X \oplus Y).$$

This implies that the sum-rate is $2H(X \oplus Y)$ which can be strictly less than $H(X, Y)$. In summary, to achieve network cooperation (decoding the sum) the users must use identical linear codes. However, if the objective is to have the full reconstruction of both the sources at the decoder, then the use of identical binning can be strictly suboptimal.

A similar observation was made recently regarding the interference channels [33]:

each cooperating transmitter using identical linear codes must pay some penalty in terms of sacrificing her/his rate for the overall good of the network. A selfish user intent on maximizing individual throughput must use essentially independent Shannon-style unstructured code ensembles. These observations indicate that the algebraic structures of coding strategies contribute to balance the trade-off between cooperation and communication/compression.

Toward addressing the role of algebraic structures in balancing the trade-off, one needs a measure for algebraic closure ("closedness") properties of codebooks. Assume the codewords of a codebook $\mathcal{C}$ are binary vectors. The size of the modulo-two sum of $\mathcal{C}$ with itself can be viewed as a measure of its algebraic closure. On one extreme, $\mathcal{C}$ is completely structured in the sense that the size of $\mathcal{C} \oplus \mathcal{C}$ equals the size of $\mathcal{C}$. This implies that $\mathcal{C}$ is closed under modulo-two addition. On the other extreme, unstructured Shannon random codes are completely unstructured in the sense that the size of $\mathcal{C} \oplus \mathcal{C}$ is close to the size of $\mathcal{C} \times \mathcal{C}$ with high probability. This gap between the completely structured codes and the completely unstructured codes leads to the following question:

Is there a spectrum of strategies involving partially structured codes or partially unstructured codes that lie between these two extremes?

In Chapter II, we investigate the existence of partially structured codes that close this gap and lie between the two extremes. To this end, we develop a new class of codes called quasi-structured code (QSC). A QSC is defined as a subset of a structured code (e.g. linear code)[3]. We show that QSCs span the spectrum from completely structured to completely unstructured. More precisely, the size of $\mathcal{C} \oplus \mathcal{C}$ is between $|\mathcal{C}|$ and $|\mathcal{C}|^2$. We provide a method for constructing specific subsets of these codes by putting single-letter distributions on the indices of the codewords. We can analyze the performance of the resulting code ensemble, and characterize the

[3]The motivation for this work comes from our earlier work on multi-level polar codes based on $\mathbb{Z}_{p^r}$ [37]. A multi-level polar code is not a group code. But it is a subset a nontrivial group code.

asymptotic performance using single-letter information quantities. By choosing the single-letter distribution on the indices one can operate anywhere in the spectrum between the two extremes: structured codes and unstructured codes. We use these class of codes to derive strictly improved achievable regions for many fundamental multi-terminal problems.

1.3 Channels with Noiseless Feedback

A challenging part of information theory is the study of channels with feedback. In this model, output symbols of a memoryless channel are available, with one unite of delay, to the transmitter. Surprisingly, the first result in this context indicates that feedback does not increase the capacity of discrete memoryless channel (DMC) [38]. Furthermore, feedback does not improve the error exponent of symmetric channels when fixed blocklength codes are used [39, 40].

For communications over channels with feedback, one can use a so-called variable-length code (VLC) whose length can depend on the channel realizations. In this context, feedback does help. Feedback reduces the complexity of the encoding and decoding required to achieve a target error probability [41]. In a remarkable work, Burnashev [42] demonstrated that the error exponent improves for DMCs with feedback and variable-length codes. The error exponent has a simple form

$$E(R) = (1 - \frac{R}{C})C_1, \tag{1.4}$$

where $0 \leq R < C$ is the (average) rate of transmission, C is the capacity of the channel, and C_1 is the maximal *relative entropy* between conditional output distributions.

In the context of communications over multi-user channels, the benefits of feedback are more prominent. Gaarder and Wolf [43] showed that feedback can expand the capacity region of discrete memoryless MAC. Similar to the study of communication

systems without feedback, three research directions are identified: 1) characterization of the capacity region, 2) structure of capacity achieving codes , and 3) closed-form expression for the error exponent. There are many partial results (namely, [44–46]) to address the first problem. The capacity region of two-user discrete memoryless MAC with feedback (MAC-FB) is characterized by Kramer in 1998 [47]. However, the characterization is in terms of multi-letter *directed mutual information* measures which is not computable in general. To the best of our knowledge, there is no closed-form expression for the error exponent of MAC with feedback. Finding a computable characterization for the capacity region and the error exponent of MAC-FB remains an open problem. In this thesis, we investigate the problem of communications over MAC with feedback to characterize 1) the error exponent, and 2) the structure of capacity achieving codes. In what follows, we explain our main contributions in this setting.

1.3.1 Coding Structures for MAC with Feedback

In MAC-FB setup, the transmitters send independent messages simultaneously to a receiver. However, conditioned on the feedback, the messages are statistically correlated. This correlation can be used to combat interference and channel noise more effectively in subsequent channel uses.

We study the problem of finding capacity achieving coding strategies for MAC-FB setup. For that, in Chapter IV, we make a connection between MAC-FB and another fundamental problem called transmission of correlated sources over MAC [48]. This problem is explained as follows:

MAC with correlated source: In this problem, there are multiple transmitters; each observing a source correlated to others. The transmitters do not communicate with each other and wish to send their observations via a MAC to a central receiver.

The receiver reconstructs the sources losslessly. This problem is studied in many works including [6, 48, 49].

In addition, we use the concept of *common information* due to Gács-Körner [50] and Witsenhausen [51] which is explained as follows:

Common information: The common information between two random variables S_1, S_2 is the maximum entropy of W which is a function of S_1 and a function of S_2, i.e., $W = f(S_1) = g(S_2)$. In other words, common information quantifies the amount of common randomness that can be extracted by knowing S_1 and S_2 separately.

We seek a more general definition of common information which incorporates the cases with more than two random variables, say S_1, S_2, and S_3. We introduce a new form of common information called *conferencing* common information. In Chapter III, we study the use of this common information to develop coding strategies for the three-user version of MAC-FB and MAC with correlated sources. It is shown, in Chapter III and IV, that exploiting conferencing common information contributes to improvements over conventional coding schemes, such as Cover - El Gamal - Salehi [48], and Cover - Leung schemes [52].

1.3.2 On the Error Exponent of MAC with Feedback

The decoding error in a MAC-FB setup consists of the union of two error events (say E_1, E_2) one for decoding each transmitter's message. Therefore, the probability of error equals to the sum of the following three terms: $P(E_1 \backslash E_2), P(E_2 \backslash E_1)$, and $P(E_1 \cap E_2)$. The main challenge in finding the error exponent in this setting is to analyze the exponential rate of decay of these probabilities and to determine which term is the dominant one.

To overcome this challenge, in Chapter VI, we make a connection between this problem and the problem of sequential hypothesis testing [53]. We use the tools from

dynamic programming and Burnashev's techniques for PtP settings [42] to derive bounds on the error exponent of MAC-FB. We derive an upper bound and a lower bound on the error exponent. In this setting, we observe that the upper bound can be expressed in the following form

$$E_u(R_1, R_2) = (1 - \frac{||\underline{R}||}{C(\theta_R)})D_u \tag{1.5}$$

where $(||\underline{R}||, \theta_R)$ denote the polar coordinate of (R_1, R_2) in $\mathbb{R}^2$. Also, $C(\theta_R)$ is the point of the capacity frontier at the angle determined by $\underline{R}$. The lower-bound is the same as E_u but with different constant D_l. The constants D_l and D_u depend only on the channel's transition probability matrix and are determined by the relative entropy between the conditional output distributions.

1.4 Communication Systems with Continuous Alphabets

Communications over channels/ networks whose input alphabets are Euclidean spaces, e.g. $\mathbb{R}$, accounts for different coding strategies than the discrete counterparts. A well-known example is the channel with additive white Gaussian noise (AWGN)

$$Y = X + N, \quad N \sim \mathcal{N}(0, 1) \tag{1.6}$$

with input power constraint $\frac{1}{n}\sum_{i=1}^{n} X_i^2 \leq P$. The works in finding low-complexity and capacity achieving codes for these channels give rise to many coding structures such as *Lattice* codes [54–56] .

Lattice codes are analogous of linear codes in Euclidean spaces. A lattice code in $\mathbb{R}^d$ is defined as the set of all linear combinations, with integer coefficients, of a given set of linearly independent vectors in $\mathbb{R}^d$. Traditionally, performance characterization of lattices is carried out using Gaussian test channels. Such techniques are known to

be suitable for Gaussian source/channel setups. The capacity of the AWGN channel in (1.6) is achievable using lattices [54]. However, for general channel/source setups, it is difficult to derive achievable rates of lattices using such techniques. Recently, a new method is introduced to overcome this challenge [57]. In the method, first the objective continuous source/ channel problem is quantized to obtain its discrete version. The performance analysis is carried out for the discrete version of the problem and inner bounds are derived in terms of discrete mutual information quantities. Then, it is shown that as the discretization process keeps refining, the mutual information terms converge to the continuous ones. Hence, inner bounds are obtained for the original continuous source/channel setup. This method is not restricted to PtP systems. Using this approach together, in Chapter V, we extend our results in discrete settings to multi-terminal systems with continuous alphabets. We introduce coding schemes based on lattices for multiple descriptions (MD) and MAC problems. We show that applications of lattices for such problems lead to performance improvements comparing to the conventional coding strategies.

CHAPTER II

Quasi-Structured Codes for Multi-Terminal Communications

Stepping beyond this conventional technique, Körner and Marton [7] proposed a technique based on statistically correlated codebooks (in particular, identical random linear codes) possessing algebraic closure properties, henceforth referred to as (random) structured codes, that outperformed all techniques based on (random) unstructured codes. This technique was proposed for the problem of distributed computation of the modulo two sum of two correlated symmetric binary sources [7]. Applications of structured codes were also studied for various multi-terminal communication systems, including, but not limited to, distributed source coding [10–13], computation over MAC [14–20], MAC with side information [11, 21–24], the joint source-channel coding over MAC [25], multiple-descriptions [26], interference channel [27–33], broadcast channel [34] and MAC with Feedback [35]. In these works, algebraic structures are exploited to design new coding schemes which outperform all coding schemes solely based on random unstructured codes. The emerging opinion in this regard is that even if computational complexity is a non-issue, algebraic structured codes may be necessary, in a deeply fundamental way, to achieve optimality in transmission and storage of information in networks.

There are several algebraic structures such as fields, ring and groups. Linear

codes are defined over finite fields. The focus of this work is on structured codes defined over the ring of modulo-m integers, that is $\mathbb{Z}_m$. Group codes are a class of structured codes constructed over $\mathbb{Z}_m$, and were first studied by Slepian [58] for the Gaussian channel. A group code over $\mathbb{Z}_m$ is defined as a set of codeswords that is closed under the element-wise modulo-m addition. Linear codes are a special case of group codes (the case when m is a prime). There are two main incentives to study group codes. First, linear codes are defined only over finite fields, and finite fields exists only when alphabet sizes equal to a prime power, i.e., $\mathbb{Z}_{p^r}$. Second, there are several communications problems in which group codes have superior performance limits compared to linear codes. As an example, group codes over $\mathbb{Z}_8$ have better error correcting properties than linear codes for communications over an additive white Gaussian noise channel with 8-PSK constellation [59]. As an another example, construction of polar codes over alphabets of size equal to a prime power p^r, is more efficient with a module structure rather than a vector space structure [37, 60–62]. Bounds on the achievable rates of group codes in PtP communications were studied in [59, 63–67]. Como [66] derived the largest achievable rate using group codes for certain PtP channels. In [63], Ahlswede showed that group codes do not achieve the capacity of a general discrete memoryless channel. In [67], Sahebi et.al., unified the previously known works, and characterized the ensemble of all group codes over finite commutative groups. In addition, the authors derived the optimum asymptotic performance limits of group codes for PtP channel/source coding problems.

Contributions

Our contributions in this Chapter are as follows. A new class of codes over groups called Quasi Group Codes (QGC) is introduced. These codes are constructed by taking subsets of group codes. This work considers QGCs over cyclic groups $\mathbb{Z}_{p^r}$. One can use the fundamental theorem of finitely generated Abelian groups to generalize

the results of this paper to QGCs over non-cyclic finite Abelian groups. Information-theoretic characterizations for the asymptotic performance limits and properties of QGCs for source coding and channel coding problems are derived in terms of single-letter information quantities. Covering and packing bounds are derived for an ensemble of QGCs. Next, a binning technique for the QGCs is developed by constructing nested QGCs. As a result of these bounds, the PtP channel capacity and optimal rate-distortion function of sources are shown to be achievable using nested QGCs. The applications of QGCs in some multi-terminal communications problems are considered. More specifically our study includes the following problems:

Distributed Source Coding A more general version of Körner-Marton problem is considered. In this problem, there are two distributed sources taking values from $\mathbb{Z}_{p^r}$. The sources are to be compressed in a distributed fashion. The decoder wishes to compute the modulo p^r-addition of the sources losslessly.

Computation over MAC In this problem, two transmitters wish to communicate independent information to a receiver over a MAC. The objective is to decode the modulo-p^r sum of the codewords sent by the transmitters at the receiver. This problem is of interest in its own right. Moreover, this problem finds applications as an intermediate step in the study of other fundamental problems such as the interference channel and broadcast channel [34, 68].

MAC with Distributed States In this problem, two transmitters wish to communicate independent information to a receiver over a MAC. The transition probability between the output and the inputs depends on states S_1, and S_2 corresponding to the two transmitters. The state sequences are generated IID according to some fixed joint probability distribution. Each encoder observes the corresponding state sequence non-causally. The objective of the receiver is to decode the messages of both

transmitters.

These problems are formally defined in the sequel. For each of these problems, a coding scheme based on (nested) QGCs is introduced. It is shown, through examples, that the coding scheme improves upon the best-known coding strategies based on unstructured codes, linear codes and group codes. In addition, for each problem a new single-letter achievable rate-region is derived. These rate-regions strictly subsume all the previously known rate-regions for each of these problems.

2.1 Preliminaries

A group is a set equipped with a binary operation denoted by "+". All groups in this paper are Abelian. Given a prime power p^r, the group of integers modulo-p^r is denoted by $\mathbb{Z}_{p^r}$, where the underlying set is $\{0, 1, \cdots, p^r - 1\}$, and the addition is modulo-p^r addition. Given a group M, a subgroup is a subset H which is closed under the group addition. For $s \in [0 : r]$, define

$$H_s = p^s \mathbb{Z}_{p^r} = \{0, p^s, 2p^s, \cdots, (p^{r-s} - 1)p^s\},$$

and $T_s = \{0, 1, \cdots, p^s - 1\}$. For example, $H_0 = \mathbb{Z}_{p^r}, T_0 = \{0\}$, whereas $H_r = \{0\}, T_r = \mathbb{Z}_{p^r}$. Note, H_s is a subgroup of $\mathbb{Z}_{p^r}$, for $s \in [0 : r]$. Given H_s and T_s, each element a of $\mathbb{Z}_{p^r}$ can be represented uniquely as a sum $a = t + h$, where $h \in H_s$ and $t \in T_s$. We denote such t by $[a]_s$. Note that $[a]_s = a \bmod p^s$, for $s \in [0, r]$. Therefore, with this notation, $[\cdot]_s$ is a function from $\mathbb{Z}_{p^r} \to T_s$. Note that this function satisfies the distributive property:

$$[a + b]_s = \Big[[a]_s + [b]_s\Big]_s$$

For any elements $a, b \in \mathbb{Z}_{p^r}$, we define the multiplication $a \cdot b$ by adding a with

itself b times. Given a positive integer n, denote $\mathbb{Z}_{p^r}^n = \bigotimes_{i=1}^n \mathbb{Z}_{p^r}$. Note that $\mathbb{Z}_{p^r}^n$ is a group, whose addition is element-wise and its underlying set is $\{0, 1, \ldots, p^r - 1\}^n$. We follow the definition of shifted group codes on $\mathbb{Z}_{p^r}$ as in [67] [10].

Definition 1 (Shifted Group Codes)**.** *An (n, k)-shifted group code over $\mathbb{Z}_{p^r}$ is defined as*

$$\mathcal{C} = \{\mathbf{uG} + \mathbf{b} : \mathbf{u} \in \mathbb{Z}_{p^r}^k\}, \tag{2.1}$$

where $\mathbf{b} \in \mathbb{Z}_{p^r}^n$ is the translation (dither) vector and $\mathbf{G}$ is a $k \times n$ generator matrix with elements in $\mathbb{Z}_{p^r}$.

We follow the definition of typicality as in [3].

Definition 2. *For any probability distribution P on $\mathcal{X}$ and $\epsilon > 0$, a sequence $\mathbf{x}^n \in \mathcal{X}^n$ is said to be ϵ-typical with respect to P if*

$$\left|\frac{1}{n}N(a|\mathbf{x}^n) - P(a)\right| \leq \frac{\epsilon}{|\mathcal{X}|}, \ \forall a \in \mathcal{X},$$

and, in addition, no $a \in \mathcal{X}$ with $P(a) = 0$ occurs in $\mathbf{x}^n$. Note that $N(a|x^n)$ is the number of the occurrences of a in the sequence $\mathbf{x}^n$. The set of all ϵ-typical sequences with respect to a probability distribution P on $\mathcal{X}$ is denoted by $A_\epsilon^{(n)}(X)$.

The above definition can be extended to define joint typicality with respect to a joint probability distribution P_{XY} on $\mathcal{X} \times \mathcal{Y}$. A pair of sequences $(\mathbf{x}^n, \mathbf{y}^n) \in \mathcal{X}^n \times \mathcal{Y}^n$ is said to be jointly ϵ-typical with respect to P_{XY} if

$$\left|\frac{1}{n}N(a, b|\mathbf{x}^n, \mathbf{y}^n) - P_{XY}(a, b)\right| \leq \frac{\epsilon}{|\mathcal{X}||\mathcal{Y}|}, \ \forall (a, b) \in \mathcal{X} \times \mathcal{Y}$$

such that none of (a, b) with $P_{XY}(a, b) = 0$ occurs in $(\mathbf{x}^n, \mathbf{y}^n)$. The set of all such pairs is denoted by $A_\epsilon^{(n)}(X, Y)$.

2.2 Quasi Group Codes

Linear codes and group codes are two classes of structured codes. These codes are closed under the addition of the underlying group or field. It is known in the literature that coding schemes based on linear codes and group codes improve upon unstructured random coding strategies [7]. In this section, we propose a new class of structured codes called *quasi-group codes*.

A QGC is defined as a subset of a group code. Therefore, QGCs are not necessarily closed under the addition of the underlying group. An (n, k) shifted group code over $\mathbb{Z}_{p^r}$ is defined as the image of a linear mapping from $\mathbb{Z}_{p^r}^k$ to $\mathbb{Z}_{p^r}^n$ as in Definition 1. Let $\mathcal{U}$ be an arbitrary subset of $\mathbb{Z}_{p^r}^k$. Then a QGC is defined as

$$\mathcal{C} = \{\mathbf{uG} + \mathbf{b} : \mathbf{u} \in \mathcal{U}\}, \tag{2.2}$$

where $\mathbf{G}$ is a $k \times n$ matrix and $\mathbf{b}$ is an element of $\mathbb{Z}_{p^r}^n$. If $\mathcal{U} = \mathbb{Z}_{p^r}^k$, then $\mathcal{C}$ is a shifted group code. As we will show, by changing the subset $\mathcal{U}$, the code $\mathcal{C}$ ranges from completely structured codes (such as group codes and linear codes) where $|\mathcal{C}+\mathcal{C}| = |\mathcal{C}|$ to completely unstructured codes where $|\mathcal{C} + \mathcal{C}| \approx |\mathcal{C}|^2$. For a general subset $\mathcal{U}$, it is difficult to derive a single-letter characterization of the asymptotic performance of such codes. To address this issue, we present a special type of subsets $\mathcal{U}$ for which single-letter characterization of their performance is possible.

Construction of $\mathcal{U}$ Given a positive integer m, consider m mutually independent random variables $U_1, U_2, \cdots, U_m$. Suppose each U_i takes values from $\mathbb{Z}_{p^r}$ with distribution $P_{U_i}, i \in [1 : m]$. For $\epsilon > 0$, and positive integers k_i, define $\mathcal{U}$ as a Cartesian product of the ϵ-typical sets of $U_i, i \in [1 : m]$. More precisely,

$$\mathcal{U} \triangleq \bigotimes_{i=1}^{m} A_\epsilon^{(k_i)}(U_i). \tag{2.3}$$

In this construction, set $\mathcal{U}$ is determined by m, k_i, ϵ, and the PMFs $P_{U_i}, i \in [1:m]$. An example of such construction for $m = 1$ is given in the following.

Example 1. Let U be a random variable over $\mathbb{Z}_{p^r}$ with PMF P_U. For $\epsilon > 0$, let $\mathcal{U}$ to be the set of all ϵ-typical sequences $\mathbf{u}^k$. More precisely, define $\mathcal{U} = A_\epsilon^{(k)}(U)$. In this case, $\mathcal{U}$ is determined by the PMF P_U and ϵ. For instance, if U is uniform over $\mathbb{Z}_{p^r}$, then $\mathcal{U} = \mathbb{Z}_{p^r}^k$.

In what follows, we provide an alternative representation for the construction given in (2.3). Let $k \triangleq \sum_{i=1}^m k_i$ and denote $q_i \triangleq \frac{k_i}{k}$. With this notation, $q_i, i \in [1, m]$ form a probability distribution; because, $q_i \geq 0$ and $\sum_i q_i = 1$. Therefore, we can define a random variable Q with $P(Q = i) = q_i$. Define a random variable U with the conditional distribution $P(U = a|Q = i) = P(U_i = a)$ for all $a \in \mathbb{Z}_{p^r}, i \in [1:m]$. With this notation the set $\mathcal{U}$ in the above construction is characterized by a finite set $\mathcal{Q}$, a pair of random variables (U, Q) distributed over $\mathbb{Z}_{p^r} \times \mathcal{Q}$, an integer k, and $\epsilon > 0$. The joint distribution of U and Q is denoted by P_{UQ}. Note that we assume $P_Q(q) > 0$ for all $q \in \mathcal{Q}$. For a more concise notation, we identify the set $\mathcal{U}$ without explicitly specifying ϵ. Q can be interpreted as a *time sharing* random variable. It determines the contribution of U_i, measured by $\frac{k_i}{k}$, in the construction of $\mathcal{U}$. With the notation given for the construction of $\mathcal{U}$, we define its corresponding QGC.

Definition 3. *An (n, k)- QGC $\mathcal{C}$ over $\mathbb{Z}_{p^r}$ is defined as in (2.2) and* (2.3)*, and is characterized by a matrix $\mathbf{G} \in \mathbb{Z}_{p^r}^{k \times n}$, a translation $\mathbf{b} \in \mathbb{Z}_{p^r}^n$, and a pair of random variables (U, Q) distributed over the finite set $\mathbb{Z}_{p^r} \times \mathcal{Q}$. The set $\mathcal{U}$ in* (2.3) *is defined as the index set of $\mathcal{C}$.*

Remark 1. Any shifted group code over $\mathbb{Z}_{p^r}$ is a QGC.

Remark 2. Let $\mathcal{C}$ be a random (n, k)-QGC constructed by selecting the elements of its generator matrix and translation vector randomly independently with uniform

distribution from $\mathbb{Z}_{p^r}, r > 1$. In contrast to linear codes, codewords of $\mathcal{C}$ are not necessarily pairwise independent.

Information theoretic analysis of coding strategies are usually carried out by constructing ensembles of randomly generated codebooks [2, 5]. Following the same approach, we construct ensembles of QGCs with different blocklengths.

Fix positive integers (n, k) and random variables (U, Q). We create an ensemble of codes by taking the collection of all (n, k)-QGCs with random variables (U, Q), for all matrices $\mathbf{G}$ and translations $\mathbf{b}$. A random codebook $\mathcal{C}$ from this ensemble is chosen by selecting the elements of $\mathbf{G}$ and $\mathbf{b}$ randomly and uniformly from $\mathbb{Z}_{p^r}$. In order to characterize the asymptotic performance limits of QGCs, we need to define sequences of ensembles of QGCs. For any positive integer n, let $k_n = cn$, where $c > 0$ is a constant. Consider the sequence of the ensembles of (n, k_n)-QGCs with random variables (U, Q). In the next two lemmas, we characterize the size of randomly selected codebooks from these ensembles. The first lemma shows that the index set $\mathcal{U}$ for an ensemble of QGCs approximately equals to $2^{kH(U|Q)}$.

Lemma 1. *Let $\mathcal{U}_n$ be the index set associated with the ensemble of (n, k_n)-QGCs with random variables (U, Q) and $\epsilon > 0$, where $k_n = cn$ for a constant $c > 0$. Then there exists $N > 0$, such that for all $n > N$,*

$$\left| \frac{1}{k_n} \log_2 |\mathcal{U}_n| - H(U|Q) \right| \leq \epsilon',$$

where ϵ' is a continuous function of ϵ, and $\epsilon' \to 0$ as $\epsilon \to 0$.

Proof. The proof is given in Appendix A.1 □

Remark 3. As an immediate consequence of Lemma 1, we provide an upper-bound on the size of a QGC. For that, let $\mathcal{C}_n$ be an (n, k_n)-QGC with random variables (U, Q).

Then, for large enough n,

$$\frac{1}{n}\log_2|\mathcal{C}_n| \leq \frac{k_n}{n}H(U|Q) + \epsilon'. \tag{2.4}$$

To explain inequality (2.4), note that a codebook $\mathcal{C}_n$ is the image of the index set $\mathcal{U}_n$ under the mapping $\Phi_n(\mathbf{u}) = \mathbf{u}\mathbf{G}_n + \mathbf{b}^n$. Therefore, the bound in (2.4) is due to the fact that Φ_n is, in general, a many-to-one mapping. In the case of linear codes ($r = 1$), it is assumed that $k < n$. In this case, for sufficiently large n, Φ_n is injective with high probability. This implies that the size of a random linear code approximately equals $\approx 2^k$. Consequently, $\frac{k}{n}$ is a relevant measure for the rate of a (k, n) linear code. However, for a QGC (general $r \geq 2$), even if $k \geq n$, under certain conditions, Φ_n is "almost" injective with high probability. In what follows, we characterize these conditions. We begin by defining α-injectivity.

Definition 4. *A mapping $\phi : \mathcal{U} \to \mathcal{X}$, defined on finite sets $(\mathcal{U}, \mathcal{X})$, is said to be α-injective, if there exists a subset $\mathcal{A} \subseteq \mathcal{U}$ with cardinality at least $\alpha|\mathcal{U}|$ such that restriction of ϕ to $\mathcal{A}$ is injective.*

By the above definition, any 1-injective map is one-to-one. The next lemma shows that under particular conditions on (U, Q) and for sufficiently large n, the mapping Φ_n is α-injective with high probability, where $\alpha \approx 1$.

Lemma 2. *Let $\mathcal{U}_n$ be the index set associated with the ensemble of (n, k_n)-QGCs with random variables (U, Q), where $k_n = cn$ for a constant $c > 0$. Define a map $\Phi_n : \mathcal{U}_n \to \mathbb{Z}_{p^r}^n$, $\Phi_n(\mathbf{u}) = \mathbf{u}\mathbf{G}_n$ for all $\mathbf{u} \in \mathcal{U}_n$, where $\mathbf{G}_n$ is a $k_n \times n$ matrix whose elements are chosen randomly and uniformly from $\mathbb{Z}_{p^r}$. Suppose $H(U|[U]_s, Q) \leq \frac{1}{c}(r-s)\log_2 p - \epsilon$ for all $s \in [0 : r-1]$. Then, for any $\gamma, \delta > 0$ and sufficiently large n, the mapping Φ_n is $(1-\delta)$-injective with probability at least $(1-\gamma)$.* [1]

[1]Note that the map Φ_n in the lemma does not have any translation, i.e., $\mathbf{b} = 0$. It is sufficient to prove the lemma for $\mathbf{b} = 0$. This is due to the fact that if Φ_n is $(1-\delta)$-injective, then so is $\Phi_n + \mathbf{b}$, for any translation $\mathbf{b}$.

Proof. The proof is provided in Appendix A.2. □

As a result, under the conditions given in Lemma 2, the rate of a random codebook selected from ensemble of (n, k)-QGCs with random variables (U, Q) approximately equals $R \approx \frac{k}{n} H(U|Q)$, with high probability. The condition in Lemma 2 can viewed as a restriction on the size of the index set, that is

$$\frac{k}{n} H(U|[U]_s, Q) \leq (r - s) \log_2 p - \epsilon, \quad 0 \leq s \leq r - 1. \tag{2.5}$$

We refer to this condition as the *injectivity* condition.

2.3 Properties of Quasi Group Codes

It is known that if $\mathcal{C}$ is a random unstructured codebook, then $|\mathcal{C} + \mathcal{C}| \approx |\mathcal{C}|^2$ with high probability. Group codes on the other hand are closed under the addition, which means $|\mathcal{C} + \mathcal{C}| = |\mathcal{C}|$. Comparing to unstructured codes, when the structure of the group codes matches with that of a multi-terminal channel/source coding problem, it turns out that higher/lower transmission rates are obtained. However, in certain problems, the structure of the group codes is too restrictive. More precisely, when the underlying group is $\mathbb{Z}_{p^r}$ for $r \geq 2$, there are several nontrivial subgroups. These subgroups cause a penalty on the rate of a group code. This results in lower transmission rates in channel coding and higher transmission rates in source coding.

Quasi group codes balance the trade-off between the structure of the group codes and that of the unstructured codes. More precisely, when $\mathcal{C}$ is a QGC, then $|\mathcal{C} + \mathcal{C}|$ is a number between $|\mathcal{C}|$ and $|\mathcal{C}|^2$. This results in a more flexible algebraic structure to match better with the structure of the channel or source. This trade-off is shown more precisely in the following lemma.

Lemma 3. *Let $\mathcal{C}_i, i = 1, 2$ be an (n, k_i)-QGC over $\mathbb{Z}_{p^r}$ with random variables (U_i, Q).*

Suppose, $P_{U_1,U_2,Q}$ is such that the Markov chain $U_1 \leftrightarrow Q \leftrightarrow U_2$ holds and that the injectivity condition in (2.5) is satisfied for (U_1, Q) and (U_2, Q).

1. *Suppose $k_1 = k_2 = k$, and the generator matrices of $\mathcal{C}_1, \mathcal{C}_2$ and $\mathcal{D}$ are identical. Let $\mathcal{D}$ be an (n, k)-QGC with random variables $(U_1 + U_2, Q)$ and the same generator matrix as for $\mathcal{C}_1$ and $\mathcal{C}_2$. Suppose $\mathbf{U}_i$ is selected randomly and uniformly from the index set (see Definition 3) of $\mathcal{C}_i, i = 1, 2$. Let $\mathbf{X}_i$ be the codeword of $\mathcal{C}_i$ corresponding to $\mathbf{U}_i, i = 1, 2$. Then, for all $\epsilon > 0$ and sufficiently large n,*

$$P\{\mathbf{X}_1 + \mathbf{X}_2 \in \mathcal{D}\} \geq 1 - \delta(\epsilon),$$

where $\delta(\epsilon) \to 0$ as $\epsilon \to 0$.

2. *$\mathcal{C}_1 + \mathcal{C}_2$ is an $(n, k_1 + k_2)$-QGC with random variables $(U_I, (Q, I))$, where $I \in \{1, 2\}$. If $I = i$, then $U_I = U_i, i = 1, 2$. In addition, $P(I = i, Q = q, U_I = a) = \frac{k_i}{k_1+k_2} P(Q = q) P(U_i = a|Q = q)$, for all $a \in \mathbb{Z}_{p^r}, q \in \mathcal{Q}$ and $i = 1, 2$.*

Proof. Suppose $\mathcal{U}_i$ is the index set, $\mathbf{G}_i$ is the matrix, and $\mathbf{b}_i$ is the translation of $\mathcal{C}_i, i = 1, 2$.

We prove the first statement for the case when time sharing random variable Q is trivial. The proof for general Q follows from similar steps. If Q is trivial, the index sets satisfy $\mathcal{U}_i = A_\epsilon^{(k)}(U_i), i = 1, 2$. Since $k_1 = k_2$ and $\mathbf{G}_1 = \mathbf{G}_2$, then $\mathbf{X}_i = \mathbf{U}_i\mathbf{G} + \mathbf{b}_i, i = 1, 2$. With this notation, $\mathbf{X}_1 + \mathbf{X}_2 = (\mathbf{U}_1 + \mathbf{U}_2)\mathbf{G} + \mathbf{b}_1 + \mathbf{b}_2$. From Lemma 28, with probability at least $1 - 2^{-n\epsilon/p^r}$, we have $(\mathbf{U}_1, \mathbf{U}_2) \in A_{\delta(\epsilon)}^{(k)}(U_1, U_2)$, where δ is a function as in Lemma 28. Therefore, $\mathbf{U}_1 + \mathbf{U}_2 \in A_{\delta(\epsilon)}^{(k)}(U_1 + U_2)$ with probability at least $1 - 2^{-n\epsilon/p^r}$. The proof is complete by noting that the index set of $\mathcal{D}$ is defined as $\mathcal{U}_d \triangleq A_{\delta(\epsilon)}^{(k)}(U_1 + U_2)$.

For the second statement, we have

$$\mathcal{C}_1 + \mathcal{C}_2 = \{[\mathbf{u}_1, \mathbf{u}_2] \begin{bmatrix} \mathbf{G}_1 \\ \mathbf{G}_2 \end{bmatrix} + \mathbf{b}_1 + \mathbf{b}_2 : \mathbf{u}_i \in \mathcal{U}_i, i = 1, 2\}.$$

Therefore, $\mathcal{C}_1 + \mathcal{C}_2$ is an $(n, k_1 + k_2)$-QGC. Note that $\mathcal{U}_1 \times \mathcal{U}_2$ is the index set associated with this codebook. The statement follows, since each subset $\mathcal{U}_i, i = 1, 2$ is a Cartesian product of ϵ-typical sets of $U_{i,q}, q \in \mathcal{Q}$. The random variables $(U_I, (Q, I))$ describes such a Cartesian product.

□

We explain the intuition behind the lemma. Suppose $\mathcal{C}_1, \mathcal{C}_2$ and $\mathcal{D}$ are QGCs with identical generator matrices and with random variables U_1, U_2 and $U_1 + U_2$, respectively. Then $\mathcal{D} = \mathcal{C}_1 + \mathcal{C}_2$ with probability approaching one.

Remark 4. If $\mathcal{C}_1$ and $\mathcal{C}_2$ are the QGCs as in Lemma 3, then from standard counting arguments we have

$$\max\{|\mathcal{C}_1|, |\mathcal{C}_2|\} \leq |\mathcal{C}_1 + \mathcal{C}_2| \leq \min\{p^{rn}, |\mathcal{C}_1| \cdot |\mathcal{C}_2|\}$$

In what follows, we derive a packing bound and a covering bound for a QGC with matrices and translation chosen randomly and uniformly. Fix a PMF P_{XY}, and suppose an ϵ-typical sequence $\mathbf{y}$ is given with respect to the marginal distribution P_Y. Consider the set of all codewords in a QGC that are jointly typical with $\mathbf{y}$ with respect to P_{XY}. In the packing lemma, we characterize the conditions under which the probability of this set is small. This implies the existence of a "good-channel" code which is also a QGC. In the covering lemma, we derive the conditions for which, with high probability, there exists at least one such codeword in a QGC. In this case a "good-source" code exists which is also a QGC. These conditions are provided in the next two lemmas.

For any positive integer n, let $k_n = cn$, where $c > 0$ is a constant. Let $\mathcal{C}_n$ be a sequence of (n, k_n)-QGCs with random variables (U, Q), $\epsilon > 0$. By R_n denote the rate of $\mathcal{C}_n$. Suppose the elements of the generator matrix and the translation of $\mathcal{C}_n$ are chosen randomly and uniformly from $\mathbb{Z}_{p^r}$.

Lemma 4 (Packing)**.** *Let $(X, Y) \sim P_{XY}$. By $\mathbf{c}_n(\theta)$ denote the θth codeword of $\mathcal{C}_n$. Let $\tilde{\mathbf{Y}}^n$ be a random sequence distributed according to $\prod_{i=1}^n P_{Y|X}(\tilde{y}_i|c_{n,i}(\theta))$. Suppose, conditioned on $\mathbf{c}_n(\theta)$, $\tilde{\mathbf{Y}}^n$ is independent of all other codewords in $\mathcal{C}_n$. Then, for any $\theta \in [1 : |\mathcal{C}_n|]$, and $\delta > 0$, $\exists N > 0$ such that for all $n > N$,*

$$P\{\exists \mathbf{x} \in \mathcal{C}_n : (\mathbf{x}, \tilde{\mathbf{Y}}^n) \in A_\epsilon^{(n)}(X, Y), \mathbf{x} \neq \mathbf{c}_n(\theta)\} < \delta,$$

if the following bounds hold

$$R_n < \min_{0 \le s \le r-1} \frac{H(U|Q)}{H(U|Q, [U]_s)} \big(\log_2 p^{r-s} - H(X|Y, [X]_s) + \eta(\epsilon) \big), \tag{2.6}$$

where $\eta(\epsilon) \to 0$ as $\epsilon \to 0$.

Proof. See Appendix A.3. □

Lemma 5 (Covering)**.** *Let $(X, \hat{X}) \sim P_{X\hat{X}}$, where $\hat{X}$ takes values from $\mathbb{Z}_{p^r}$. Let $\mathbf{X}^n$ be a random sequence distributed according to $\prod_{i=1}^n P_X(x_i)$. Then, for any $\delta > 0$, $\exists N > 0$ such that for all $n > N$,*

$$P\{\exists \hat{\mathbf{x}} \in \mathcal{C}_n : (\mathbf{X}^n, \hat{\mathbf{x}}) \in A_\epsilon^{(n)}(X, \hat{X})\} > 1 - \delta$$

if the following inequalities hold

$$R_n > \max_{1 \le s \le r} \frac{H(U|Q)}{H([U]_s|Q)} \big(\log_2 p^s - H([\hat{X}]_s|X) + \eta(\epsilon) \big). \tag{2.7}$$

Proof. See Appendix A.4. □

Remark 5. The covering and packing bounds for the special of $r = 1$ are simplified to

$$\text{Packing: } R_n < \log_2 p - H(X|Y), \qquad \text{Covering: } R_n > \log_2 p - H(\hat{X}|X)$$

Lemma 3, 4 and Lemma 5 provide a tool to derive inner bounds for achievable rates using quasi group codes in multi-terminal channel coding and source coding problems.

2.4 Binning Using QGC

Note that in a randomly generated QGC, all codewords have uniform distribution over $\mathbb{Z}_{p^r}^n$. However, in many communication setups we require application of codes with non-uniform distributions. In addition, we require binning techniques for various multi-terminal communications. In this section, we present a method for random binning of QGCs. In the next sections, we will use random binning of QGCs to propose coding schemes for various multi-terminal problems.

We introduce nested quasi group codes using which we propose a random binning technique. A QGC $\mathcal{C}_I$ is said to be nested in a QGC $\mathcal{C}_O$, if $\mathcal{C}_I \subset \mathcal{C}_O + \mathbf{b}$, for some translation $\mathbf{b}$. Suppose $\mathcal{C}_O$ is an $(n, k+l)$-QGC with the following structure,

$$\mathcal{C}_O \triangleq \{\mathbf{u}\mathbf{G} + \mathbf{v}\tilde{\mathbf{G}} + \mathbf{b} : \mathbf{u} \in \mathcal{U}, \mathbf{v} \in \mathcal{V}\}, \tag{2.8}$$

where $\mathcal{U}$ and $\mathcal{V}$ are subsets of $\mathbb{Z}_{p^r}^k$, and $\mathbb{Z}_{p^r}^l$, respectively. Define the inner-code as

$$\mathcal{C}_I \triangleq \{\mathbf{u}\mathbf{G} + \mathbf{b} : \mathbf{u} \in \mathcal{U}\}.$$

By Definition 3, $\mathcal{C}_I$ is an (n, k)-QGC. In addition, there exists $\mathbf{a} \in \mathbb{Z}_{p^r}^n$ such that $\mathcal{C}_I \subset \mathcal{C}_O + \mathbf{a}$. The pair $(\mathcal{C}_I, \mathcal{C}_O)$ is called a nested QGC. For any fixed element $\mathbf{v} \in \mathcal{V}$,

we define its corresponding bin as the set

$$\mathcal{B}(\mathbf{v}) \triangleq \{\mathbf{u}\mathbf{G} + \mathbf{v}\tilde{\mathbf{G}} + \mathbf{b} : \mathbf{u} \in \mathcal{U}\}. \tag{2.9}$$

Definition 5. *An (n, k, l)-nested QGC is defined as a pair $(\mathcal{C}_I, \mathcal{C}_O)$, where $\mathcal{C}_I$ is an (n, k)-QGC, and $\mathcal{C}_O = \{\mathbf{x}_I + \bar{\mathbf{x}} : \mathbf{x}_I \in \mathcal{C}_I, \bar{\mathbf{x}} \in \bar{\mathcal{C}}\}$, where $\bar{\mathcal{C}}$ is an (n, l)-QGC. Let the random variables corresponding to $\mathcal{C}_I$ and $\bar{\mathcal{C}}$ are (U, Q) and (V, Q), respectively. $\mathcal{C}_I, \mathcal{C}_O$ and $\bar{\mathcal{C}}$ are called the inner, the outer and the shift codes, respectively. Then, $\mathcal{C}_O$ is characterized by (U, V, Q).*

In a nested QGC both the outer-code and the inner-code are themselves QGCs. More precisely we have the following remark.

Remark 6. Let $(\mathcal{C}_I, \mathcal{C}_O)$ be an (n, k_1, k_2)-nested QGC with random variables (U_1, U_2, Q). Suppose the joint distribution among (U_1, U_2, Q) is the one that satisfies the Markov chain $U_1 \leftrightarrow Q \leftrightarrow U_2$. Then by Lemma 3 $\mathcal{C}_O$ is an $(n, k_1 + k_2)$-QGC with random variables $(U_I, (Q, I))$.

Note that with equation (2.9), $\mathcal{B}(\mathbf{v}) = \mathcal{C}_I + \mathbf{v}\tilde{\mathbf{G}}$. As a result, each bin is a shifted version of the inner-code. Thus, each bin in an (n, k, l)-nested QGC is also an (n, k)-QGC.

Remark 7. Suppose $(\mathcal{C}_I, \mathcal{C}_O)$ is an (n, k_1, k_2)-nested QGC with random matrices and translations. Assume the injectivity condition (2.5) holds for C_I and C_O. By R_O and R_I denote the rates of $\mathcal{C}_O$ and $\mathcal{C}_I$, respectively. Let ρ denote the binning rate (the rate of $\bar{\mathcal{C}}$ as in Definition 5). Using Remark 6 and 3, for large enough n, with probability close to one, $|R_O - R_I - \rho| \leq o(\epsilon)$.

Intuitively, as a result of this remark, $R_O \approx R_I + \rho$. Furthermore, since the injectivity condition holds, then with probability close to one,

$$R_O \approx \frac{k}{n} H(U|Q) + \frac{l}{n} H(V|Q), \qquad R_I \approx \frac{k}{n} H(U|Q), \quad \text{and} \quad \rho \approx \frac{l}{n} H(V|Q).$$

This implies that the bins $\mathcal{B}(\mathbf{v})$ corresponding to different $\mathbf{v} \in \bar{\mathcal{C}}$ are "almost disjoint". In this method for binning, since both the inner-code and the outer-code are QGCs, the structure of the inner-code, bins and the outer-code can be determined using the PMFs of the related random variables (that is U, V and Q as in Definition 5).

PtP Communications

We established a set of lemmas (Lemma 1- 5) that are used to derive achievable rates for coding strategies based on QGCs. In the following, we introduce a coding strategy using QGCs and show the achievability of the Shannon performance limits for PtP channel and source coding problem. For that, we first provide a set of definitions to model PtP channel and source coding problem.

Channel Model: A discrete memoryless channel is characterized by the triple $(\mathcal{X}, \mathcal{Y}, P_{Y|X})$, where the two finite sets $\mathcal{X}$ and $\mathcal{Y}$ are the input and output alphabets, respectively, and $P_{Y|X}$ is the channel transition probability matrix.

Definition 6. *An (n, Θ)-code for a channel $(\mathcal{X}, \mathcal{Y}, P_{Y|X})$ is a pair of mappings (e, f) where $e : [1 : \Theta] \to \mathcal{X}^n$ and $f : \mathcal{Y}^n \to [1 : \Theta]$.*

Definition 7. *For a given channel $(\mathcal{X}, \mathcal{Y}, P_{Y|X})$, a rate R is said to be achievable if for any $\epsilon > 0$ and for all sufficiently large n, there exists an (n, Θ)-code such that :*

$$\frac{1}{\Theta} \sum_{i=1}^{\Theta} P_{Y|X}^n(f(Y^n) \neq i | X^n = e(i)) < \epsilon, \quad \frac{1}{n} \log \Theta > R - \epsilon.$$

The channel capacity is defined as the supremum of all achievable rates.

Source Model: A discrete memoryless source is a tuple $(\mathcal{X}, \hat{\mathcal{X}}, P_X, d)$, where the two finite sets $\mathcal{X}$ and $\hat{\mathcal{X}}$ are the source and reconstruction alphabets, respectively, P_X is the source probability distribution, and $d : \mathcal{X} \times \hat{\mathcal{X}} \to \mathbb{R}^+$ is the (bounded) distortion function.

Definition 8. *An (n, Θ)-code for a source $(\mathcal{X}, \hat{\mathcal{X}}, P_X, d)$ is a pair of mappings (e, f) where $f : \mathcal{X}^n \to [1 : \Theta]$ and $e : [1 : \Theta] \to \hat{\mathcal{X}}^n$.*

Definition 9. *For a given source $(\mathcal{X}, \hat{\mathcal{X}}, P_X, d)$, a rate-distortion pair (R, D) is said to be achievable if for any $\epsilon > 0$ and for all sufficiently large n, there exists an (n, Θ)-code such that :*

$$\frac{1}{n}\sum_{i=1}^{n} d(X_i, \hat{X}_i) < D + \epsilon, \quad \frac{1}{n}\log\Theta < R + \epsilon,$$

where $\hat{X}^n = e(f(X^n))$. The optimal rate-distortion region is defined as the set of all achievable rate-distortion pairs.

Definition 10. *An (n, Θ)-code is said to be based on nested QGCs, if there exists an (n, k, l)-nested QGC with random variables (U, V, Q) such that a) $\Theta = |\mathcal{V}|$, where $\mathcal{V}$ is the index set associated with the codebook $\bar{\mathcal{C}}$ (see Definition 5), b) for any $\mathbf{v} \in \mathcal{V}$, the output of the mapping $e(\mathbf{v})$ is in $\mathcal{B}(\mathbf{v})$, where $\mathcal{B}(\mathbf{v})$ is the bin associated with $\mathbf{v}$, and is defined as in* (2.9).

Definition 11. *For a channel, a rate R is said to be achievable using nested QGCs if for any $\epsilon > 0$ and all sufficiently large n, there exists an (n, Θ)-code based on nested QGCs such that:*

$$\frac{1}{\Theta}\sum_{i=1}^{\Theta} P(f(Y^n) \neq i | X^n = e(i)) < \epsilon, \quad \frac{1}{n}\log\Theta > R - \epsilon.$$

For a source, a rate-distortion pair (R, D) is said to be achievable using nested QGSs, if for any $\epsilon > 0$ and for all sufficiently large n, there exists an (n, Θ)-code based on nested QGCs such that:

$$\frac{1}{n}\sum_{i=1}^{n} d(X_i, \hat{X}_i) < D + \epsilon, \quad \frac{1}{n}\log\Theta < R + \epsilon,$$

where $\hat{X}^n = e(f(X^n))$.

Theorem II.1. *The PtP channel capacity and the optimal rate-distortion region of sources are achievable using nested QGCs.*

In what follows, we introduce an achievable scheme using nested QGCs and provide an outline of the proof for the theorem.

Channel coding using QGCs Consider a memoryless channel with input alphabet $\mathcal{X}$ and conditional distribution $P_{Y|X}$. Let the prime power p^r be such that $|\mathcal{X}| \leq p^r$. Fix a PMF P_X on $\mathcal{X}$, and set $l = nR$, where R will be determined later. Let $(\mathcal{C}_I, \mathcal{C}_O)$ be an (n, k, l)-nested QGC with random variables (U, V, Q). Let Q be a trivial random variable, and U and V be independent with uniform distribution over $\{0, 1\}$. The elements of the generator matrix and the translation used for the nested QGC are drawn randomly and uniformly from $\mathbb{Z}_{p^r}$. Let R_I and R_O denote the rate of the inner-code $\mathcal{C}_I$ and the outer-code $\mathcal{C}_O$, respectively. According to Remark 7, with probability close to one, $R_O \approx R_I + R$ and the binning rate approximately equals to $\frac{l}{n}H(V) = R$.

Suppose the messages are drawn randomly and uniformly from $\{0, 1\}^l$. Upon receiving a message $\mathbf{v}$, the encoder first calculates its bin, that is $\mathcal{B}(\mathbf{v})$. Then it finds $\mathbf{x} \in \mathcal{B}(\mathbf{v})$ such that $\mathbf{x} \in A_\epsilon^{(n)}(X)$. If $\mathbf{x}$ was found, it is transmitted to the channel. Otherwise, an encoding error is declared. Upon receiving $\mathbf{y}$ from the channel, the decoder finds all $\tilde{\mathbf{c}} \in \mathcal{C}_O$ such that $(\tilde{\mathbf{c}}, \mathbf{y}) \in A_\epsilon^{(n)}(X, Y)$. Then, the decoder lists the bin number for any of such $\tilde{\mathbf{c}}$. If the bin number is unique, it is declared as the decoded message. Otherwise, an encoding error will be declared.

The effective transmission of the above coding strategy equals the binning rate, i.e., R. Using the covering lemma (Lemma 5), the probability of the error at the encoder approaches zero, if $R_I \geq \log p^r - H(X)$. Using the packing lemma (Lemma 4), the probability of error at the decoder approaches zero, if $R_O \leq \log p^r - H(X|Y)$.

As a result, the effective transmission rate $R \leq I(X;Y)$ is achievable.

Source coding using QGCs We use the same nested QGC constructed for the channel coding problem. Given a distortion level D, consider a random variable $\hat{X}$ such that $\mathbb{E}\{d(X,\hat{X})\} \leq D$. Let $\mathbf{x}$ be a typical sequence from the source. The encoder finds a codeword $\mathbf{c} \in \mathcal{C}_O$ such that $(\mathbf{x}, \mathbf{c})$ is jointly ϵ-typical with respect to $P_X P_{\hat{X}|X}$. If no such $\mathbf{c}$ was found, an encoding error will be declared. Otherwise, the encoder sends the bin index $\mathbf{v}$ for which $\mathbf{c} \in \mathcal{B}(\mathbf{v})$. Given $\mathbf{v}$, the decoder finds $\tilde{\mathbf{c}} \in \mathcal{B}(\mathbf{v})$ such that $\tilde{\mathbf{c}}$ is ϵ-typical with respect to $P_{\hat{X}}$. An error occurs, if no unique codeword $\tilde{\mathbf{c}}$ was found.

Note that with high probability the effective transmission rate approximately equals to R. Using Lemma 5, the encoding error approaches zero, if $R_O \geq \log p^r - H(\hat{X}|X)$. Using Lemma 4, the decoding error approaches zero, if $R_I \leq \log p^r - H(\hat{X})$. As a result the rate $R \geq I(X;\hat{X})$ and distortion D is achievable.

2.5 Distributed Source Coding

In this section, we consider a distributed source coding problem described as follows. Suppose X_1 and X_2 are sources with alphabet $\mathbb{Z}_{p^r}$ and with joint PMF $P_{X_1X_2}$. The jth encoder compresses X_j and sends it to a central decoder. The decoder wishes to reconstruct $X_1 + X_2$ losslessly, where the addition is modulo-p^r. Figure 2.1 depicts the diagram of this setup.

It is assumed that n IID copies of the sources are made available at the encoders, where n is called the blocklength. In what follows, we define the encoding and decoding processes and formulate the problem setup.

Definition 12. *An (n, Θ_1, Θ_2)-code consists of two encoding functions*

$$f_i : \mathbb{Z}_{p^r}^n \to \{1, 2, \cdots, \Theta_i\}, \quad i = 1, 2,$$

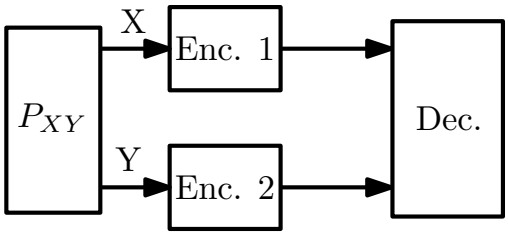

Figure 2.1: An example for the problem of distributed source coding. In this setup, the sources X_1 and X_2 take values from $\mathbb{Z}_{p^r}$. The decoder reconstructs $X_1 + X_2$ losslessly.

and a decoding function

$$g : \{1, 2, \cdots, \Theta_1\} \times \{1, 2, \cdots, \Theta_2\} \to \mathbb{Z}_{p^r}^n$$

Definition 13. *Given a pair of sources $(X_1, X_2) \sim P_{X_1X_2}$ with values over $\mathbb{Z}_{p^r} \times \mathbb{Z}_{p^r}$, a pair (R_1, R_2) is said to be achievable if for any $\epsilon > 0$ and sufficiently large n , there exists an (n, Θ_1, Θ_2)-code such that,*

$$\frac{1}{n} \log_2 M_i < R_i + \epsilon, \quad for \;\; i = 1, 2, \quad and \quad P\{\mathbf{X_1}^n + \mathbf{X_2}^n \neq g(f_1(\mathbf{X_1}^n), f_2(\mathbf{X_2}^n))\} \leq \epsilon.$$

For this problem, we adopt nested QGCs and propose a new coding scheme. The following theorem presents an achievable rate region for the defined setup.

Theorem II.2. *For a pair of sources $(X_1, X_2) \sim P_{X_1X_2}$ with values from $\mathbb{Z}_{p^r}$, lossless reconstruction of the modulo-p^r sum $X_1 + X_2$ is possible with transmission rate-pair (R_1, R_2), if there exist random variables (W_1, W_2, Q) such that the following bound holds*

$$R_i \geq \log_2 p^r - \min_{0 \leq s \leq r-1} \frac{H(W_i|Q)}{H(W_1 + W_2|[W_1 + W_2]_s, Q)} (\log_2 p^{(r-s)} - H(X_1 + X_2|[X_1 + X_2]_s)), \tag{2.10}$$

where $i = 1, 2$, (W_1, W_2) take values from $\mathbb{Z}_{p^r}$, the Markov chain $W_1 - Q - W_2$ holds,

and the injectivity condition (2.5) *is satisfied for each pair* (W_1, Q) *and* (W_2, Q)*. In addition,* $|\mathcal{Q}| \leq r$ *is sufficient to achieve the above bounds.*

Proof. See Appendix A.5. □

Remark 8. The intuition for the rate-region can be briefly explained as follows. Each source is encoded using a nested QGC. The source covering task constrains the rate of the outer code. The packing task induced by the need to recover the sum $(X_1 + X_2)$ at the decoder constrains the rate of the inner code. The overall rates of transmission is given by the difference between these two rates.

Every linear code and group code is a QGC. Therefore, the achievable rate region given in Theorem II.2 subsumes the one achieved using linear codes or group codes with jointly typical encoding/decoding techniques. We show, through the following example, that the inclusion is strict.

Example 2. Consider a distributed source coding problem in which X_1 and X_2 are sources over $\mathbb{Z}_4$ and lossless reconstruction of $X_1 \oplus_4 X_2$ is required at the decoder. Assume X_1 is uniform over $\mathbb{Z}_4$. X_2 is related to X_1 via the equation $X_2 = N - X_1$, where N is a random variable which is independent of X_1. The distribution of N is presented in Table 3.1.

Table 2.1: Distribution of N

N	0	1	2	3
P_N	0.06	0.54	0.04	0.36

Using random unstructured codes, the rates (R_1, R_2) such that $R_1 + R_2 \geq H(X_1, X_2)$ are achievable [36]. It is also possible to use linear codes for the reconstruction of $X_1 \oplus_4 X_2$. For that, the decoder first reconstructs the modulo-7 sum of X_1 and X_2, then from $X_1 \oplus_7 X_2$ the modulo-4 sum is retrieved. This is because linear codes are built only over finite fields, and $\mathbb{Z}_7$ is the smallest field in which the modulo-4

addition can be embedded. Therefore, the rates $R_1 = R_2 \geq H(X_1 \oplus_7 X_2)$ is achievable using linear codes over the field $\mathbb{Z}_7$ [7]. As is shown in [67], group codes in this example outperform linear codes. The largest achievable region using group codes is described by all rate pair (R_1, R_2) such that $R_i \geq \max\{H(Z), 2H(Z|[Z]_1)\},\ i = 1, 2$, where $Z = X_1 \oplus_4 X_2$. It is shown in [16] that using transversal group codes the rates (R_1, R_2) such that $R_i \geq \max\{H(Z), 1/2H(Z) + H(Z|[Z]_1)\}$ are achievable. An achievable rate region using nested QGC's can be obtained from Theorem II.2. Let Q be a trivial random variable and set $P(W_1 = 0) = P(W_2 = 0) = 0.95$ and $P(W_1 = 1) = P(W_2 = 1) = 0.05$. As a result one can verify that the following is achievable:

$$R_j \geq 2 - \min\{0.6(2 - H(Z)), 5.7(2 - 2H(Z|[Z]_1)\}.$$

Note that the factors 0.6 and 5.7 are determined by the specific choice of the probability distribution on (W_1, Q) and (W_2, Q). Different factor are obtained by changing the probability distributions. We compare the achievable rates of these schemes. The result are presented in Table 2.2.

Table 2.2: Achievable sum-rate using different coding schemes for Example 2. Note that $Z \triangleq X_1 \oplus_4 X_2$.

Scheme	Achievable Rate	
Unstructured Codes	$H(X_1, X_2)$	3.44
Linear Codes	$H(X_1 \oplus_7 X_2)$	4.12
Group Codes	$\max\{H(Z), 2H(Z\|[Z]_1)\}$	3.88
QGCs	$2 - \min\{0.6(2 - H(Z)), 5.7(2 - 2H(Z\|[Z]_1)\}$	3.34

2.6 Computation Over MAC

In this section, we consider the problem of computation over MAC. Figure 2.2 depicts an example of this problem. In this setup X_1 and X_2 are the channel's inputs, and take values from $\mathbb{Z}_{p^r}$. Two distributed encoders map their messages to X_1^n and

X_2^n. Upon receiving the channel output the decoder wishes to decode $X_1^n + X_2^n$ losslessly. The definition of a code for computation over MAC, and an achievable rate are given in Definition 15 and 16, respectively. Applications of this problem are found in various multi-user communication setups such as interference and broadcast channels.

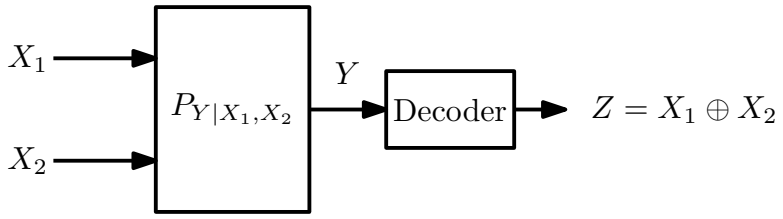

Figure 2.2: An example for the problem of computation over MAC. The channel input alphabets belong to $\mathbb{Z}_{p^r}$. The receiver decodes $X_1 + X_2$ which is the modulo-p^r sum of the inputs of the MAC.

Definition 14. *A two-user MAC is a tuple $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}, P_{Y|X_1X_2})$, where the finite sets $\mathcal{X}_1, \mathcal{X}_2$ are the inputs alphabets, $\mathcal{Y}$ is the output alphabet, and $P_{Y|X_1X_2}$ is the channel transition probability matrix. Without loss of generality, it is assumed that $\mathcal{X}_1 = \mathcal{X}_2 = \mathbb{Z}_{p^r}$, for a prime-power p^r.*

Definition 15 (Codes for computation over MAC)**.** *An (n, Θ_1, Θ_2)-code for computation over a MAC $(\mathbb{Z}_{p^r}, \mathbb{Z}_{p^r}, \mathcal{Y}, P_{Y|X_1X_2})$ consists of two encoding functions and one decoding function $f_i : [1 : \Theta_i] \to \mathbb{Z}_{p^r}^n$, for $i = 1, 2$, and $g : \mathcal{Y}^n \to \mathbb{Z}_{p^r}^n$, respectively.*

Definition 16 (Achievable Rate)**.** *(R_1, R_2) is said to be achievable, if for any $\epsilon > 0$, there exists for all sufficiently large n an (n, Θ_1, Θ_2)-code such that*

$$P\{g(Y^n) \neq f_1(M_1) + f_2(M_2)\} \leq \epsilon, \quad R_i - \epsilon \leq \frac{1}{n} \log \Theta_i,$$

where M_1 and M_2 are independent random variables and $P(M_i = m_i) = \frac{1}{\Theta_i}$ for all $m_i \in [1 : \Theta_i], i = 1, 2$.

For the above setup, we use QGCs to derive an achievable rate region.

Theorem II.3. *Given a MAC* $(\mathbb{Z}_{p^r}, \mathbb{Z}_{p^r}, \mathcal{Y}, P_{Y|X_1X_2})$*, rate-pair* (R_1, R_2) *is achievable according to Definition 16, if there exist random variables* $(Q, X_1, X_2, V_1, V_2, W_1, W_2)$ *such that the following bounds hold*

$$R_i \leq \min_{0\leq s\leq r} \frac{H(V_i|Q)}{H(V|[V]_s, Q)} \Bigg(\log_2 p^{r-s} - H(X|Y, [X]_s) \\ - \max_{\substack{1\leq t\leq r \\ j=0,1}} \frac{H(W|[W]_s, Q)}{H([W_j]_t|Q)} \left(\log_2 p^t - H([X_j]_t)\right) \Bigg)$$

where $i = 1, 2$*,* (V_1, V_2, W_1, W_2) *take values from* $\mathbb{Z}_{p^r}$*, and* $W = W_1 + W_2, V = V_1 + V_2, X = X_1 + X_2$*. Moreover, the injectivity condition* (2.5) *is satisfied for each pair* $(W_1, Q), (W_2, Q), (V_1, Q)$*, and* (V_2, Q) *and the joint PMF of all the random variables factors as*

$$P_{QX_1X_2V_1V_2W_1W_2Y} = P_{X_1}P_{X_2}P_Q P_{Y|X_1X_2} \prod_{i=1}^{2} P_{V_i|Q} P_{W_i|Q}.$$

Remark 9. The cardinality bound $|\mathcal{Q}| \leq r^2$ is sufficient to achieve the rate region in the theorem.

Proof. See Appendix A.6. □

Corollary 1. *A special case of the theorem is when* X_1 *and* X_2 *are distributed uniformly over* $\mathbb{Z}_{p^r}$*. In this case, the following is achievable*

$$R_i \leq \min_{0\leq s\leq r} \frac{H(V_i|Q)}{H(V_1 + V_2|[V_1 + V_2]_s, Q)} I(X_1 + X_2; Y|[X_1 + X_2]_s), \quad i = 1, 2, \tag{2.11}$$

We show, through the following example, that QGC outperforms the previously known schemes.

Example 3. Consider the MAC described by $Y = X_1 + X_2 + N$, where X_1 and X_2 are the channel inputs with alphabet $\mathbb{Z}_4$. N is independent of X_1 and X_2 with the distribution given in Table 3.1.

Using standard unstructured codes the rate pair (R_1, R_2) satisfying $R_1 + R_2 \leq I(X_1X_2;Y)$ are achievable. Note that the modulo-4 addition can be embedded in a larger field such as $\mathbb{Z}_7$. For that linear codes over $\mathbb{Z}_7$ can be used. In this case, the following rates are achievable:

$$R_1 = R_2 = \max_{P_{X_1}P_{X_2}:X_1,X_2\in\mathbb{Z}_4} \min\{H(X_1), H(X_2)\} - H(X_1 \oplus_7 X_2|Y),$$

where the maximization is taken over all probability distribution $P_{X_1}P_{X_2}$ on $\mathbb{Z}_7 \times \mathbb{Z}_7$ such that $P(X_i \in \mathbb{Z}_4) = 1, , i = 1, 2$. This is because, $\mathbb{Z}_4$ is the input alphabet of the channel.

It is shown in [67] that the largest achievable region using group codes is

$$R_i \leq \min\{I(Z;Y), 2I(Z;Y|[Z]_1)\},$$

where $Z = X_1 + X_2$ and X_1 and X_2 are uniform over $\mathbb{Z}_4$. Using Corollary 1, QGC's achieve $R_i \leq \min\{0.6I(Z;Y), 5.7I(Z;Y|[Z]_1)\}$. This can be verified by checking (2.11) when Q is a trivial random variable, $P(V_1 = 0) = P(V_2 = 0) = 0.95$ and $P(V_1 = 1) = P(V_2 = 1) = 0.05$. Note that the factors 0.6 and 5.7 are determined by the specific choice of the probability distribution on (W_1, Q) and (W_2, Q). Different factors can be obtained by changing the probability distributions. We compare the achievable rates of these schemes for the explained setup. The result are presented in Table 2.3.

Table 2.3: Achievable rates using different coding schemes for Example 3. Note that $Z \triangleq X_1 + X_2$.

Scheme	Achievable Rate $(R_1 = R_2)$	
Unstructured Codes	$I(X_1X_2;Y)/2$	0.28
Linear codes	$\min\{H(X_1), H(X_2)\} - H(X_1 \oplus_7 X_2\|Y)$	0.079
Group Codes	$\min\{I(Z;Y), 2I(Z;Y\|[Z]_1)\}$	0.06
QGCs	$\min\{0.6I(Z;Y), 5.7I(Z;Y\|[Z]_1)\}$	0.33

2.7 MAC with States

2.7.1 Model

Consider a two-user discrete memoryless MAC with input alphabets $\mathcal{X}_1, \mathcal{X}_2$, and output alphabet $\mathcal{Y}$. The transition probabilities between the input and the output of the channel depends on a random vector (S_1, S_2) which is called state. Figure 2.3 demonstrates such setup. Each state S_i takes values from a set $\mathcal{S}_i$, where $i = 1, 2$. The sequence of the states is generated randomly according to the probability distribution $\prod_{i=1}^{n} P_{S_1 S_2}$. The entire sequence of the state S_i is known at the ith transmitter, $i = 1, 2$, non-causally. The conditional distribution of Y given the inputs and the state is $P_{Y|X_1 X_2 S_1 S_2}$. Each input X_i is associated with a state dependent cost function $c_i : \mathcal{X}_i \times \mathcal{S}_i \to [0, +\infty)$[2]. The cost associated with the sequences x_i^n and s_i^n is given by

$$\bar{c}_i(x_i^n, s_i^n) = \frac{1}{n} \sum_{j=1}^{n} c_i(x_{ij}, s_{ij}).$$

Definition 17. *An (n, Θ_1, Θ_2)-code for reliable communication over a given two-user MAC with states is defined by two encoding functions*

$$f_i : \{1, 2, \ldots, \Theta_i\} \times \mathcal{S}_i^n \to \mathcal{Y}^n, \quad i = 1, 2,$$

and a decoding function

$$g : \mathcal{Y}^n \to \{1, 2, \ldots, \Theta_1\} \times \{1, 2, \ldots, \Theta_2\}.$$

Definition 18. *For a given MAC with state, the rate-cost tuple$(R_1, R_2, \tau_1, \tau_2)$ is said to be achievable, if for any $\epsilon > 0$, and for all large enough n there exists an (n, Θ_1, Θ_2)-*

[2]We use a cost function for this problem because, in many cases without a cost function the problem has a trivial solution.

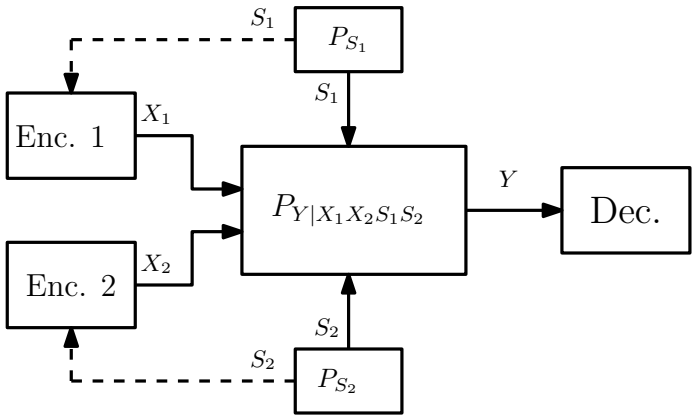

Figure 2.3: A two-user MAC with distributed states. The states (S_1, S_2) are generated randomly according to $P_{S_1S_2}$. The entire sequence of each state S_i is available non-casually at the ith transmitter, where $i = 1, 2$.

code such that

$$P\{g(Y^n) \neq (M_1, M_2)\} \leq \epsilon, \quad \frac{1}{n}\log \Theta_i \geq R_i - \epsilon, \quad \mathbb{E}\{\bar{c}_i(f_i(M_i), S_i^n)\} \leq \tau_i + \epsilon,$$

for $i = 1, 2$, where a) M_1, M_2 are independent random variables with distribution $P(M_i = m_i) = \frac{1}{\Theta_i}$ for all $m_i \in [1 : \Theta_i]$, b) (M_1, M_2) is independent of the states (S_1, S_2). Given τ_1, τ_2, the capacity region $\mathcal{C}_{\tau_1,\tau_2}$ is defined as the set of all rates (R_1, R_2) such that the rate-cost $(R_1, R_2, \tau_1, \tau_2)$ is achievable.

2.7.2 Achievable Rates

We propose a structured coding scheme that builds upon QGC. Then we present the single-letter characterization of the achievable region of this coding scheme. Using this binning method, a rate region is given in the following theorem.

Theorem II.4. *For a given MAC $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}, P_{Y|X_1X_2})$ with independent states (S_1, S_2)*

and cost functions c_1, c_2 the following rates are achievable using nested-QGC

$$R_1 + R_2 \leq r \log_2 p - H(Z_1 + Z_2|Y, Q) - \max_{\substack{i=1,2 \\ 1\leq t\leq r}} \Big\{ \frac{H(V_1 + V_2|Q)}{H([V_i]_t|Q)} \Big(\log_2 p^t - H([Z_i]_t|Q, S_i) \Big) \Big\},$$

where the joint distribution of the above random variables factors as

$$P_{S_1S_2} P_Q P_{Y|X_1X_2} \prod_{i=1,2} P_{V_i|Q} P_{Z_i|QS_i} P_{X_i|QZ_iS_i}.$$

Proof. Let $\mathcal{C}_{I,j}$ be an (n, k)-QGC with matrix $\mathbf{G}_j$, translation $\mathbf{b}_j$, and random variables (W_j, Q), where W_j is uniform over $\{0, 1\}$, and $j = 1, 2$. Denote $\mathcal{W}_1$ and $\mathcal{W}_2$ as the index sets associated with $\mathcal{C}_{I,1}$ and $\mathcal{C}_{I,1}$, as in (2.2). Let $\bar{\mathcal{C}}_1, \bar{\mathcal{C}}_2$ and $\bar{\mathcal{D}}$ be three (n, l) QGC with identical matrices $\bar{\mathbf{G}}$ and identical translations $\bar{\mathbf{b}}$. Suppose (V_j, Q) are the random variables associated with $\bar{\mathcal{C}}_j$, where $j = 1, 2$. Furthermore, let $(V_1 + V_2, Q)$ is the random variable associated with $\bar{\mathcal{D}}$. Suppose that the elements of all the matrices and the translations are selected randomly and uniformly from $\mathbb{Z}_{p^r}$. Rate of $\bar{\mathcal{C}}_i$ is denoted by ρ_i, rate of $\bar{\mathcal{D}}$ is denoted by ρ, and that of $\mathcal{C}_{I,i}$ is $R_i, i = 1, 2$. For each, sequence $\mathbf{z}_i$ and $\mathbf{s}_i$, generate a sequence $\mathbf{x}_i$ randomly with IID distribution according to $P^n_{X_i|Z_iS_i}, i = 1, 2$. Denote such sequence by $x_i(\mathbf{s}_i, \mathbf{z}_i)$.

Codebook Construction: For each encoder we use a nested QGC. For the first encoder, we use the (n, k, l)nested QGC generated by $\mathcal{C}_{I,1}$ and $\bar{\mathcal{C}}_1$. For the second encoder, we use the (n, k, l)-nested QGC characterized by $\mathcal{C}_{I,2}$ and $\bar{\mathcal{C}}_2$. The codebook used in the decoder is $\mathcal{C}_{I,1} + \mathcal{C}_{I,2} + \bar{\mathcal{D}}$. By Lemma 3 , this codebook is an $(n, 2k + l)$-QGC. In addition, the rate of such code is $R_1 + R_2 + \rho$

Encoding: For $i = 1, 2$, the ith encoder is given a message θ_i, and an state sequence $\mathbf{s}_i$. The encoder first calculates the bin associated with θ_i. Then it finds a codeword $\mathbf{z}_i$ in that bin such $(\mathbf{z}_i, \mathbf{s}_i)$ are jointly ϵ-typical with respect to $P_{Z_iS_i}$. If no such sequence was found, the error event E_i will be declared. The encoder calculates $\mathbf{x}_i(\mathbf{s}_i, \mathbf{z}_i)$, and sends it through the channel. Define the event E_c as the event in

which $(\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{s}_1, \mathbf{s}_2)$ are not jointly ϵ'- typical with respect to the joint distribution $P_{Z_1 Z_2 S_1 S_2}$.

Decoding: The decoder receives y^n from the channel. Then it finds $\tilde{\mathbf{w}}_1 \in \mathcal{W}_1, \tilde{\mathbf{w}}_2 \in \mathcal{W}_2$, and $\tilde{\mathbf{v}} \in A_\epsilon^{(n)}(V_1 + V_2)$ such that the corresponding codeword defined as

$$\tilde{\mathbf{z}} = \tilde{\mathbf{w}}_1 \mathbf{G}_1 + \tilde{\mathbf{w}}_2 \mathbf{G}_2 + \tilde{\mathbf{v}} \bar{\mathbf{G}} + \mathbf{b}_1 + \mathbf{b}_2 + \bar{\mathbf{b}}$$

is jointly $\tilde{\epsilon}$-typical with $\mathbf{Y}$ with respect to $P_{Z_1+Z_2,Y}$. If $\tilde{\mathbf{w}}_1, \tilde{\mathbf{w}}_2$ are unique, then they are considered as the decoded messages. Otherwise an error event E_d will be declared.

Error Analysis: We use Lemma 5 for E_1 and E_2. For that in the covering bound given in (2.7) set $R = \rho_i, U = V_i, Q = \bar{Q}, \hat{X} = X_i$, and $X = S_i$, where $i = 1, 2$. As a result, $P(E_1)$ and $P(E_2)$ approaches zero as $n \to \infty$, if the covering bound holds:

$$\rho_i > \max_{1 \le t \le r} \frac{H(V_i|\bar{Q})}{H([V_i]_t|\bar{Q})} (\log_2 p^t - H([Z]_t|S_i)).$$

Note that by Remark 3, $\rho_i \le \frac{l}{n} H(V_i|\bar{Q}) + \delta(\epsilon)$. Thus, the above bound gives the following bound

$$\frac{l}{n} H([V_i]_t|\bar{Q}) > \log_2 p^t - H([Z]_t|S_i), \ 1 \le t \le r, \ i = 1, 2. \tag{2.12}$$

Analysis of $E_c \cap E_1^c \cap E_2^c$ Define the set

$$\mathcal{E}_{\mathbf{s}_1,\mathbf{s}_2} \triangleq \Big\{ (\mathbf{z}_1, \mathbf{z}_2) \in \mathbb{Z}_{p^r}^n \times \mathbb{Z}_{p^r}^n : (\mathbf{z}_i, \mathbf{s}_i) \in A_\epsilon^{(n)}(Z_i, S_i), \\ (\mathbf{z}_1, \mathbf{z}_2, \mathbf{s}_1, \mathbf{s}_2) \notin A_\epsilon^{(n)}(Z_1, Z_2, S_1, S_2), i = 1, 2 \Big\}.$$

Therefore, probability of $E_c \cap E_1^c \cap E_2^c$ can be written as

$$P(E_c \cap E_1^c \cap E_2^c) = \sum_{(\mathbf{s}_1,\mathbf{s}_2) \in A_\epsilon^{(n)}(S_1,S_2)} P_{S_1,S_2}^n(\mathbf{s}_1, \mathbf{s}_2) \sum_{(\mathbf{z}_1,\mathbf{z}_2) \in \mathcal{E}_{\mathbf{s}_1,\mathbf{s}_2}} P(e_1(\Theta_1, \mathbf{s}_1) = \mathbf{x}_1, e_2(\Theta_2, \mathbf{s}_2) = \mathbf{x}_2),$$

where e_i is the output of the ith encoder, and Θ_i is the random message to be transmitted by encoder i, where $i = 1, 2$. To bound $P(E_c \cap E_1^c \cap E_2^c)$, we use a similar argument as in the proof of Theorem II.3. We can show that, $\mathbb{E}\{P(E_c \cap E_1^c \cap E_2^c)\} \to 0$ as $n \to \infty$.

Analysis of $E_d \cap (E_c \cup E_1 \cup E_2)^c$ Next, we use Lemma 4 to provide an upper-bound on $P(E_d \cap (E_c \cup E_1 \cup E_2)^c)$. Conditioned on $E_1^c \cap E_2^c$, the event E_d is the same as the event of interest in Lemma 4. Set $\mathcal{C}_n = \mathcal{C}_{I,1} + \mathcal{C}_{I,2} + \bar{\mathcal{D}}$, and $R = R_1 + R_2 + \rho$. It can be shown that $P(E_d \cap (E_c \cup E_1 \cup E_2)^c)$ approaches zero, if the packing bound in (2.6) holds. Since W_i is uniform over $\{0, 1\}$, then $H(W_i|Q, [W_i]_t) = 0$ for all $t > 0$. Therefore, the packing bound is simplified to

$$R_1 + R_2 + \rho \leq \log_2 p^r - H(Z_1 + Z_2|Y). \tag{2.13}$$

Note that $\rho \leq \frac{l}{n} H(V_1 + V_2|Q)$. Therefore, if the bound

$$R_1 + R_2 \leq \log_2 p^r - H(Z_1 + Z_2|Y) - \frac{l}{n} H(V_1 + V_2|Q), \tag{2.14}$$

holds on $R_1 + R_2$, then (2.13) holds too. Using (2.12), we establish a lower-bound on $\frac{l}{n} H(V_1 + V_2|Q)$. We have

$$\frac{l}{n} H(V_1 + V_2|Q) > \frac{H(V_1 + V_2|Q)}{H([V_i]_t|\bar{Q})} \left(\log_2 p^t - H([Z]_t|S_i),\right) \ 1 \leq t \leq r, \ i = 1, 2. \tag{2.15}$$

Then combining (2.14) and (2.15) gives the following:

$$R_1 + R_2 \leq \log_2 p^r - H(Z_1 + Z_2|Y) - \frac{H(V_1 + V_2|Q)}{H([V_i]_t|\bar{Q})} \left(\log_2 p^t - H([Z]_t|S_i)\right).$$

Since these bounds hold for $i = 1, 2$, and $1 \leq t \leq r$, we get the bound in the theorem.

□

Lemma 6. *The rate region given in Theorem II.4 contains the achievable rate region using group codes and linear codes. For that let $V_i, i = 1, 2$ be distributed uniformly over $\mathbb{Z}_{p^r}$. Therefore, we get the bound*

$$R_1 + R_2 \leq \min_{\substack{i=1,2 \\ 1\leq t\leq r}} \{\frac{r}{t} H([Z_i]_t|QS_i)\} - H(Z_1 + Z_2|YQ).$$

Jafar [69] used the Gel'fand-Pinsker approach for the point-to-point channel coding with states, and proposed a coding scheme using unstructured random codes. Using this scheme a single-letter and computable rate region is characterized.

Definition 19. *For a MAC $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}, P_{Y|X_1X_2})$ with states (S_1, S_2) and cost functions c_1, c_2, define $\mathscr{R}_{GP}$ as*

$$\max \Big\{ I(U_1, U_2; Y|Q) - I(U_1; S_1|Q) - I(U_2; S_2|Q) \Big\}, \tag{2.16}$$

where the maximization is taken over all joint probability distributions $P_{S_1S_2QU_1U_2X_1X_2Y}$ satisfying $\mathbb{E}\{c_i(X_i, S_i)\} \leq \tau_i$ for $i = 1, 2$, and factoring as

$$P_Q P_{S_1S_2} P_{Y|X_1X_2} \prod_{i=1,2} P_{U_iX_i|S_iQ}.$$

The collection of all such PMFs $P_{S_1S_2QU_1U_2X_1X_2Y}$ is denoted by $\mathscr{P}_{GP}$.

To the best of our knowledge, $\mathscr{R}_{GP}$ is the current largest achievable rate region using unstructured codes for the problem of MAC with states [69].

2.7.3 An Example

We present a MAC with state setup for which $\mathscr{R}_{GP}$ is strictly contained in the region characterized in Theorem II.4.

Example 4. Consider a noiseless MAC given in the following

$$Y = X_1 \oplus_4 S_1 \oplus_4 X_2 \oplus_4 S_2,$$

where X_1, X_2 are the inputs, Y is the output, and S_1, S_2 are the states. All the random variables take values from $\mathbb{Z}_4$. The states S_1 and S_2 are mutually independent, and are distributed uniformly over $\mathbb{Z}_4$. The cost function at the first encoder is defined as

$$c_1(x) \triangleq \begin{cases} 1 & \text{if } \ x \in \{1,3\} \\ 0 & \text{otherwise,} \end{cases}$$

whereas, for the second encoder the cost function is

$$c_2(x) \triangleq \begin{cases} 1 & \text{if } \ x \in \{2,3\} \\ 0 & \text{otherwise.} \end{cases}$$

We are interested in satisfying the cost constraints $\mathbb{E}\{c_1(X_1)\} = \mathbb{E}\{c_2(X_2)\} = 0$. This implies that, with probability one, $X_1 \in \{0,2\}$, and $X_2 \in \{0,1\}$.

Lemma 7. *For the setup in Example 4, an outer-bound for $\mathscr{R}_{GP}$ is the set of all rate pairs (R_1, R_2) such that $R_1 + R_2 < 1$.*

Proof. See Appendix A.7. □

Using numerical analysis, we can provide a tighter bound on the sum-rate which is $R_1 + R_2 \leq 0.32$. However, the bound in Lemma 7 is sufficient for the purpose of this paper.

Corollary 2. *For the MAC with states problem in Example 4, the rate pairs (R_1, R_2) satisfying $R_1 + R_2 = 1$ is achievable.*

Proof. The proof follows using Theorem II.4 with appropriately selected distributions $P_{V_i|Q}, P_{Z_i|QS_i}$, and $P_{X_i|QZ_iS_i}$ for $i = 1, 2$. For that, let Q be a trivial random variable

and (V_1, V_2) be IID random variables uniform distribution over $\{0, 1\}$. Conditioned on S_1, the distributions of Z_1 is given by

$$P_{Z_1|S_1}(z_1|s_1) \triangleq \begin{cases} 1/2 & \text{if } z_1 = -s_1, \text{or } z_1 = -s_1 + 2 \\ 0 & \text{otherwise,} \end{cases}$$

The distribution of Z_2 conditioned on S_2 is

$$P_{Z_2|S_2}(z_2|s_2) \triangleq \begin{cases} 1/2 & \text{if } z_2 = s_2, \text{or } z_2 = s_2 + 1 \\ 0 & \text{otherwise,} \end{cases}$$

The conditional distributions of X_i given $(S_i, Z_i), i = 1, 2$, are governed by the relation $X_i = Z_i \ominus S_i, i = 1, 2$. As a result, $X_1 \in \{0, 2\}$, and $X_2 \in \{0, 1\}$, with probability one. Hence, the cost constraints for (c_1, c_2) are satisfied. Therefore, for the defined distributions, the sum-rate given in the Theorem is simplified to $R_1 + R_2 \leq 1$. As a result the sum-rate $R_1 + R_2 = 1$ is achievable. □

CHAPTER III

Joint Source-Channel Coding in MAC

The separation principle of Shannon plays a fundamental role to reinforce the notion of modularity. This in turn allows separate development of source and channel code design. However, as shown by Shannon [1], the separation does not generalize to multi-terminal communications. For instance, this phenomenon was observed in many-to-one communications involving transmission of correlated sources over MAC [48].

In the problem of MAC with correlated sources, there are multiple transmitters, each observing a source correlated to others. The transmitters do not communicate with each other and wish to send their observations via a MAC to a central receiver. The receiver reconstructs the sources losslessly. The separate coding approach involves a source coding part and a channel coding part. In the channel coding part, Ahlswede [6] and Liao [70] studied the case where the transmitters have independent information and derived the capacity region for channel coding over MAC. In the source coding part, the distributed source coding problem was studied in which transmitters can communicate to the receiver error-free. Slepian and Wolf showed that lossless reproduction of the sources is possible with rates close to the joint entropy [36].

Due to suboptimality of the separation based strategies, the joint source-channel

coding approach has been of great interest. Cover-El Gamal-Salehi (CES) scheme introduced in [48], is a generalization of the results in [6] and [49]. Using this scheme a single-letter characterization of the set of sources that can be reliably transmitted was derived. It was shown that this scheme strictly improves upon the previously known strategies. However, Dueck [71] proved that this approach only gives a sufficient condition and not a necessary one. The joint source-channel coding problem is well studied in other settings such as: source coding with side information via a MAC [11], broadcast channels with correlated sources [72] and interference channels [73].

Recently, structured codes were used to design coding strategies for joint source-channel coding problems, [14,16,18,74]. A graph-theoretic framework was introduced in [74] to improve the joint source-channel coding schemes both in the MAC and the broadcast channel.

In this chapter, we investigate the shortcomings of coding strategies based on unstructured codes such as CES scheme. We observe that further improvements are possible when the sources impose an algebraic structure. One example is when one of the sources is the modulo sum of the other two. In this scenario, a structured coding strategy is needed for the codebooks to match with the structure of the sources. With this intuition, first we characterize existing algebraic structures in correlated sources. In particular, we define a new class of common information called "conferencing common information". Next, we propose new coding strategies that exploit such structures and contribute to improvements in terms of achievable rates.

3.1 Preliminaries and Problem Formulation

3.1.1 Notations

Calligraphic letters are used to denote sets such as $\mathcal{X}, \mathcal{Y}$. For any set $\mathcal{A}$, let $S_{\mathcal{A}} = \{S_a\}_{a \in \mathcal{A}}$. If $\mathcal{A} = \emptyset$, then $S_{\mathcal{A}} = \emptyset$. As a shorthand, we sometimes denote

a triple (s_1, s_2, s_3) by $\underline{s}$. We also denote a triple of sequences $(\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3)$ by $\underline{\mathbf{s}}$. By $\mathbb{F}_q$, we denote the field of integers modulo-q, where q is a prime number. For any mapping $\Phi : \mathcal{A} \mapsto \mathcal{B}$ and any integer n, define the mapping $\Phi^n : \mathcal{A}^n \mapsto \mathcal{B}^n$ such that $\Phi^n(a^n) \triangleq (\Phi(a_1), \Phi(a_2), ..., \Phi(a_n))$ for all $a^n \in \mathcal{A}^n$.

3.1.2 Randomized Coding Strategy

In a multi-terminal communication system a block coding scheme identifies a set of t block encoding functions $e_i : \mathcal{S}_i^k \mapsto \mathcal{X}_i^n$, where $i \in [1 : t]$, k is the input blocklength, n is the output blocklength, and $(\mathcal{S}_i^k, \mathcal{X}_i^n)$ are the input-output alphabets. Let $\mathcal{H}_k^n$ denote the set of all such encoding functions that can be used for a multi-terminal system.

To characterize performance limits (in terms of achievable rates or error exponent) of a communication system, it is necessary to show the existence of an optimality achieving coding scheme. A conventional method in information theory to show the existence of an optimal coding scheme is the so-called *random coding* technique. In this approach, the encoders are generated randomly according to a predefined probability measure on the set of all encoders. Then, it will be shown that the expectation of the performance criterion for such random encoders approaches the optimal performance limit of the communication setup. With this notion, one can identify a (random) coding strategy by the corresponding probability measure on the set of all encoders. In what follows, we formalize the definition of a randomized coding strategy.

Definition 20. *A randomized coding strategy for the set $\mathcal{H}_k^n$ of all encoders in a communications system is characterized by a probability distribution P_k^n on $\mathcal{H}_k^n$.*

There are several well-known examples of randomized coding strategies including: standard unstructured random codes for PtP communications [1], CES for MAC with correlated sources [48], random linear codes, etc.

Example 5. CES scheme with blocklength n is a randomized coding strategy with probability measure P_n^n that factors as

$$\mathsf{P}_n^n(e_1, e_2) = \prod_{(\mathbf{s}_1,\mathbf{s}_2)\in\mathcal{S}_1^n\times\mathcal{S}_2^n} \prod_{i=1}^{n} P_{X_1X_2|S_1S_2}(e_{1,i}(\mathbf{s}_1), e_{2,i}(\mathbf{s}_2)|s_{1,i}, s_{2,i})$$

for all encoding functions $e_j : \mathcal{S}_j^n \mapsto \mathcal{X}_j^n, j = 1, 2$, where $e_{1,i}$ and $e_{2,i}$ is the ith output of e_1 and e_2, respectively. Also, the conditional distribution $P_{X_1X_2|S_1S_2}$ is the marginal of a distribution of the form $P_{UX_1X_2|S_1S_2} = P_U P_{X_1|US_1} P_{X_2|US_2}$.

Example 6. (identical) random linear codes over $\mathbb{F}_q$ with t-encoders are randomized coding strategies where P_k^n is the uniform probability measure on the set of all encoding functions $e_i, i \in [1, t]$ of the form $e_i(\mathbf{a}) = \phi(\mathbf{a}) + \mathbf{b}_i, \forall \mathbf{a} \in \mathbb{F}_q^k$, where $\phi : \mathbb{F}_q^k \mapsto \mathbb{F}_q^n$ is a linear transformation and $\mathbf{b}_i \in \mathbb{F}_q^n$ are vectors satisfying $\mathbf{b}_1 \oplus \mathbf{b}_2 \oplus ... \oplus \mathbf{b}_t = 0$

Some of the known coding strategies are considered to be "structured" such as linear code; whereas some are considered to be "unstructured" such as unstructured random codes. In what follows, we aim to develop a measure to identify when a coding strategy is "unstructured".

By $(E_1, E_2, ..., E_t)$ denote random encoders of a randomized coding strategy P_k^n. For each random encoder E_j, let $E_{j,i}$ denote the ith output of the encoder, where $i \in [1, n]$.

Definition 21. *A randomized coding strategy P_k^n is said to be δ- unstructured for input distribution P_{S^k}, if $\delta \geq 0$ is the largest number for which the following inequalities hold for any non-constant mapping $\Phi : \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_t \mapsto \{0, 1\}$*

$$\mathbb{P}\left\{\Phi^n\Big(E_1(S_1^k), E_2(S_2^k)), ..., E_t(S_t^k)\Big) = \mathbf{0}^n\right\} \leq 1 - \delta, \quad \forall i \in [1, n]. \tag{3.1}$$

A well-known example of a δ-unstructured coding strategy is CES scheme.

Lemma 8. *The CES scheme for which* $P_{X_1,X_2|S_1,S_2}(x_1,x_2|s_1,s_2) > \epsilon, \forall x_i \in \mathcal{X}_i, \forall s_i \in \mathcal{S}_i, i = 1,2$ *is* δ*-unstructured for any source distribution, where* $\delta \geq 1-(1-\epsilon)^n$.

Proof. We need to find δ that satisfies (3.1). Let (E_1, E_2) represent random encoders in CES scheme. Given any non-constant mapping Φ we have:

$$\mathbb{P}\Big\{\Phi^n\Big(E_1(S_1^n), E_2(S_2^n)\Big) = 0\Big\} = \sum_{s_1^n, s_2^n} P^n_{S_1,S_2}(s_1^n, s_2^n)\mathbb{P}\Big\{\Phi^n\Big(E_1(s_1^n), E_2(s_2^n)\Big) = 0\Big|s_1^n, s_2^n\Big\}$$

Note that conditioned on (s_1^n, s_2^n) the outputs of the encoders (X_1^n, X_2^n) are IID. Hence, the probability in the right-hand side term above equals

$$\prod_{i=1}^n \mathbb{P}\big\{\Phi(X_{1,i}, X_{2,i}) = 0\big|s_{1,i}, s_{2,i}\big\}$$

Let Null(Φ) represent the null set of Φ. Then, the above term equals

$$\begin{aligned}\prod_{i=1}^n \mathbb{P}\big\{(X_{1,i}, X_{2,i}) \in \text{Null}(\Phi)\big|s_{1,i}, s_{2,i}\big\} &\leq \prod_{i=1}^n \sup_{\mathcal{A}\subsetneq\mathcal{X}_1\times\mathcal{X}_2} P_{X_1,X_2|S_1,S_2}(\mathcal{A}|s_{1,i}, s_{2,i}) \\ &< \prod_{i=1}^n (1-\epsilon) = (1-\epsilon)^n,\end{aligned}$$

where the last inequality follows as $P_{X_1,X_2|S_1,S_2}(x_1,x_2|s_1,s_2) > \epsilon$. As a result for any mapping Φ we obtain

$$\mathbb{P}\Big\{\Phi^n\Big(E_1(S_1^n), E_2(S_2^n)\Big) = 0\Big\} < (1-\epsilon)^n,$$

which implies that $\delta \geq 1-(1-\epsilon)^n$. □

According to Definition 21, when $S_1 \oplus S_2 \oplus ... \oplus S_t = 0$ with probability one, coding strategies such as random linear/group codes with identical generating matrix are not δ-unstructured for any $\delta \geq 0$.

3.1.3 Conferencing Common Information

Joint source-channel coding techniques exploit the existing statistical correlations in the sources. This has been done in [48] through the notion of "common information". A well-known definition of common information is due to Gács-Körner-Witsenhausen (GKW) [50], [51]. In subsection, we define a new class of common information called conferencing common information. Let us begin with the definition of GKW common information.

Definition 22 (GKW Common part)**.** *A common part between random variables (X, Y) is a random variable W for which there exist functions f, g such that $W = f(X)$, and $W = g(Y)$ with probability one. In this work, such a random variable W is sometimes called a uni-variate common part.*

Definition 23 (GKW Common Information)**.** *The common information between random variables (X, Y) is defined as the maximum entropy of W where W is a common part between (X, Y).*

One can generalize the above definitions for more than two random variables. With this notion, we can define a common part between random variables $(S_1, S_2, \ldots, S_k)$ as a random variable W for which there exist functions $f_i, i \in [1 : k]$ such that $W = f_i(S_i)$ holds with probability one.

Definition 24. *The conferencing common part among three random variables (S_1, S_2, S_3) is a triplet (T_1, T_2, T_3) for which there exist functions $f_i, g_i, i \in \{1, 2, 3\}$ such that the inequalities $T_i = f_i(X_i) = g_i(X_j, X_k)$ hold with probability one for all distinct $i, j, k \in \{1, 2, 3\}$.*

As a result of Definition 22 and 24, the common parts defined among three random variables (S_1, S_2, S_3) are $(W_{12}, W_{13}, W_{23}, W_{123}, T_1, T_2, T_3)$, where W_{ij} is the pairwise common part between (S_i, S_j), W_{123} is the mutual common part (all in the sense of

Definition 22), and (T_1, T_2, T_3) are conferencing common parts (as in Definition 24) among (S_1, S_2, S_3).

In this work, we focus on a special class of conferencing common part which is defined as follows.

Definition 25. *Given m and random variables (S_1, S_2, S_3), an m-additive common part is defined as a triple (T_1, T_2, T_3) each taking values in $\mathbb{Z}_m$ such that $T_i = f_i(S_i), i = 1, 2, 3$, and $T_1 \oplus_m T_2 \oplus_m T_3 = 0$ hold with probability one.*

The following example provides a triplet of binary sources with a 2-additive common part.

Example 7. Let S_1, S_2 and S_3 be three Bernoulli random variables. Suppose S_1 and S_2 are independent and $S_3 = S_1 \oplus_2 S_2$ with probability one. It is not difficult to show that uni-variate common parts are trivial, i.e., $(W_{12}, W_{13}, W_{23}, W_{123})$ are constant. As for the conferencing common parts, set $T_i = S_i, i = 1, 2, 3$. Then (T_1, T_2, T_3) satisfies the conditions in Definition 25 for $m = 2$. Therefore, (T_1, T_2, T_3) is a 2-additive common part of (S_1, S_2, S_3).

3.1.4 Problem Formulation

As depicted in Figure 3.1, the problem of MAC with correlated sources consists of multiple transmitters, each observing a source sequence statistically correlated to others. The source sequences are sent by the transmitters via a MAC to a central receiver. The objective of the receiver is to reconstruct the source sequences losslessly. It is assumed that the channel is a discrete memoryless MAC and the source sequences are discrete and generated IID according to a known joint PMF. In what follows, we formulate this problem more precisely.

Definition 26. *A discrete memoryless MAC with t users is defined by input alphabets $\mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_t$, output alphabet $\mathcal{Y}$, and a transition probability matrix $P_{Y|X_1,X_2,\cdots,X_t}$.*

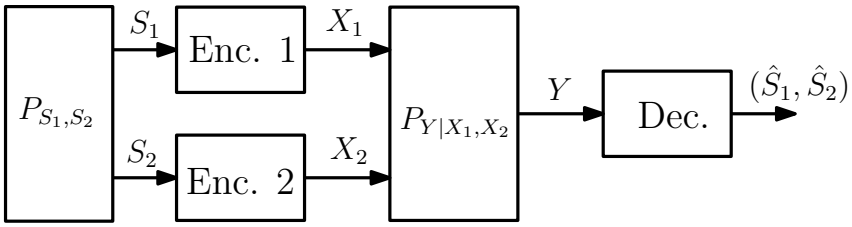

Figure 3.1: The diagram of a two-user MAC with correlated sources. In this Setup, the source sequences (S_1^n, S_2^n) are observed by the corresponding encoders. The encoders produce (X_1^n, X_2^n) which are channel's input sequences. Upon observing the channel output Y^n, the decoder produces an estimate for the sources. The design objective is to provide a lossless estimate of the source sequences at the receiving end of the channel.

The input and output alphabets are assumed to be finite sets. This setup is denoted by $P_{Y|X_1,X_2,\cdots,X_t}$.

Definition 27. *The input sources for a* t*-user MAC are defined as* t *sequences of random variables* $(S_1^n, S_2^n, \cdots, S_t^n)$ *generated IID according to a joint distribution* $P_{S_1,S_2,\cdots,S_t}$. *Such input sources are denoted by the underlying random variables* $(S_1, S_2, \cdots, S_t)$.

In this paper, no bandwidth expansion is considered for transmission of the sources. In other words, the input and output sequences at each transmitter have identical blocklength.

Definition 28. *A coding scheme (without bandwidth expansion) for transmission of the sources* $(S_1, S_2, \ldots, S_t)$ *over a MAC consists of encoding functions* $e_i : \mathcal{S}_i^n \to \mathcal{X}_i^n, i \in [1:t]$, *and a decoding function* $d : \mathcal{Y}^n \to \mathcal{S}_1^n \times \mathcal{S}_2^n \times \cdots \times \mathcal{S}_t^n$. *The parameter* n *is called blocklength.*

Definition 29. *Given a MAC* $P_{Y|X_1,X_2,\cdots,X_t}$ *and for an* $\epsilon > 0$, *a source* $(S_1, S_2, \ldots, S_t)$ *is said to be* ϵ*-transmissible using a coding scheme with encoders* $e_i, i \in [1:t]$ *and a decoder* g, *if*

$$\mathbb{P}\Big\{d(Y^n) \neq (S_1^n, S_2^n, \ldots S_t^n) | X_i^n = e_i(S_i^n), i \in [1:t]\Big\} \leq \epsilon,$$

where n is the blocklength of the coding scheme.

For a given MAC with correlated sources setup, let $\mathcal{H}_n$ denote the set of all encoders that map the source sequences to the channel's input sequences. Based on Definition 20, a randomized coding strategy is identified by a probability measure on $\mathcal{H}_n$. [1]

Definition 30. *Given a MAC $P_{Y|X_1,X_2,\cdots,X_t}$ and for an $\epsilon > 0$, a source $(S_1, S_2, \ldots, S_t)$ is said to be ϵ-transmissible using a randomized coding strategy P_n, if*

$$\mathbb{E}_{\mathsf{P}_n}\left[\mathbb{P}\Big\{d(Y^n) \neq (S_1^n, S_2^n, \ldots S_t^n) | X_i^n = e_i(S_i^n), i \in [1:t]\Big\}\right] \leq \epsilon.$$

Definition 31. *A source $(S_1, S_2, \ldots, S_t)$ is said to be reliably transmissible over a MAC $P_{Y|X_1,X_2,\cdots,X_t}$, if for any $\epsilon > 0$ there exists a randomized coding strategy (or equivalently a coding scheme) using which it is ϵ-transmissible.*

3.2 Applications of Common Information in MAC with Correlated Sources

In this section, we investigate techniques to exploit common information in the the problem of MAC with correlated sources. We show how algebraic structures in the statistic of the sources can be exploited using structured codes.

3.2.1 Encoding of Uni-Variate Common Information

The two-user version of MAC with correlated sources is investigated in [48] and CES scheme is introduced. It is observed that common information can be transmitted more efficiently and treating it separately may lead to achieving higher rates. This coding strategy is explained here.

[1] Note that here $k = n$ because no bandwidth expansion is considered.

Let W be a uni-variate common part between the input sources (S_1, S_2) as in Definition 22. In CES scheme, as shown in Figure 3.2, first the common part W is calculated at each encoder. The common part is available at both transmitters. Therefore, it is transmitted using identical encoders (as it is done in PtP communications). Next, at each transmitter, each source is encoded using a codebook that is "super-imposed" on the common codebook.

It is shown in [48] that using CES scheme reliable transmission of (S_1, S_2) is possible if the following conditions are satisfied,

$$\begin{aligned} H(S_1|S_2) &\leq I(X_1; Y|X_2, S_2, U), \\ H(S_2|S_1) &\leq I(X_2; Y|X_1, S_1, U), \\ H(S_1, S_2|W) &\leq I(X_1X_2; Y|W, U), \\ H(S_1, S_2) &\leq I(X_1X_2; Y), \end{aligned}$$

where W is the common part between (S_1, S_2), and the joint distribution of all the random variables factors as

$$P_{S_1,S_2,U,X_1,X_2,Y} = P_{S_1,S_2} P_U P_{X_1|S_1,U} P_{X_2|S_2,U} P_{Y|X_1,X_2}.$$

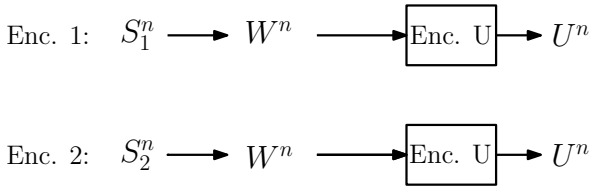

Figure 3.2: In CES scheme uni-variate common parts are encoded using identical encoders. Random variable U^n represents the encoded version of the common part at each transmitter.

3.2.2 Encoding of Conferencing Common Information

Unlike uni-variate common information, conferencing common parts are not available at any Transmitter. This is due to the fact that conferencing common parts are bi-variate functions of the sources. As a result, to exploit conferencing common information, different coding techniques need to to be developed. The focus of this work is on q-additive common information, where q is a prime number. Such class of common information can be exploited using linear (or affine) maps.

For a fixed prime number q, suppose (T_1, T_2, T_3) are non-trivial q-additive common parts of given sources (S_1, S_2, S_3). We construct three affine maps for encoding of such common parts. Let $\mathbf{G}$ be a n by n matrix with elements in $\mathbb{F}_q$. Also, select vectors $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3 \in \mathbb{F}_q^n$ such that $\mathbf{b}_1 \oplus \mathbf{b}_2 \oplus \mathbf{b}_3 = \mathbf{0}$. Define the encoded version of the common parts as $V_i^n = T_i^n \mathbf{G} \oplus \mathbf{b}_i$, where $i = 1, 2, 3$. Since in this approach an affine map is used to encode each q-additive common part, the equality $V_1^n \oplus V_2^n \oplus V_3^n = \mathbf{0}$ holds with probability one. One may adopt a randomized affine map to encode the q-additive common parts. For that, we can select the matrix $\mathbf{G}$ and the vectors $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$ randomly and uniformly from the set of all matrices and vectors with elements in $\mathbb{F}_q$.

In what follows, we show that applications of affine maps for transmission of q-additive common information improves upon CES scheme.

Example 8. Suppose (S_1, S_2, S_3) are as in Example 7. The sources are to be transmitted via a MAC with binary inputs $\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3$, binary outputs $\mathcal{Y}_1 \times \mathcal{Y}_2$, and a conditional probability distribution that satisfies

$$(Y_1, Y_2) = \begin{cases} (X_1 \oplus N_\delta, X_1 \oplus N'_\delta, & \text{if } X_3 = X_1 \oplus X_2, \\ (N_{1/2}, N'_{1/2}), & \text{if } X_3 \neq X_1 \oplus X_2, \end{cases} \tag{3.2}$$

where $N_\delta, N'_\delta, N_{1/2}$ and $N'_{1/2}$ are independent Bernoulli random variables with parameter $\delta, \delta, \frac{1}{2}$, and $\frac{1}{2}$, respectively.

As explained in Example 7, the uni-variate common parts are trivial, and the 2-additive common parts are $T_i = S_i, i = 1, 2, 3$. For such setup, we use random affine maps explained above. For that set $X_i^n = S_i^n \mathbf{G} \oplus \mathbf{B}_i, i = 1, 2, 3$, where $\mathbf{G}, \mathbf{B}_1, \mathbf{B}_2, \mathbf{B}_3$ are selected randomly, and satisfying $\mathbf{B}_1 \oplus \mathbf{B}_2 \oplus \mathbf{B}_3 = \mathbf{0}$. The following lemma provides a necessary and sufficient condition for reliable transmission of (S_1, S_2, S_3). The achievability is obtained using the above approach.

Lemma 9. *For the setup defined in Example 8, the sources are reliably transmissible, if and only if $H(S_i) \leq 1 - h_b(\delta)$, where $i = 1, 2$. Furthermore, such sources are ϵ-transmissible for any $\epsilon > 0$ and using identical random linear codes.*

Proof. The proof for the direct part follows using random affine maps and from the standard arguments. For the converse part, suppose (S_1, S_2, S_3) are ϵ-transmissible using a coding scheme with (e_1, e_2, e_3) as the encoders and g as the decoder. From Fano's inequality

$$\begin{aligned}
\frac{1}{n} H(S_1^n, S_2^n) &\leq \frac{1}{n} I(S_1^n, S_2^n; Y_1^n, Y_2^n) + 2\epsilon + \frac{1}{n} h_b(\epsilon) \\
&\overset{(a)}{=} \frac{1}{n} I(X_1^n, X_2^n, X_3^n; Y_1^n, Y_2^n) + 2\epsilon + \frac{1}{n} h_b(\epsilon) \\
&\overset{(b)}{\leq} 2 - h_b(\delta) + 2\epsilon + \frac{1}{n} h_b(\epsilon),
\end{aligned}$$

where (a) follows because of the Mrkov chain $(S_1, S_2, S_3) \leftrightarrow (X_1, X_2, X_3) \leftrightarrow (Y_1, Y_2)$. Inequality (b) holds as the mutual information does not exceed the sum-capacity of the MAC which equals to $2 - h_b(\delta)$. The proof is complete as the inequalities hold for arbitrary $\epsilon > 0$. □

3.3 Three-User MAC with Correlated Sources

In this section, we investigate coding strategies in three-user MAC with correlated sources. We first present an extension of CES scheme for such problem and then

propose a new coding strategy. A new sufficient condition is characterized which improves upon the one derived using CES scheme.

3.3.1 A Three-User Extension of CES Scheme

For the case of multiple sources, say (S_1, S_2, S_3), a similar idea as in CES can be used to encode the uni-variate common parts. In what follows we provide a natural extension of CES scheme to three-use MAC with correlated sources.

The uni-variate common parts among the sources (S_1, S_2, S_3) are defined as follows. There are four components denoted by $(W_{12}, W_{13}, W_{23}, W_{123})$. For more convenience, we denote the pairwise common parts either by W_{ij} or W_{ji} (we simply drop the condition $j > i$, as it is understood that $W_{ij} = W_{ji}$).

The first step in CES scheme is to capture the common parts among the sources. By observing S_1 at the first transmitter, three common parts can be calculated: W_{123}, W_{12}, and W_{13}. Similarly, at the ith transmitter W_{123}, W_{ij}, and W_{ik} are calculated, where i, j, k are distinct elements of $\{1, 2, 3\}$. The three-user extension of CES involves three layers of coding. In the first layer W_{123} is encoded at each encoder. Next, based on the output of the first layer, the W_{ij}'s are encoded. Finally, based on the output of the first and the second layers, S_1, S_2 and S_3 are encoded. Figure 3.3 shows the random variables involved in the extension of CES.

$$\begin{aligned}
&\text{Enc.1} \qquad S_1^n \rightarrow W_{123}^n\ W_{12}^n\ W_{13}^n \longrightarrow U_{123}^n\, U_{12}^n\, U_{13}^n \\
&\text{Enc.2} \qquad S_2^n \rightarrow W_{123}^n\ W_{12}^n\ W_{23}^n \longrightarrow U_{123}^n\, U_{12}^n\, U_{23}^n \\
&\text{Enc.3} \qquad S_3^n \rightarrow W_{123}^n\ W_{13}^n\ W_{23}^n \longrightarrow U_{123}^n\, U_{13}^n\, U_{23}^n
\end{aligned}$$

Figure 3.3: The random variables involved in the three-user extension of CES.

As a result, the extension of CES scheme is a randomized coding strategy with the following probability measure.

Remark 10. The three-user extension of CES scheme is a coding strategy with a

probability measure P_n^n that factors as

$$\mathsf{P}_n^n(e_1, e_2, e_3) = \prod_{\underline{\mathbf{s}} \in \underline{\mathcal{S}}^n} \prod_{i=1}^{n} P_{\underline{X}|\underline{S}}\Big(e_{1,i}(\mathbf{s}_1), e_{2,i}(\mathbf{s}_2), e_{3,i}(\mathbf{s}_3) \;\Big|\; s_{1,i}, s_{2,i}, s_{3,i}\Big),$$

where $P_{\underline{X}|\underline{S}}$ is the conditional and marginal distribution obtained from the joint PMF in (3.3).

The following proposition determines sufficient conditions for which correlated sources can be transmitted using the above scheme.

Proposition 1. *The reliable transmission of the sources (S_1, S_2, S_3) over a three-user MAC is possible if for any distinct $i, j, k \in \{1, 2, 3\}$ and any $\mathcal{B} \subseteq \{12, 13, 23\}$ the following inequalities hold*

$$\begin{aligned}
H(S_i|S_jS_k) &\leq I(X_i; Y|S_jS_kX_jX_kU_{123}U_{12}U_{13}U_{23}),\\
H(S_iS_j|S_kW_{\mathcal{B}}) &\leq I(X_iX_j; Y|S_kW_{\mathcal{B}}U_{123}U_{ik}U_{jk}U_{\mathcal{B}}X_k),\\
H(S_1S_2S_3|W_{123}W_{\mathcal{B}}) &\leq I(X_1X_2X_3; Y|W_{123}W_{\mathcal{B}}U_{123}U_{\mathcal{B}}),\\
H(S_1S_2S_3) &\leq I(X_1X_2X_3; Y),
\end{aligned}$$

where $U_{ij} = U_{ji}$ and the joint distribution of $(\underline{S}, \underline{X}, U_{123}, U_{12}, U_{13}, U_{23})$ factors as

$$P_{S_1,S_2,S_3}P_{U_{123}}\left[\prod_{b\in\{12,13,23\}} P_{U_b|W_bU_{123}}\right] P_{X_1|S_1U_{123}U_{12}U_{13}}P_{X_2|S_2U_{123}U_{12}U_{23}}P_{X_3|S_3U_{123}U_{13}U_{23}}. \tag{3.3}$$

Outline of the proof. Suppose $(S_1, S_2, S_3, X_1, X_2, X_3, U_{123}, U_{12}, U_{13}, U_{23})$ is distributed according to the joint PMF given in (3.3). Let the sequence $\mathbf{s}_i \in \mathcal{S}_i^n$ be a realization of the ith source, where $i = 1, 2, 3$.

Codebook Generation At each Transmitter three different codebooks are defined, one for the mutual common part, one for the pairwise common parts, and one for the input source. The construction of these codebooks is given below:

1. For each realization $\mathbf{w}_{123}$ of the mutual common part, a sequence $\mathbf{u}_{123}$ is generated randomly according to the marginal PMF $\prod_{l\in[1,n]} P_{U_{123}}$. Such sequence is indexed by $u_{123}(\mathbf{w}_{123})$.

2. Given $b \in \{12, 13, 23\}$ and for each $\mathbf{u}_{123}$ and $\mathbf{w}_b$ a sequence $\mathbf{u}_b$ is generated randomly according to the conditional PMF $\prod_{l\in[1,n]} P_{U_b|W_bU_{123}}$. Such sequence is indexed by $u_b(\mathbf{w}_b, \mathbf{u}_{123})$.

3. Given distinct elements $i, j, k \in \{1, 2, 3\}$, any realization $\mathbf{s}_i$ of the source, the common parts $(\mathbf{w}_{123}, \mathbf{w}_{ij}, \mathbf{w}_{ik})$, and the corresponding sequences $u_{123}(\mathbf{w}_{123})$, $u_{ij}(\mathbf{w}_{ij}, \mathbf{u}_{123})$ and $u_{ik}(\mathbf{w}_{ik}, \mathbf{u}_{123})$ generate a random IID sequence $\mathbf{x}_i$ according to $\prod_{l\in[1,n]} P_{X_i|S_iU_{123}U_{ij}U_{ik}}$. For shorthand, such sequence is denoted by $x_i(\mathbf{s}_i, \mathbf{u}_{123}, \mathbf{u}_{ij}, \mathbf{u}_{ik})$.

Encoding Upon observing a realization $\mathbf{s}_i$ of the ith source, Transmitter i first calculates the common part sequences $(\mathbf{w}_{123}, \mathbf{w}_{ij}, \mathbf{w}_{ik})$, where $i, j, k \in \{1, 2, 3\}$ are distinct. Then, the transmitter finds the corresponding sequences

$$(u_{123}(\mathbf{w}_{123}), u_{ij}(\mathbf{w}_{ij}, \mathbf{u}_{123}), u_{ik}(\mathbf{w}_{ik}, \mathbf{u}_{123}))$$

and sends $x_i(\mathbf{s}_i, \mathbf{u}_{123}, \mathbf{u}_{ij}, \mathbf{u}_{ik})$ to the channel.

Decoding Upon receiving $\mathbf{y}$ from the channel, the decoder finds source realizations $(\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2, \tilde{\mathbf{s}}_3)$ such that

$$(\underline{\tilde{\mathbf{s}}}, \tilde{\mathbf{u}}_{123}, \tilde{\mathbf{u}}_{12}, \tilde{\mathbf{u}}_{13}, \tilde{\mathbf{u}}_{23}, \tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \tilde{\mathbf{x}}_3, \mathbf{y}) \in A_\epsilon^{(n)}(\underline{S}, U_{123}, U_{12}, U_{13}, U_{23}, X_1, X_2, X_3, Y),$$

where $\tilde{\mathbf{u}}_{123} = u_{123}(\tilde{\mathbf{w}}_{123}), \tilde{\mathbf{u}}_{ij} = u_{ij}(\tilde{\mathbf{w}}_{ij}, \tilde{\mathbf{u}}_{123})$, $\mathbf{x}_i = x_i(\tilde{\mathbf{s}}_i, \tilde{\mathbf{u}}_{123}, \tilde{\mathbf{u}}_{ij}, \tilde{\mathbf{u}}_{ik})$, and $i, j, k \in \{1, 2, 3\}$ are distinct. Note that $(\tilde{\mathbf{w}}_{123}, \tilde{\mathbf{w}}_{12}, \tilde{\mathbf{w}}_{13}, \tilde{\mathbf{w}}_{23})$ are the corresponding common part sequences of $(\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2, \tilde{\mathbf{s}}_3)$.

A decoding error will be occurred, if no unique $(\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2, \tilde{\mathbf{s}}_3)$ is found. Using a standard argument as in [48], it can be shown that the error probability is sufficiently small for large enough n, if the conditions in Preposition 1 are satisfied.

□

3.3.2 New Sufficient Condition

We use the intuition behind the argument above and propose a new coding strategy in which a combination of random linear codes (as in Example 6) and the extension CES scheme (as in Definition 10) is used. The coding scheme uses both uni-variate and q-additive common information among the sources. In the next Theorem, we derive sufficient conditions for transmission of correlated sources over three-user MAC.

Theorem III.1. *A source (S_1, S_2, S_3) is reliably transmissible over a MAC $P_{Y|X_1,X_2,X_3}$, if for any distinct $i, j, k \in \{1, 2, 3\}$ and for any $\mathcal{A} \subseteq \{1, 2, 3\}, \mathcal{B} \subseteq \{12, 13, 23\}$ the followings hold:*

$$H(S_i|S_jS_k) \leq I(X_i; Y|S_jS_kU_{123}U_{12}U_{13}U_{23}V_1V_2V_3X_jX_k) \tag{3.4}$$

$$H(S_iS_j|S_kW_{\mathcal{B}}T_{\mathcal{A}}) \leq I(X_iX_j; Y|S_kW_{\mathcal{B}}U_{123}U_{ik}U_{jk}U_{\mathcal{B}}T_{\mathcal{A}}V_kV_{\mathcal{A}}X_k) \tag{3.5}$$

$$H(S_iS_jS_k|W_{123}W_{\mathcal{B}}T_{\mathcal{A}}) \leq I(X_iX_jX_k; Y|W_{123}W_{\mathcal{B}}U_{123}U_{\mathcal{B}}T_{\mathcal{A}}V_{\mathcal{A}}) \tag{3.6}$$

$$H(S_iS_jS_k|T_{\mathcal{A}}) \leq I(X_iX_jX_k; Y|T_{\mathcal{A}}V_{\mathcal{A}}) \tag{3.7}$$

where the joint distribution of all the random variables factors as

$$P_{S_1,S_2,S_3}P_{U_{123}}\left[\prod_{b\in\{12,13,23\}} P_{U_b|W_bU_{123}}\right] P_{V_1V_2V_3}\left[\prod_{\substack{i,j,k\in\{1,2,3\}\\ j<k}} P_{X_i|S_iU_{123}U_{ij}U_{ik}V_i}\right], \tag{3.8}$$

where the random variables $(W_{123}, W_{12}, W_{13}, W_{23})$ *are the uni-variate common parts,* (T_1, T_2, T_3) *are* q*-additive common parts for a prime* q*, and* $P_{V_1V_2V_3} = \frac{1}{q^2}\mathbb{1}\{V_3 = V_1 \oplus_q V_2\}$.

Remark 11. The set of sufficient conditions given in Theorem III.1 includes the one in Proposition 1. For that select the joint distribution in (3.8) such that X_i be independent of V_i for all $i = 1, 2, 3$.

Outline of the proof. We use affine maps to encode q-additive common parts and build upon the coding scheme described in the proof of Proposition 1. Suppose the random variables $(\underline{S}, \underline{X}, U_{123}, U_{12}, U_{13}, U_{23}, \underline{V})$ are distributed according to a joint distribution that factors as in (3.8).

Codebook Generation At each transmitter five different codebooks are defined, one codebook for the q-additive common part T_i, three codebooks for uni-variate common parts $(W_{123}, W_{ij}, W_{ik})$, where i, j, k are distinct elements of $\{1, 2, 3\}$, and one codebook for generating the total output X_i^n.

1. The codebooks for encoding of uni-variate common parts are as in the proof of Proposition 1.

2. The codebook for encoding of (T_1, T_2, T_3) is defined using affine maps. Generate two vectors $\mathbf{b}_1, \mathbf{b}_2$ of length n, and an $n \times n$ matrix $\mathbf{G}$ with elements selected randomly, uniformly and independently from $\mathbb{F}_q$. Set $\mathbf{b}_3 = \mathbf{b}_1 \oplus \mathbf{b}_2$. For each sequence $\mathbf{t}_i \in \mathbb{F}_q^n$, define $v_i(\mathbf{t}_i) = \mathbf{t}_i\mathbf{G} \oplus \mathbf{b}_i$, where $i = 1, 2, 3$, and all the additions and multiplications are modulo-q.

3. Given distinct $i, j, k \in \{1, 2, 3\}$, any realization $\mathbf{s}_i$ of the source, the common parts $(\mathbf{w}_{123}, \mathbf{w}_{ij}, \mathbf{w}_{ik}, \mathbf{t}_i)$, and the corresponding sequences

$$\big(u_{123}(\mathbf{w}_{123}), u_{ij}(\mathbf{w}_{ij}, \mathbf{u}_{123}), u_{ik}(\mathbf{w}_{ik}, \mathbf{u}_{123}), v_i(\mathbf{t_i})\big)$$

generate a random IID sequence $\mathbf{x}_i$ according to $\prod_{l\in[1,n]} P_{X_i|S_iU_{123}U_{ij}U_{ik}V_i}$. For shorthand, such sequence is denoted by $x_i(\mathbf{s}_i, \mathbf{u}_{123}, \mathbf{u}_{ij}, \mathbf{u}_{ik}, \mathbf{v}_i)$.

Encoding Assume $\mathbf{s}_i$ is a realization of the ith source, where $i = 1, 2, 3$. Transmitter i first calculates the common part sequences $(\mathbf{w}_{123}, \mathbf{w}_{ij}, \mathbf{w}_{ik}, \mathbf{t}_i)$, where $i, j, k \in \{1, 2, 3\}$ are distinct. Next, the transmitter finds the corresponding sequences

$$\big(u_{123}(\mathbf{w}_{123}), u_{ij}(\mathbf{w}_{ij}, \mathbf{u}_{123}), u_{ik}(\mathbf{w}_{ik}, \mathbf{u}_{123}), v_i(\mathbf{t_i})\big)$$

and sends $x_i(\mathbf{s}_i, \mathbf{u}_{123}, \mathbf{u}_{ij}, \mathbf{u}_{ik}, \mathbf{v}_i)$ to the channel.

Decoding Upon receiving $\mathbf{y}$ from the channel, the decoder finds sequences $\tilde{\mathbf{s}}_i \in \mathcal{S}_i^n, i = 1, 2, 3$, such that

$$(\underline{\tilde{\mathbf{s}}}, \tilde{\mathbf{u}}_{123}, \tilde{\mathbf{u}}_{12}, \tilde{\mathbf{u}}_{13}, \tilde{\mathbf{u}}_{23}, \underline{\tilde{\mathbf{v}}}, \underline{\tilde{\mathbf{x}}}, \mathbf{y}) \in A_\epsilon^{(n)}(\underline{S}, U_{123}, U_{12}, U_{13}, U_{23}, \underline{V}, \underline{X}, Y),$$

where $\tilde{\mathbf{u}}_{123} = u_{123}(\tilde{\mathbf{w}}_{123}), \tilde{\mathbf{u}}_{ij} = u_{ij}(\tilde{\mathbf{w}}_{ij}, \tilde{\mathbf{u}}_{123}), , \tilde{\mathbf{v}}_i = v_i(\tilde{\mathbf{t}}_i)$, $\tilde{\mathbf{x}}_i = x_i(\tilde{\mathbf{s}}_i, \tilde{\mathbf{u}}_{123}, \tilde{\mathbf{u}}_{ij}, \tilde{\mathbf{u}}_{ik}, \tilde{\mathbf{t}}_i)$, and $i, j, k \in \{1, 2, 3\}$ are distinct. Note that $(\tilde{\mathbf{w}}_{123}, \tilde{\mathbf{w}}_{12}, \tilde{\mathbf{w}}_{13}, \tilde{\mathbf{w}}_{23})$ and $(\tilde{\mathbf{t}}_1, \tilde{\mathbf{t}}_2, \tilde{\mathbf{t}}_3)$ are the uni-variate and q-additive common part sequences of $(\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2, \tilde{\mathbf{s}}_3)$, respectively.

A decoding error will be occurred, if no unique $(\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2, \tilde{\mathbf{s}}_3)$ is found. It is shown in Appendix B.1 that the probability of error approaches zero as $n \to \infty$, if (3.4)-(3.7) are satisfied. □

Remark 12. The coding strategy explained in the proof of Theorem III.1 subsumes the extension of CES scheme and identical random linear coding strategy.

3.3.3 Suboptimnality of CES Scheme

It is noted in Remark 11 that the sufficient conditions in Theorem III.1 includes the one derived using CES in Proposition 1. In this section we show that this inclusion

is strict and that CES scheme is suboptimal when applied to three-user MAC with correlated sources. The argument starts by introducing an example which is given below.

Example 9. Consider the sources denoted by (S_1, S_2, S_3), where S_1 and S_3 are independent independent Bernoulli random variables with parameter σ and γ, respectively. Suppose the third source satisfies $S_3 = S_1 \oplus_2 S_2$ with probability one. For shorthand we associate such sources with the parameters (σ, γ). The sources are to be transmitted trough a MAC with binary inputs as shown in Figure 3.4. In this channel the noise random variable N is assumed to be independent of other random variables. The PMF of N is given in Table 3.1, where the parameter $0 \leq \delta \leq \frac{1}{2}, \delta \neq \frac{1}{4}$.

Table 3.1: Distribution of N

N	0	1	2	3
P_N	$\frac{1}{2} - \delta$	$\frac{1}{2}$	δ	0

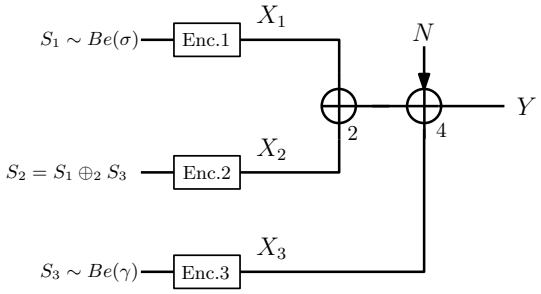

Figure 3.4: The diagram the setup introduced in Example 9. Note the input alphabets of this MAC are restricted to $\{0, 1\}$.

For this setup, we show that there exist parameters (σ, γ) whose corresponding sources in Example 9 cannot be transmitted reliably using the CES scheme. However, according on Theorem III.1, such sources can be reliably transmitted. This emphasizes the fact that efficient encoding of conferencing common information contributes to improvements upon coding schemes solely based on uni-variate common

information. In what follows, we explain the steps to show the existence of such parameters.

Remark 13. For a especial case in which $\sigma = 0$, the qualities $S_1 = 0$ and $S_2 = S_3$ hold with probability one. From Proposition 1, such (S_1, S_2, S_3) can be transmitted using CES scheme, if $h_b(\gamma) \leq 2 - H(N)$ holds.

Let $\gamma^* \in [0, \frac{1}{2}]$ be such that $\gamma^* = h_b^{-1}(2-H(N))$. Such γ^*, exists as $0 \leq 2-H(N) \leq 1$. By Remark 13, the sources (S_1, S_2, S_3) with parameter $(\sigma = 0, \gamma = \gamma^*)$ can be transmitted reliably using CES scheme. However, we argue that for small enough $\epsilon > 0$, the sources with parameter $(\sigma = \epsilon, \gamma = \gamma^* - \epsilon)$ cannot be transmitted using this scheme (Lemma 11). Whereas, from Theorem III.1, this source can be transmitted reliably (Lemma 12). The existence of such ϵ is investigated in the next two lemmas.

Lemma 10. *For the MAC in Example 9, $I(X_1, X_2, X_3; Y) \leq 2 - H(N)$, with equality if and only if $X_3 = X_1 \oplus_2 X_2$ with probability one, and X_3 is uniform over $\{0, 1\}$.*

Proof. Note $I(X_1, X_2, X_3; Y) = H(Y) - H(N)$. We proceed by finding all the necessary and sufficient conditions on P_{X_1,X_2,X_3} for which Y is uniform over $\mathbb{Z}_4$. From Figure 3.4, $Y = (X_1 \oplus_2 X_2) \oplus_4 X_3 \oplus_4 N$. Denote $X_2' = X_1 \oplus_2 X_2$. Let $P(X_2' \oplus_4 X_3 = i) = q(i)$ where $i = 1, 2, 3, 4$. Since X_2' and X_3 are binary, $q(3) = 0$. Given the distribution of N is Table 3.1, the distribution of Y is as follows:

$$P(Y=0) = q(0)(\frac{1}{2} - \delta) + q(2)\delta, \qquad P(Y=1) = q(0)\frac{1}{2} + q(1)(\frac{1}{2} - \delta)$$
$$P(Y=2) = q(0)\delta + q(2)(\frac{1}{2} - \delta), \qquad P(Y=3) = q(2)\frac{1}{2} + q(1)\delta$$

Assume $\delta \neq \frac{1}{4}$. By comparing the first and third bounds, we can show that Y is uniform, if and only if $q(1) = 0$ and $q(0) = q(2) = \frac{1}{2}$. Note

$$q(1) = P(X_2' = 0, X_3 = 1) + P(X_2' = 1, X_3 = 0)$$

Therefore, $q(1) = 0$ implies that $X_3 = X_2'$ with probability one. If this condition is satisfied, then $q(0) = P(X_3 = 0)$ and $q(2) = P(X_3 = 1)$. Since $q(0) = q(2) = \frac{1}{2}$ then X_3 is uniform over $\{0, 1\}$. To sum up, we proved that Y is uniform, if and only if 1) $X_3 = X_1 \oplus_2 X_2$. 2) X_3 is uniform over $\{0, 1\}$. □

Lemma 11. *For the setup in Example 9, there exists $\epsilon > 0$ such that the sources (S_1, S_2, S_3) with parameters $(\sigma > 0, \gamma \geq \gamma^* - \epsilon)$ cannot be transmitted reliably using the three-user CES scheme.*

Proof. We first derive an outer bound for the CES scheme. Consider the fourth inequality in Proposition 1. Since $\sigma > 0$ there is no common part. Let $U' = U_{123}U_{12}U_{13}U_{23}$. Suppose the source (S_1, S_2, S_3) in Example 9 can be transmitted using CES, then the following holds

$$h(\gamma) + h(\sigma) \leq \max_{p(u')p(\underline{x}|u'\underline{s})} I(X_1X_2X_3; Y|U'), \tag{3.9}$$

where

$$p(\underline{s}, \underline{x}, u') = p(\underline{s})p(u')p(x_1|s_1, u')p(x_2|s_2, u')p(x_3|s_3, u').$$

It can be shown that the inequality in (3.9) is equivalent to

$$h(\gamma) + h(\sigma) \leq \max_{p(\underline{x}|\underline{s})} I(X_1X_2X_3; Y), \tag{3.10}$$

where $p(\underline{s}, \underline{x}) = p(\underline{s})p(x_1|s_1)p(x_2|s_2)p(x_3|s_3)$.

Next, we argue that the right-hand side in (3.10) is strictly less than $h(\gamma^*) = 2 - H(N)$. For the moment assume this argument is true. Then by the bound above, $h(\gamma) + h(\sigma) < h(\gamma^*)$. This implies that $\exists \epsilon_0 > 0$ such that for any σ, $h(\gamma^*) - h(\gamma) > \epsilon_0$. Hence, as the entropy function is continuous, $\exists \epsilon > 0$ such that any source with $\sigma > 0$ and $\gamma \geq \gamma^* - \epsilon$ cannot be transmitted using the CES scheme.

It remains to show that the right-hand side in (3.10) is strictly less than $2 - H(N)$.

Note that Lemma 10 characterizes the set of all distributions P_{X_1,X_2,X_3} for which $I(X_1X_2X_3;Y) = 2 - H(N)$. The distributions induced in CES scheme for this case satisfy the Markov chain $X_3 - S_3 - X_1, X_2$. Hence, we can show for these distributions, the condition $X_3 = X_1 \oplus_2 X_2$ hold if and only if X_3 is a function of S_3. However, as $\gamma < 1/2$, X_3 cannot be uniform over $\{0, 1\}$. This contradicts with the second condition and completes the proof.

□

Lemma 12. $\exists\ \epsilon' > 0$ *such that the sources with parameters* (σ, γ)*, satisfying* $\sigma \leq \epsilon'$ *and* $|\gamma - \gamma^*| \leq \epsilon'$*, are transmissible reliably.*

Proof. The proof is given in Appendix B.2. □

The final step in our argument is as follows. Take $\epsilon'' = \min\{\epsilon, \epsilon'\}$, where ϵ and ϵ' are as in Lemma 11 and 12, respectively. Then, as a result of these lemmas, the sources (S_1, S_2, S_3) with parameters $\sigma = \epsilon''$ and $\gamma = \gamma^* - \epsilon''$ are transmissible reliably; while they cannot be transmitted using CES scheme.

CHAPTER IV

Structured codes for Communications over MAC with Feedback

The problem of three user MAC with noiseless feedback is depicted in Figure 4.1. This communication channel consists of one receiver and multiple transmitters. After each channel use, the output of the channel is received at each transmitter noiselessly. Gaarder and Wolf [43] showed that the capacity region of the MAC can be expanded through the use of the feedback. This was shown in a binary erasure MAC. Cover and Leung [52] studied the two-user MAC with feedback, and developed a coding strategy using unstructured random codes.

The main idea behind the CL scheme is to use superposition block-Markov encoding. The scheme operates in two stages. In stage one, the transmitters send the messages with a rate outside of the no-feedback capacity region (i.e. higher rates than what is achievable without feedback). The transmission rate is taken such that each user can decode the other user's message using feedback. In this stage, the receiver is unable to decode the messages reliably, since the transmission rates are outside the no-feedback capacity region. Hence, the decoder only is able to form a list of "highly likely" pairs of messages. In the second stage, the encoders fully cooperate to send the messages (as if they are sent by a centralized transmitter). The receiver decodes the message pair from its initial list. After the initiation block, superposition coding

is used to transmit the sequences corresponding to the two stages.

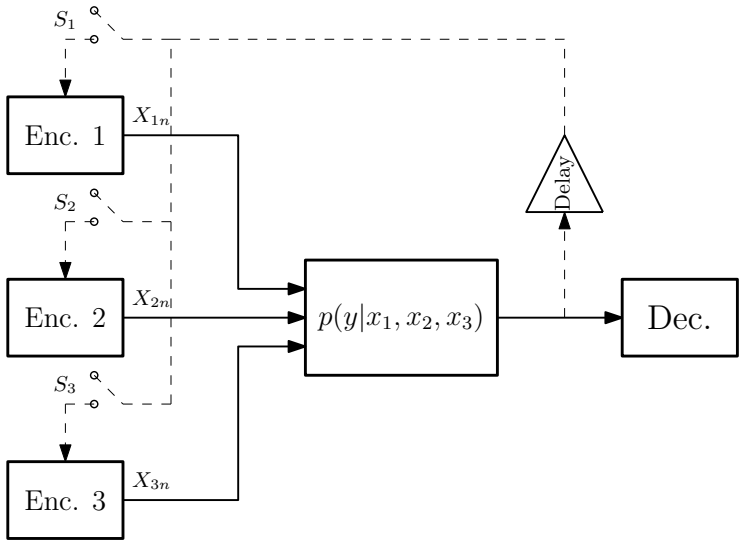

Figure 4.1: The three-user MAC with noiseless feedback. If the switch S_i is closed, the feedback is available at the ith encoder, where $i = 1, 2, 3$.

The single-letter achievable rate region for the CL scheme was characterized in [52]. Later, it was shown that the CL scheme achieves the feedback capacity for a class of MAC with feedback [75]. However, this is not the case for the general MAC with feedback [76]. Several improvements to the CL achievable region were derived [77], [78]. In [77] and [78], additional stages are appended to the CL scheme. In these schemes, the encoders decode each others' messages in several stages. Kramer [47], used the notion of *directed information* to derive the capacity region of the two-user MAC with feedback. However, the characterization is not computable, since it is an infinite letter characterization. Finding a computable characterization of the capacity region remains an open problem.

In this chapter, we study the problem of three-user MAC with feedback. We propose a new coding scheme which builds upon the CL scheme. We derive a computable single-letter achievable rate region for this scheme, and show that the new region improves upon the previous known achievable regions for this problem. Recently, we

showed that the application of structured codes results in improved performance for the problem of transmission of sources over the MAC [25]. Here, we use the ideas proposed in [25] to prove the necessity of structured codes in the problem of MAC with feedback. Specifically, we use *quasi-linear* codes that are proposed in [79].

The coding scheme operates in three stages. In stage one, the encoders send independent messages with rates outside of the CL region. Therefore, encoders are unable to decode each others' messages. However, each encoder can decode the binary sum of the messages of the other two encoders. In stage two, the messages are superimposed on the summation which is decoded in the previous stage. At the end of this stage, the encoders decode each others' messages. Stage three is similar to the second stage in CL scheme. We provide an example where the new coding scheme achieves optimal performance, whereas the previous schemes are suboptimal. Finally, we prove that any optimality achieving coding scheme must use encoders whose set of output sequences is linearly closed.

4.1 Preliminaries and Model

In what follows, we formulate the problem of communications over MAC-FB. We restrict ourselves to MAC with noiseless feedback in which all or a subset of the transmitters have access to the feedback perfectly. Consider a t-user MAC identified by a transition probability matrix $P_{Y|X_1,X_2,...,X_t}$ as in Definition 26. Let y^n be a realization of the output of the channel after n uses, where x_i^n is the ith input sequence of the channel, $i \in [1,t]$. Then, the following condition is satisfied:

$$p(y_n|y^{n-1}, x_i^{n-1}, i \in [1,t]) = p(y_n|x_{1n}, x_{2n}, ..., x_{tn}). \tag{4.1}$$

It is assumed that noiseless feedback is made available, with one unite of delay, to a subset $\mathcal{T} \subseteq [1,t]$ of the transmitters. Figure 4.1 illustrates an example of this

setup. In the figure, the switches $S_i, i = 1, 2, 3$ determine which transmitter receives the feedback. A formal definition of a MAC-FB setup is given in the following.

Definition 32. *A t-user MAC-FB setup is characterized by a t-user MAC $P_{Y|X_1,X_2,...,X_t}$ and a subset $\mathcal{T} \subseteq [1, t]$ determining the transmitters which have access to the feedback.*

Definition 33. *For a t-user MAC-FB setup with a subset $\mathcal{T} \subseteq [1, t]$, an $(N, \Theta_1, ..., \Theta_t)$ coding scheme consists of t sequences of encoding functions defined as,*

$$e_{i,n} : [1, \Theta_i] \times \mathcal{Y}^{n-1} \to \mathcal{X}_i, \quad \textit{for } i \in \mathcal{T}, \quad \textit{and} \quad e_{j,n} : [1, \Theta_j] \to \mathcal{X}_j, \quad \textit{for } j \in [1, t] \backslash \mathcal{T}.$$

where $n \in [1, N]$ and a decoding function denoted by

$$d : \mathcal{Y}^N \to [1, \Theta_1] \times [1, \Theta_2] \times \cdots \times [1, \Theta_t].$$

We use a unified notation $e_{i,n}(m, y^{n-1})$ to denote the encoders, as it is understood that for $i \notin \mathcal{T}$ the encoder $e_{i,n}$ is only a function of the message m. Moreover, for shorthand, the encoders of the coding scheme are denoted by $\underline{e}$.

It is assumed that, Transmitter i receives a message index M_i which is drawn randomly and uniformly from $[1, \Theta_i]$, where $i \in [1, t]$. Furthermore, the message indexes $(M_1.M_2, ..., M_t)$ are assumed to be mutually independent. For this setup, the average probability of error is defined as

$$P_{err}(\underline{e}) \triangleq \mathbb{P}\{d(Y^N) \neq (M_1, M_2, ..., M_t)\}, \tag{4.2}$$

where $\underline{e}$ denotes the encoders of the coding scheme.

Definition 34. *For a t-user MAC-FB, a rate-tuple $(R_1, R_2, ..., R_t)$ is said to be*

achievable using an $(N, \Theta_1, \Theta_2, ..., \Theta_t)$ *coding scheme, if for any* $\epsilon > 0$

$$P_{err}(\underline{e}) < \epsilon, \quad \frac{1}{N} \log_2 \Theta_i \geq R_i - \epsilon, \quad \textit{where} \;\; i \in [1, t].$$

Based on our earlier discussion in Section 3.1, one can consider a randomized coding strategy for which the encoding functions are selected randomly according to a predefined probability measure. For that we take a similar approach as in Definition 20 and define a randomized coding strategy as in the following.

Definition 35. *For a* t*-user MAC-FB setup, an* $(N, \Theta_1, \Theta_2, ..., \Theta_t)$ *randomized coding strategy is characterized by a probability measure* P_N *on the set of all encoding functions* $(e_{i,n})$ *with* $i \in [1, t]$ *and* $n \in [1, N]$ *as in Definition 33.*

Let $\underline{E}$ denote random encoders of a randomized coding strategy with probability measure P_N, then the expected error probability is

$$\bar{P}_{err} \triangleq \mathbb{E}_{\mathsf{P}_N}[P_{err}(\underline{E})\}],$$

where $P_{err}(\cdot)$ is defined as in (4.2).

Definition 36. *A rate-tuple* $(R_1, R_2, ..., R_t)$ *is said to be achievable using an* $(N, \Theta_1, \Theta_2, ..., \Theta_t)$ *randomized coding strategy with probability measure* P_N *if for any* $\epsilon > 0$

$$\bar{P}_{err} < \epsilon, \quad \frac{1}{N} \log_2 \Theta_i \geq R_i - \epsilon, \quad i \in [1, t].$$

Remark 14. The capacity region is defined as the closure of the set of all rate tuples $(R_1, R_2, ..., R_t)$ that are achievable using a coding scheme or a randomized coding strategy.

We extend the results of Kramer for t-user MAC with feedback. We derive a multi-letter characterization for the capacity region. We use the notion of *directed*

information presented in [47].

$$H(\mathbf{Y}^n||\mathbf{X}^n) = \sum_{k=1}^{n} H(\mathbf{Y}_k|\mathbf{Y}^{k-1}, \mathbf{X}^k).$$

Directed information from a sequence $\mathbf{X}^n$ to a sequence $\mathbf{Y}^n$ is defined as

$$I(\mathbf{X}^n \to \mathbf{Y}^n) = H(\mathbf{Y}^n) - H(\mathbf{Y}^n||\mathbf{X}^n).$$

Directed information from a sequence $\mathbf{X}^n$ to a sequence $\mathbf{Y}^n$ when causally conditioned on $\mathbf{Z}^n$ is defined by

$$I(\mathbf{X}^n \to \mathbf{Y}^n||\mathbf{Z}^n) = H(\mathbf{Y}^n||\mathbf{Z}^n) - H(\mathbf{Y}^n||\mathbf{X}^n\mathbf{Z}^n).$$

Let $I(\mathbf{X}^n \to \mathbf{Y}^n)$ be the directed information from $\mathbf{X}^n$ to $\mathbf{Y}^n$. Define

$$I_n(X \to Y) \triangleq \frac{1}{n} I(\mathbf{X}^n \to \mathbf{Y}^n). \tag{4.3}$$

Definition 37. *Given a positive integer N and a t-user MAC with feedback, define $\mathcal{R}_L$ as the convex hull of the set of all rates $(R_1, R_2, ..., R_t)$ such that,*

$$R_{\mathcal{A}} \leq I_L(X_{\mathcal{A}} \to Y||X_{\mathcal{A}^c}), \quad \textit{for all} \;\; \mathcal{A} \subseteq [1, t]. \tag{4.4}$$

where the conditional distribution $p(x_{1,l}, x_{2,l}, ..., x_{t,l}|x_1^{l-1}, x_2^{l-1}, ..., x_t^{l-1}, y^{l-1})$ factors as $\prod_{i=1}^{t} p(x_{i,l}|x_i^{l-1}y^{l-1})$.

Proposition 2. *The capacity region of t-user MAC with feedback is characterized by $\mathcal{C}_{\mathcal{FB}} = \bigcup_{L=1}^{\infty} \mathcal{R}_L$.*

Proof. The proof is a streightforward generalization of the proof in [47]. □

Note that this is a multi-letter characterization, and is not computable.

4.2 Conferencing Common Information in MAC-FB

Prior to the start of communications over a MAC-FB setup, the messages are mutually independent. During the communication, after multiple uses of the channel, the messages are statistically correlated conditioned on the feedback. In this section, we make a connection to the problem of MAC with correlated sources to design coding strategies that exploit the statistical correlation among the messages. We use the notion of conferencing common information to propose a new coding strategy for 3-user MAC-FB.

We begin by explaining the intuition behind Cover-Leung (CL) scheme [52]. In CL scheme, the message indexes are transmitted in N channel uses. The communications take place in B blocks, each of length n, where $N = Bn$. The message for Transmitter i is divided into B sub-messages denoted by $(M_{i,1}, M_{i,2}, ..., M_{i,B})$, where $i = 1, 2$. At the first block of the communication ($b = 1$), the transmitters send the sub-messages $(M_{1,1}, M_{2,1})$ with a rate outside of the no-feedback capacity region (i.e. higher rates than what is achievable without feedback). Therefore, the receiver is unable to decode the sub-messages reliably - rather it is only able to form a list of "highly likely" pairs of messages. However, the transmission rates are taken to be sufficiently low such that each user can reliably decode the other user's sub-message using the feedback. Therefore, at the end of this block, $(M_{1,1}, M_{2,1})$ is known at the two transmitters with "high" probability. Hence, at the next block of the transmission ($b = 2$), one can view $(M_{1,1}, M_{2,1})$ as a common information that is known at the transmitters. At this block, a superposition block-Markov encoding is used to send the common information as well the new sub-messages $(M_{1,2}, M_{2,2})$. At the end of this block, the receiver decodes $(M_{1,1}, M_{2,1})$ from its initial list. The scheme is repeated for the next blocks $b > 2$. Using this approach, the following rate-region is achievable for communications over a MAC $P_{Y|X_1,X_2}$ with noiseless feedback available at the two

transmitters [52]:

$$R_1 \leq I(X_1; Y|X_2, U), \qquad R_2 \leq I(X_2; Y|X_1, U), \qquad R_1 + R_2 \leq I(X_1, X_2; Y).$$

Where, the joint distribution of the random variables (U, X_1, X_2, Y) factors as

$$P_U P_{X_1|U} P_{X_2|U} P_{Y|X_1,X_2}.$$

As explained, the decoded sub-messages $(M_{1,b}, M_{2,b})$ are used as a common information for the next block of transmission. One can extend CL scheme for a multi-user MAC-FB setup (say a three-user MAC-FB). In this setup, the transmitters send the messages with rates outside of the no-feedback capacity region. Hence, the receiver is not able to decode the messages. However, the transmission rates are taken to be sufficiently low so that each user can decode the sub-messages of the other users. The decoded sub-messages at the end of each block b are used as uni-variate common parts for the next block of transmission. Also, one can use the notion of conferencing common information as defined in Section 3.1.3 to design a more sophisticated coding scheme.

In what follows, we gave the intuition behind the use of conferencing common information in MAC-FB. Consider a three-user MAC-FB setup as depicted in Figure 4.2. Similar to the two-user version of the problem, the communications take place in B blocks each of length n. Moreover, the message at Transmitter i is divided into B sub-messages denoted by $(M_{i,1}, M_{i,2}, ..., M_{i,B})$, where $i = 1, 2, 3$. Suppose, the transmission rates are such that neither the decoder nor the transmitters can decode the messages. However, at each block b, the rates are sufficiently low so that each transmitter is able to decode the modulo-two sum of the other two sub-messages. For instance, Transmitter 1 can decode $M_{2,b} \oplus M_{3,b}$ with high probability. Let $T_{i,b}$ denote the decoded sum at Transmitter i, where $i = 1, 2, 3$. Then, for binary messages,

$T_{1,b} \oplus T_{2,b} \oplus T_{3,b} = 0$ with high probability. As a result, (T_1, T_2, T_3) can be interpreted as 2-additive conferencing common parts (see Definition 25). Building upon this intuition, in what follows, we propose a coding strategy for communications over 3-user MAC-FB. Further, we derive a new commutable achievable rate region for the three-user MAC with feedback problem.

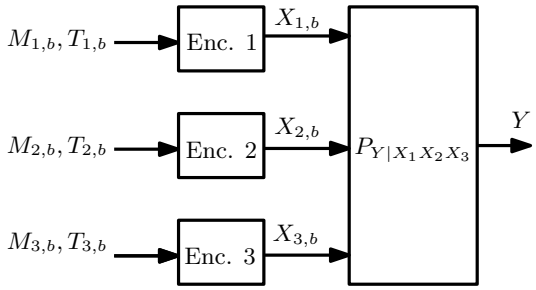

Figure 4.2: Applications of conferencing common information for communications over MAC-FB. The new sub-messages at block b are denoted by $M_{i,b}$. At the end of block $b-1$, each Transmitter decodes the modulo-two sum of the other two transmitters. The decoded sums are denoted by $T_{i,b}, i = 1, 2, 3$. Note that $T_{1,b} \oplus T_{2,b} \oplus T_{3,b} = 0$ with probability close to one.

Definition 38. *For a given set $\mathcal{U}$ and a three-user MAC with feedback $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3, \mathcal{Y}, P_{Y|X_1X_2X_3})$, define $\mathscr{P}$ as the collection of all distributions P of the form*

$$P_U P_{V_1V_2V_3} \prod_{i=1}^{3} P_{T_i} P_{X_i|UT_iV_i} P_{Y|X_1X_2X_3},$$

where (T_1, T_2, T_3) are mutually independent with uniform distribution over $\mathbb{F}_2$, (V_1, V_2, V_3) are pairwise independent each with uniform distribution over $\mathbb{F}_2$, and $P_{V_1V_2V_3}(v_1, v_2, v_3) = \frac{1}{4}\mathbb{1}\{v_1 \oplus v_2 \oplus v_3 = 0\}$.

Fix a distribution $P \in \mathscr{P}$. Denote $S_i = (X_i, T_i, V_i)$ for $i = 1, 2, 3$. Consider two sets of random variables (U, S_1, S_2, S_3, Y) and $(\tilde{U}, \tilde{S}_1, \tilde{S}_2, \tilde{S}_3, \tilde{Y})$. Suppose the distribution of each set of the random variables is P. Then with this notation we

have

$$P_{US_1S_2S_3Y} = P_{\tilde{U}\tilde{S}_1\tilde{S}_2\tilde{S}_3\tilde{Y}} = P$$

Theorem IV.1. *Consider a MAC* $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3, \mathcal{Y}, P_{Y|X_1X_2X_3})$*, and a distribution* $P \in \mathscr{P}$*. For any subset* $\mathcal{A} \subseteq \{1,2,3\}$*, and for any distinct elements* $i, j, k \in \{1,2,3\}$ *the following bounds hold*

$$\begin{aligned}
R_{\mathcal{A}} &\leq I(X_{\mathcal{A}}; Y|US_{\mathcal{A}^c}\tilde{V}_1\tilde{V}_2\tilde{V}_3) + I(U; Y|\tilde{U}\tilde{Y}) \\
R_i + R_j &\leq I(T_i \oplus T_j; Y|UT_kX_k\tilde{V}_1\tilde{V}_2\tilde{V}_3) \\
&\quad + I(\tilde{X}_i\tilde{X}_j; \tilde{Y}|\tilde{U}\tilde{S}_k\tilde{V}_1\tilde{V}_2\tilde{V}_3V_k) \\
&\quad + I(\tilde{X}_i\tilde{X}_j; Y|\tilde{U}\tilde{S}_k\tilde{V}_1\tilde{V}_2\tilde{V}_3US_k\tilde{Y}) \\
R_i + R_j &\leq \frac{H(W_i) + H(W_j)}{H(W_i \oplus W_j)} I(T_i \oplus T_j; Y|UT_kX_k),
\end{aligned}$$

where 1) W_i*, is a Bernoulli random variable that is independent of all other random variables, 2) the equality* $V_i = \tilde{T}_j \oplus \tilde{T}_k$ *holds with probability one, and 3) the Markov chain*

$$\tilde{U}, \tilde{S}_1, \tilde{S}_2, \tilde{S}_3 \leftrightarrow V_1, V_2, V_3 \leftrightarrow U, T_i, X_i,$$

holds for $i = 1, 2, 3$*.*

Proof. The proof is given in Appendix C.1. □

Remark 15. The rate region in Theorem IV.1 contains the three-user extension of the CL region. For that set V_1, V_2, V_3 to be independent of all other random variables. This gives a distribution in $\mathscr{P}$.

4.3 Necessity of Structured Codes for MAC-FB

In this section, we show that coding strategies based on structured codes are necessary for the problem of MAC with feedback. We first provide an example of a MAC with feedback. Then, we propose a coding scheme using linear codes, and show that such coding scheme achieves optimality in terms of achievable rates.

Example 10. Consider the three-user MAC with feedback problem depicted in Figure 4.3. In this setup, there is a MAC with three binary inputs, where the ith input is denoted by the pair (X_{i1}, X_{i2}) for $i = 1, 2, 3$. The output of the channel is denoted by a binary vector (Y_1, Y_{21}, Y_{22}). Assume that noiseless feedback is available only at the third transmitter.

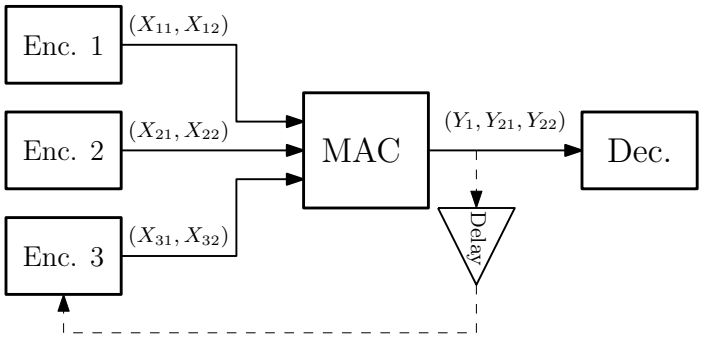

Figure 4.3: The MAC with feedback setup for Example 10.

The MAC in this setup consists of two parallel channels. The first channel is a three-user binary additive MAC with inputs (X_{11}, X_{21}, X_{31}), and output Y_1. The transition probability matrix of this channel is described by the following relation:

$$Y_1 = X_{11} \oplus X_{21} \oplus X_{31} \oplus \tilde{N}_\delta,$$

where $\tilde{N}_\delta$ is a Bernoulli random variable with bias δ, and is independent of the inputs.

The second channel is a MAC with (X_{12}, X_{22}, X_{32}) as the inputs, and (Y_{21}, Y_{22}) as

the output. The conditional probability distribution of this channel satisfies

$$(Y_{21}, Y_{22}) = \begin{cases} (X_{12} \oplus N_\delta, X_{22} \oplus N'_\delta), & \text{if } X_{32} = X_{12} \oplus X_{22}, \\ (N_{1/2}, N'_{1/2}), & \text{if } X_{32} \neq X_{12} \oplus X_{22}, \end{cases} \tag{4.5}$$

where $N_\delta, N'_\delta, N_{1/2}$ and $N'_{1/2}$ are independent Bernoulli random variables with parameter $\delta, \delta, \frac{1}{2}$, and $\frac{1}{2}$, respectively. The relation between the output and the input of the channel is depicted in Figure 4.4. The channel operates in two states. If the condition $X_{31} = X_{12} \oplus X_{22}$ holds, the channel would be in the first state (the left channel in Figure 4.4); otherwise it would be in the second state (the right channel in Figure 4.4). In this channel, N_δ and N'_δ are Bernoulli random variables with identical bias δ. Whereas, $N_{1/2}$ and $N'_{1/2}$ are Bernoulli random variables with bias $\frac{1}{2}$. We assume that $\tilde{N}_\delta, N_\delta, N'_\delta, N_{1/2}$, and $N'_{1/2}$ are mutually independent, and are independent of all the inputs.

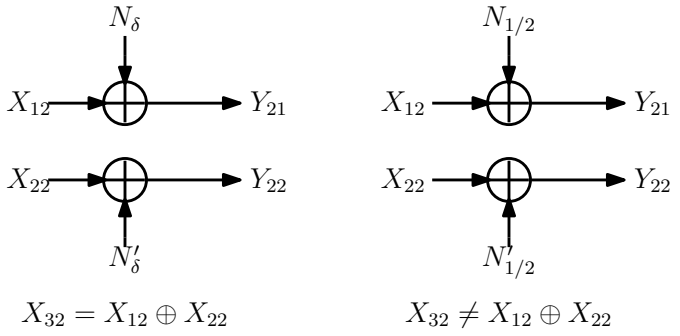

Figure 4.4: The second channel for Example 10. If the condition $X_{31} = X_{12} \oplus X_{22}$ holds, the channel would be the one on the left; otherwise it would be the right channel.

We use linear codes to propose a new coding strategy for the setup given in Example 10. The scheme uses a large number L of blocks. The length of each block is n. Each encoder has two outputs, one for each channel. We use identical linear codes with length n and rate $\frac{k}{n}$ for each transmitter. The coding scheme at each block is performed in two stages. In the first stage, each transmitter encodes the

fresh message at the beginning of the block l, where $1 \leq l \leq L$. The encoding process is performed using the identical linear codes. At the end of the block l, the feedback is received by the third user. In stage 2, the third user uses the feedback from the first channel (that is Y_1) to decode the binary sum of the messages of the other encoders. Then, it encodes the summation, and sends it through its second output. If the decoding process is successful at the third user, then the relation $X_{32} = X_{12} \oplus X_{22}$ holds with probability one. This is because identical linear codes are used to encode the messages. As a result of this equality, the channel in Figure 4.4 is in the first state with probability one. In the next Lemma, we show that the rate

$$(1 - h(\delta), 1 - h(\delta), 1 - h(\delta))$$

is achievable using this strategy.

Lemma 13. *For the channel given in Example 10, the rate triple* $(1 - h(\delta), 1 - h(\delta), 1 - h(\delta))$ *is achievable.*

Proof. The proof is given in Appendix C.2. □

Remark 16. Based on Preposition 2, the triple $(1-h(\delta), 1-h(\delta), 1-h(\delta))$ is a corner point in the capacity region of the channel in Example 10. This implies the optimality of the above coding strategy in terms of achievable rates.

The above coding strategy is different from known schemes in two ways: 1) Identical linear codes are used to encode the messages, 2) The third user uses feedback to decode only the binary sum others' messages.

One implication of Remark 16 is that the proposed coding scheme achieves optimality. We show a stronger result in this Subsection. We prove that every coding scheme that achieves $(1 - h(\delta), 1 - h(\delta), 1 - h(\delta))$, should carry certain algebraic structures such as closeness under the binary addition.

Suppose there exists a (N, M_1, M_2, M_3) transmission system with rates close to $R_i = 1 - h(\delta)$, and average probability of error close to 0, in particular

$$\bar{P} < \epsilon, \quad \frac{1}{n} \log_2 M_i \geq 1 - h(\delta) - \epsilon, \quad i = 1, 2, 3,$$

where $\epsilon > 0$ is sufficiently small. Since there is no feedback at the first and second encoder, the transmission system predetermines a codebook for user 1 and 2. Note that there are two outputs for encoder 1 and 2. Suppose $\mathcal{C}_{12}$ and $\mathcal{C}_{22}$ are the codebooks assigned to the second output of encoder 1 and encoder 2, respectively.

Let $\mathbf{X}_{i2}^N$ be the second output of encoder i, where $i = 1, 2, 3$. Let $X_{i2,l}$ denote the lth component of X_{i2}^N, where $1 \leq l \leq N,\ i = 1, 2, 3$. The following lemmas hold for this transmission system.

Lemma 14. *For any fixed $c > 0$, define*

$$\mathcal{I}_c^N := \{l \in [1 : N] : P(X_{32,l} \neq X_{12,l} \oplus X_{22,l}) \geq c\}.$$

Then, the inequality $\frac{|\mathcal{I}_c^N|}{N} \leq \frac{\eta(\epsilon)}{2c(1-h(\delta))}$ holds, where $\eta(\epsilon)$ is a function such that, $\eta(\epsilon) \to 0$, as $\epsilon \to 0$.

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

The Lemma implies that in order to achieve $(1-h(\delta), 1-h(\delta), 1-h(\delta))$, the third user needs to decode $X_{12,l} \oplus X_{22,l}$ for "almost all" $l \in [1 : N]$. This requirement is necessary to insure that the channel given in Figure 4.4 is in the first state.

In the next step, we use the results of Lemma 14, and drive two necessary conditions for decoding $X_{12} \oplus X_{22}$.

Lemma 15. *The following holds*

$$\frac{1}{N}\Big|\ \log||\mathcal{C}_{12}\oplus\mathcal{C}_{22}||-\log||\mathcal{C}_{12}||\ \Big|\leq\lambda_1(\epsilon),$$
$$\frac{1}{N}\Big|\ \log||\mathcal{C}_{12}\oplus\mathcal{C}_{22}||-\log||\mathcal{C}_{22}||\ \Big|\leq\lambda_2(\epsilon),$$

where $\lambda_j(\epsilon)\to 0$, *as* $\epsilon\to 0, j=1,2$.

Proof. The proof is given in Appendix C.4. □

As a result of this lemma, $\log||\mathcal{C}_{12}\oplus\mathcal{C}_{22}||$ needs to be close to $\log||\mathcal{C}_{12}||$ and $\log||\mathcal{C}_{22}||$. This implies that $\mathcal{C}_{12}$ and $\mathcal{C}_{22}$ possesses an algebraic structure, and are *almost* close under the binary addition. Not that for the case of unstructured random codes $||\mathcal{C}_{12}\oplus\mathcal{C}_{22}||\approx||\mathcal{C}_{12}||\times||\mathcal{C}_{22}||$. Hence, unstructured random coding schemes are suboptimal in this example.

Remark 17. The three-user extension of CL scheme is suboptimal. Because, the conditions in Lemma 15 are not satisfied.

CHAPTER V

Algebraic Structures for Multiple Descriptions

5.1 Introduction

In Chapter II- IV, we investigated algebraic structures in multi-terminal communication systems with discrete alphabets. In this chapter, we build upon the results in discrete setting and investigate the applications of algebraic structured codes (such as Lattices) in communication systems with continuous alphabets such as MD. Lattice codes are analogous of linear codes in Euclidean spaces, e.g. $\mathbb{R}^d$. A lattice code in $\mathbb{R}^d$ is defined as the set of all linear combinations, with integer coefficients, of a given set of linearly independent vectors in $\mathbb{R}^d$. With the recent developments in Gaussian network information theory, lattices have received significant attentions due to their applications for efficient quantization, channel coding and cryptography in continuous settings [54] [55]. Lattices have become a standard tool to design block codes for communications over AWGN channels. Lattice quantizers have been of great interest in compression of continuous sources [80] [81]. In the PtP communication settings, the interest towards such codes is mainly due to reduced complexity of encoding and decoding. In multi-terminal commumcations, the significance of lattice codes is augmented because they give performance gains over unstructured codes in terms of achievable rates. These gains are observed in a variety of multi-terminal settings such as dirty MAC [21, 22], and interference channel [27, 29, 32]

Traditionally, performance characterization of lattices was carried out using Gaussian test channels. Such techniques are known to be suitable for Gaussian source/channel setups. However, for general distributions in channel/source coding, it is difficult to derive achievable rates of lattices using such techniques. Recently, a new method is introduced in [57] to overcome this challenge. In the method, first the objective continuous source/ channel problem is quantized to obtain its discrete version. The performance analysis is carried out for the discrete version of the problem and inner bounds are derived in terms of discrete mutual information quantities. Then, it is shown that as the discretization process keeps refining, the mutual information terms converge to the continuous ones. Hence, inner bounds are obtained for the original continuous source/channel setup. Using this approach, the authors in [57] show achievability of the Wyner-Ziv rate region in the PtP set-up.

In this chapter, we investigate the applications of lattice-based strategies for MD problem. A MD problem, as depicted in Figure 5.1, consists of one encoder and several decoders. The encoder compresses the source into several descriptions and transmits them through noiseless links. Each decoder receives a subset of these descriptions. The decoder then finds a reconstruction of the source using the descriptions it has received. The problem arises naturally when a transmitter wishes to send data to different receivers with varying quality of service demands. Another instance of the MD problem emerges when dealing with channel blackouts. In this situation, satisfactory source reconstruction is ensured via transmitting different descriptions of the source through multiple paths. In the latter perspective, each decoder represents the actual receiver in a specific blackout situation where a subset of the transmission links are experiencing failures; these failures are known at the decoder, but not at the encoder.

The best known achievablity scheme for discrete MD problem with two descriptions is due to Zhang and Berger [82]. Zhang-Berger scheme consists of a base layer codebook and several refinement layer codebooks. The base layer is transmitted over

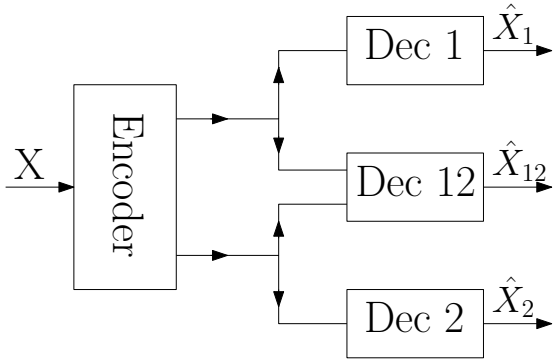

Figure 5.1: An example of a MD problem with two-descriptions. The problem consists of one encoder with three decoders. Encoder produces two descriptions of the source. Decoder 1 and 2 receive only one description of the source; whereas Decoder 12 has access to the two descriptions sent by the encoder.

all descriptions and is decoded at every decoder, while the refinement layer contains codebooks decoded at individual decoders. There has been several attempts to generalize Zhang-Berger scheme for MD with L-descriptions [83–86]. CMS scheme introduced in [86] unifies all of the previous schemes. The strategy is based on random unstructured codes and binning. This scheme is enhanced by adding a strututured coding layer [87]. In this Chapter, we provide a new achievable RD region for the L-descriptions problem using random lattice codes. We show that using a pair of *nested lattice* quantizers with the same inner code gives strict improvements over CMS scheme in terms of achievable rates.

5.2 Preliminaries

Coset linear codes: For a prime q, let $\mathbb{F}_q$ denote the modulo-q field of integers. A linear code over $\mathbb{F}_q$ with blocklength n is a subspace of $\mathbb{F}_q^n$. A k-dimensional linear subspace of $\mathbb{F}_q^n$ can be viewed as the image of a linear transformation from $\mathbb{F}_q^k$ to $\mathbb{F}_q^n$. A linear code with generator matrix $\mathbf{G_{k\times n}}$ is defined as

$$\mathcal{C} = \{u^k \mathbf{G_{k\times n}} | \mathbf{u^k} \in \mathbb{F}_\mathbf{q}^\mathbf{k}\}.$$

Such a characterization is unique up to invertible linear transformations of $\mathbf{G_{k\times n}}$. A coset linear code is a shifted version of a linear code and is characterized by a generator matrix $\mathbf{G_{k\times n}}$ and a dither $\mathbf{b}^n$:

$$\mathcal{C} = \{u^k \mathbf{G_{k\times n}} + \mathbf{b^n} | \mathbf{u^k} \in \mathbb{F}_\mathbf{q}^\mathbf{k}\}.$$

The rate of the linear code is given by $R \triangleq \frac{k}{n} \log_2 q$.

Nested Linear Codes: A pair of coset linear codes $(\mathcal{C}_i, \mathcal{C}_o)$, are called nested if $\mathcal{C}_i \subseteq \mathcal{C}_o$. $\mathcal{C}_o$ and $\mathcal{C}_i$ are called the outer and inner codes, respectively. A nested linear code over $\mathbb{F}_q$ is characterized by two generator matrices $\mathbf{G_{k\times n}}$ and $\mathbf{\Delta G_{l\times n}}$ and a dither $\mathbf{b}^n$, all with elements in $\mathbb{F}_q$. With this notation the inner code is characterized as

$$\mathcal{C}_i \triangleq \{u^k \mathbf{G_{k\times n}} + \mathbf{b^n} | \mathbf{u^k} \in \mathbb{F}_\mathbf{q}^\mathbf{k}\}.$$

and the outer code is defined as

$$\mathcal{C}_o \triangleq \{u^k \mathbf{G_{k\times n}} + \mathbf{v^l \Delta G_{l\times n}} + \mathbf{b^n} | \mathbf{u^k} \in \mathbb{F}_\mathbf{q}^\mathbf{k}, \mathbf{v^l} \in \mathbb{F}_\mathbf{q}^\mathbf{l}\}$$

Here $(\mathbf{G}_{k\times n}, \mathbf{b}^n)$ is a characterization for $\mathcal{C}_{in}$ and $([\mathbf{G}, \mathbf{\Delta G}]^t, \mathbf{b}^n)$ characterizes $\mathcal{C}_o$. Note that since $\mathcal{C}_i \subset \mathcal{C}_o$, one can always find such a characterization where the dithers are equal. $\mathcal{C}_o$ can be viewed as a union of shifted versions of $\mathcal{C}_i$, where the shift vector is chosen from the linear subspace generated by $\mathbf{\Delta G}$. Each of these shifted version of $\mathcal{C}_i$ is called a bin of $\mathcal{C}_o$ and is shown by $\mathcal{B}_m$:

$$\mathcal{B}_m = \{a\mathbf{G} + m\mathbf{\Delta G} + B | a \in \mathbb{F}_p^k\}.$$

Lattice Code Generation: A lattice code is a subspace of $\mathbb{R}^n$ which is closed under real addition. Also a coset lattice code is defined as a shifted version of a lattice code.

A method for generating such constructions using linear codes was presented in [57], here we present a brief summary of the method. Take an arbitrary coset linear code $\mathcal{C}$ over $\mathbb{F}_p$. Choose $\gamma \in \mathbb{R}^+$. Such γ is called the step size of the lattice code and determines the distance between codewords in the lattice construction. First the linear code is symmetrized with respect to the origin and scaled for the step-size γ. Define this symmetrized and scaled version as follows:

$$\Lambda(\mathcal{C}, \gamma, p) = \{\gamma(\mathbf{c}^n - \frac{p-1}{2}) | \mathbf{c}^n \in \mathcal{C}\}.$$

$\Lambda(\mathcal{C}, \gamma, p)$ is used as the building block for constructing the lattice code. We generate the lattice code by considering disjoint shifted copies of $\Lambda(\mathcal{C}, \gamma, p)$:

$$\bar{\Lambda}(\mathcal{C}, \gamma, p) = \bigcup_{v \in \gamma p \mathbb{Z}^n} \{v + \Lambda(\mathcal{C}, \gamma, p)\}.$$

Coset and nested lattice codes are also defined in a similar fashion by constructing a pair (Λ_i, Λ_o) from an underlying pair of nested linear codes $(\mathcal{C}_i, \mathcal{C}_o)$. Similar to nested linear codes, for $m \in \mathbb{F}_p^l$ bin m can be defined as:

$$\bar{\mathcal{B}}_m = \{\gamma(\mathbf{c}^n - \frac{p-1}{2}) | \mathbf{c}^n \in \mathcal{B}_m\}, \tag{5.1}$$

where $\mathcal{B}_m$ is a bin in the underlying nested linear code.

Measures of Information: We use the notion of [57] to define Kullback-Leibler divergence and the mutual information. Consider random variables U, V on $\mathbb{R}^d$ with probability measure P_{UV}. Take an arbitrary finite measurable partition $\mathcal{A} = \{A_i\}_{i \in 1:n}$ of R^d. Define random variables $U_\mathcal{A}, V_\mathcal{A}$ taking values from the set $[1 : n]$ with the following probability measure:

$$P_{U_\mathcal{A}, V_\mathcal{A}}(i, j) = P_{U,V}(A_i, A_j).$$

The Kullback-Leibler divergence between U and V is defined as follows:

$$D(P_U \| P_V) := \sup_{\mathcal{A} \in A_{\mathbb{R}^d}} D(P_{U_{\mathcal{A}}} \| P_{V_{\mathcal{A}}}),$$

where $A_{\mathbb{R}^n}$ is the set of all finite measurable partitions of $\mathbb{R}^n$.

Typicality: We use the definition of weak* typicality in [57]. For a subset A of $\mathbb{R}^n$, the set $A^\epsilon = \{x \in \mathbb{R}^d \mid \exists y \in A, \|x - y\| \leq \epsilon\}$ is called the ϵ-neighborhood of A. For a given probability measure P_1, P_2 the Prokhorov distance is given as follows:

$$\pi(P_1, P_2) = \inf\{\epsilon \mid P_1(A) < P_2(A^\epsilon) + \epsilon, P_2(A) < P_1(A^\epsilon) + \epsilon, \forall A \in \mathbb{B}(\mathbb{R}^d)\},$$

where $\mathbb{B}(\mathbb{R}^d)$ denotes the Borel σ- algebra on $\mathbb{R}^d$.

For a pair $x, y \in \mathbb{R}^d$ define the empirical probability measure induced by (x, y) on the set of Borel sets in $\mathbb{R}^d$ as:

$$\bar{P}_{xy}(A, B) := \sum_{i=1}^{n} \mathbb{1}\{x_i \in A, y_i \in B\}, \forall A, B \in \mathbb{B}(\mathbb{R}^d). \tag{5.2}$$

A sequence x is weak* ϵ-typical with distribution P_X if

$$\pi(\bar{P}_x, P_X) < \epsilon. \tag{5.3}$$

Also sequences x, y are said to be weak* joint ϵ-typical with distribution P_{XY}, if

$$\pi(\bar{P}_{xy}, P_{XY}) < \epsilon. \tag{5.4}$$

5.3 Random Coding Improvements for Discrete Sources

The CMS scheme with binning was recently improved upon. Here we give a brief summary of the improved scheme for the general L-descriptions problem.

Codebook Generation: Let $C_M, M \in 2^{2^{[1:L]}}$ be the set of codebooks used in the improved version, where 2^A is the set of all subsets of A. C_M is decoded at decoder N, if $N \in M$. U_M is the underlying random variable for codebook C_M. Define a joint probability distribution $P_{\mathbf{U}}$ on random variables $U_M \in 2^{2^{[1:L]}}$. Each codebook C_M is generated randomly and independently based on P_{U_M} with rate r_M. The i^{th} description bins the codebook randomly, uniformly and independently with binning rate $\rho_{M,i}$. These bin numbers are sent through the description. Decoder N, upon receiving the descriptions finds a unique vector $(u_M^n)_{N\in M}$ of jointly typical sequences. If the vector does not exist or is not unique, the decoder declares error.

Covering Bounds: Since codebooks are generated randomly and independently, in order to be able to find a jointly typical set of sequences U_M^n with the source vector X^n, the following mutual covering bounds need to be satisfied:

$$H(U_{\mathcal{M}}|X) \geq \sum_{M\in\mathcal{M}}(H(U_A)-r_A), \forall \mathcal{M} \subset 2^{2^{[1:L]}}.$$

Packing Bounds: For decoder N, description i is received if $i \in N$. Since binning is done independently and uniformly, in order to find a unique set of jointly typical sequences $(u_M^n)_{N\in M}$, we need to have the following packing bounds:

$$H(U_{\mathcal{L}}) \leq \sum_{M\in\mathcal{L}}(H(U_M) + \sum_{i\in[1:L]} \rho_{M,i} - r_M), \forall \mathcal{L} \subset \mathcal{M}_N,$$

where $\mathcal{M}_N = \{M|N \in M\}$. The resulting RD vector is:

$$R_i = \sum_M \rho_{M,i}$$
$$D_N = E\big\{d_N(h_N(U_{\mathcal{M}_N}, X))\big\}.$$

It can be shown that the codebook C_M is non-redundant if and only if M is a Sperner family in $[1:L]$ and it is not any of $\{\}, \{\{\}\}$ or $2^{[1:L]}$. This observation decreases the number of necessary codebooks significantly. For example in the three descriptions problem $|2^{2^{[1:3]}}| = 256$ whereas there are only 20 Sperner families (i.e. 17 necessary codebooks).

Next we show that the region is achievable using linear codes. Use the same coding scheme as described above, except that we use random linear codes instead of random unstructured codes.

Lemma 16. *The following RD vectors are achievable using linear codes:*

$$H(U_{\mathcal{M}}|X) \geq \sum_{M\in\mathcal{M}} (\log q - r_{o,M}) \tag{5.5}$$

$$H(U_{\mathcal{L}}) \leq \sum_{M\in\mathcal{L}} (\log q + \sum_{j\in[1:L]} \rho_{M,i} - r_{o,M}) \quad \forall \mathcal{L} \subset \mathcal{M}_N \tag{5.6}$$

$$r_{i,M} \leq \log q - H(U_{Q,M}) \quad \forall M \in \mathcal{M} \tag{5.7}$$

$$R_i = \sum_M \rho_{M,i} \tag{5.8}$$

$$D_N = E\big\{d_N(h_N(U_{\mathcal{M}_N}, X))\big\}. \tag{5.9}$$

Remark 18. If we take $r_{i,M} = \log q - H(U_{Q,M})$, the region reduces to the improved CMS region.

5.4 Improvements Using Random Codes for Continuous Sources

In this section we use the new lattice construction to show that the bounds in the previous section are achievable for general continuous sources.

Theorem V.1. *The rate distortion region in the previous section is achievable for continuous sources if we replace entropies with differential entropies.*

Proof. The proof involves two steps, first we approximate the n length vector of X with an n-length vector of the random variable $X_{q,\gamma}$. $X_{q,\gamma}$ is a discrete random variable defined on $\mathcal{X}_{q,\gamma} = [-\gamma\frac{q-1}{2}, \gamma\frac{q-1}{2}] \cap \gamma\mathbb{Z}$. In the next step, we use the fact that for $X_{q,\gamma}$, the definition of typicality in section II reduces to the definition in the discrete case, to show that by the same arguments as in the previous section, the improved CMS with binning is achievable for continuous $\mathbf{U}_M$ and X.

Here we give a more detailed version of the proof. Fix n, q and γ. Fix a probability distribution $P_{\mathbf{U}_M,X}$, where $\mathbf{U}_M$ is the set of random variables from the last section. The underlying alphabet for $\mathbf{U}_M$ is $\mathcal{X}_{q,\gamma}$. The definition of $X_{q,\gamma}$ is as follows:

$$X^n_{q,\gamma} = argmin\big\{d_2(X^n, x^n_{q,\gamma})|x^n_{q,\gamma} \in \mathcal{X}^n_{q,\gamma}\big\},$$

here $d_2(a^n, b^n) = \sum_{i=1}^n (a_i - b_i)^2$. Note that if $\gamma \to 0$ and $\gamma q \to \infty$ then $d_2(X^n, X^n_{q,\gamma}) \to 0$.

Codebook Generation For $M \in 2^{2^{[1:L]}}$ randomly, uniformly and independently generate a nested linear code $(C_{M,o}, C_{M,i})$ over Z_q^n. Let the rates of the nested code be $(r_{M,o}, r_{M,i})$. Fix γ and construct the corresponding nested lattice $\bar{\Lambda}_{M,i},\ \bar{\Lambda}_{M,o}$. For each description, bin the outer code with binning rate $\rho_{M,i}$.

Encoding For a source vector x^n, find the corresponding $x^n_{q,\gamma}$. Find a set of $weak^*$-typical sequences $\mathbf{u}^n_M$. Note since the underlying alphabet for all these variables is $\mathcal{X}_{q,\gamma}$, $weak^*$ typicality is equivalent to strong typicality in the discrete case. Each description carries the bin numbers for the corresponding vectors.

Decoding Upon receiving the bin numbers, decoder N finds a unique set of $weak^*$-typical sequences $\hat{u}_M, N \in M$.

Since $weak^*$ typicality reduces to strong typicality in this case, it is clear that

$X_{q,\gamma}$ can be reconstructed in the decoder with the rate distortions in Section III. If the distortion function at all decoders is bounded and continuous, then we can take $\gamma \to 0$ and $\gamma q \to \infty$ in the same manner as in [57], then since $d_2(X^n, X^n_{q,\gamma}) \to 0$ by continuity of the distortion measures, the RD vectors in Section III are achievable for continuous sources and discrete $\mathbf{U}_M$.

Theorem III.6 in [57] shows that for continuous $\mathbf{U}_M$ we can take $\mathbf{U}_{M,q,\gamma}$ such that as $\gamma \to 0$ and $\gamma q \to \infty$ mutual information terms and distortions containing $\mathbf{U}_{M,q,\gamma}$, converge to the terms containing $\mathbf{U}_M$. We showed that the improved CMS with binning RD region is achievable using linear codes. Since the bounds in that scheme can be written only in terms of mutual informations, the proof is complete.

□

5.5 Achievable RD Using Lattice Quantizers

It was shown in [26] that in the L-descriptions problem with discrete sources and reconstructions, when $L \geq 3$, it is beneficial to use a nested linear code for a pair of random variables. Here we calculate the rate-distortion region resulting from such a scheme. We first calculate the achievable region using linear codes in the discrete case and then show convergence for lattices in the continuous case. Note that since we are planning to use a nested code, the mutual covering bounds are not enough to ensure existence of jointly typical sequences. Hence we need a new covering lemma. Let $(\mathcal{C}, \mathcal{C}')$ be a pair of nested linear codes over $\mathbb{F}_q$ such that their inner codebooks are the same. Since the inner code is the same we need new covering bounds to insure the existence of jointly typical sequences. The following theorem gives the required covering bounds.

Theorem V.2. *Let U and V be the underlying random variables for a pair of nested linear codes $(\mathcal{C}, \mathcal{C}')$. Suppose the two nested codes share an identical inner code. Let*

the joint probability distribution on (U, V, X) *be given by* $P_{U,V,X}$. *For a given* ϵ*-typical sequence* x^n *the probability of finding a jointly* ϵ*-sequence* (u^n, v^n) *in the codebooks* $\mathcal{C}$ *and* $\mathcal{C}'$ *goes to* 1 *as long as the following covering bounds are satisfied:*

$$r_o \geq \log q - H(U|X))$$
$$r'_o \geq \log q - H(V|X))$$
$$r_o + r'_o \geq 2\log q - H(U, V|X)$$
$$r_o + r'_o - r_i \geq \log q - H(iU + V|X), \forall i \in Z_q$$

Where r_o, r'_o *are the rate of outer code of* $\mathcal{C}$ *and* $\mathcal{C}'$, *respectively and* r_i *denotes the rate of the inner code in* $\mathcal{C}$ *and* $\mathcal{C}'$.

Proof. See Appendix D.1. □

Remark 19. Note in the above lemma if we take $r_i = 0$, then the two codes are generated independently, so the covering bounds reduce to mutual covering bounds in the original scheme. However transmitting the linear combination of U and V would require a larger rate since they are coming from two independent codebooks (i.e. the rate of the codebook for $U + iV$ would be equal to $r_U + r_V$ is the packing bounds). On the other hand if we take $r_o = r_i$, then the covering bounds become tighter, since we are using the exact same linear code.

The packing bounds are also affected. We partition the decoders into three sets:

Case 1: Decoder $\underline{s}$ reconstructs both U and V. In this case, the decoder receives one bin number for each of u^n, v^n and $u^n + jv^n$. Using this the decoder can restrict the search over vectors of u^n and v^n to bin sizes $2^{r_U - \rho_U - t\rho_{U+jV}}$ for u^n and $2^{r_V - \rho_V - (1-t)\rho_{U+jV}}$ for v^n where $t \in [0, 1]$ (the choice of t is arbitrary since the bin number for $U + jV$

can be interpreted as either information about U or V). So in the packing bounds ρ_U is replaced with $\rho_U + t\rho_{U+jV}$ and ρ_V is replaced with $\rho_V + (1-t)\rho_{U+jV}$,
Case 2: The decoder only reconstructs U (or V), in which case reconstructing $U+jV$ is equivalent to reconstructing both U and V. So we write the packing bounds as if U and V are reconstructed at the decoder and V is sent with binning rate ρ_{U+jV}.
Case 3: The decoder does not reconstruct U or V. In this case in the packing bound, $U+jV$ is treated as the underlying random variable for a codebook of rate $r_o + r'_o - r_i$ and bin rate ρ_{U+jV} and the mutual packing bounds are written for the decoder.

Based on the arguments in the previous section we only need to show that the bounds in this scheme can also be written in terms of mutual informations. First note that after the Fourier-Motzkin elimination the $\log q$ terms would cancel. Lastly the term $H(U+jV|X)$ differs with $H(U|X)$ only in mutual information terms:

$$
\begin{aligned}
&H(U{+}jV|X) - H(U|X) \\
&= H(U{+}jV|X) - H(U,V|X) + H(V|X,U) \\
&= H(U{+}jV|X) - H(U{+}jV,V|X) + H(V|X,U) \\
&= -H(V|X,U{+}jV) + H(V|X,U) \\
&= I(U{+}jV;V|X) - I(U;V|X)
\end{aligned}
$$

So, the terms in the covering bound differ with the one in the improved CMS only in mutual information quantities. Hence, the region is achievable for continuous sources.

CHAPTER VI

On the Error Exponent of MAC with Noiseless Feedback

Many existing communication systems with feedback (such as ARQ) have variable length. Therefore, in the analysis of fundamental limits for channels with feedback, it is more relevant to allow codes whose length can depend on the channel behavior. In the regime of asymptotically large average block-length, the error exponent, defined as the exponential rate of decay of the probability of error with respect to the average block-length, has been an important performance measure for variable-length codes with feedback.

In this Chapter, we study the error exponent of discrete memoryless MAC with noiseless feedback. In particular, we derive an upper-bound and a lower-bound. We make a connection between this problem and the problem of *sequential hypothesis testing*. We use the tools from *dynamic programming* and Burnashev's techniques for the PtP case to derive the bounds on the error exponent of MAC-FB. The bounds have a similar expression. In this setting, the upper bound is described below

$$E_u(R_1, R_2) = (1 - \frac{||\underline{R}||}{C(\theta_R)})D_u \tag{6.1}$$

where $(||\underline{R}||, \theta_R)$ denote the polar coordinate of (R_1, R_2) in $\mathbb{R}^2$. Also, $C(\theta_R)$ is the

point of the capacity frontier at the angle determined by $\underline{R}$. The lower-bound is the same as E_u but with different constant D_l. The constants D_l and D_u are determined by the relative entropy between the conditional output distributions. An interesting observation is that the bounds increase linearly with respect to a specific Euclidean distance measure defined between the transmission rate pair and the capacity boundary. The lower and upper bounds match for a class of MACs.

6.1 Problem Formulation and Definitions

Consider a discrete memoryless MAC with input alphabets $\mathcal{X}_1, \mathcal{X}_2$, and output alphabet $\mathcal{Y}$. The channel conditional probability distribution is denoted by $Q(y|x_1, x_2)$ for all $(y, x_1, x_2) \in \mathcal{Y} \times \mathcal{X}_1 \times \mathcal{X}_2$. Such setup is denoted by $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}, Q)$. Let y^t and x_i^t, $i = 1, 2$, be the channel output and the inputs sequences after t uses of the channel, respectively. Then, the following condition is satisfied:

$$P(y_t|y^{t-1}, x_1^{t-1}, x_2^{t-1}) = Q(y_t|x_{1t}, x_{2t}). \tag{6.2}$$

We assume that the output of the channel as a feedback is available at the encoders with one unit of delay.

Definition 39. *An (M_1, M_2, N)- variable-length code (VLC) for a MAC $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}, Q)$ with feedback is defined by*

- *A pair of messages W_1, W_2 selected randomly with uniform distribution from $\{1, 2, \ldots, M_i\}, i = 1, 2$.*
- *Two sequences of encoding functions*

$$e_{i,t} : \{1, 2, \ldots, M_i\} \times \mathcal{Y}^{t-1} \to \mathcal{X}_i, \quad t \in \mathbb{N},\ i = 1, 2,$$

one for each transmitter.

- *A sequence of decoding functions*

$$d_t : \mathcal{Y}^t \to \{1, 2, ..., M_1\} \times \{1, 2, ..., M_2\}, \quad t \in \mathbb{N}.$$

- *A stopping time T with respect to (w.r.t) the filtration $\mathcal{F}_t$ defined as the σ-algebra of Y^t for $t \in \mathbb{N}$. Furthermore, it is assumed that T satisfies $\mathbb{E}[T] \leq N$.*

For each $i = 1, 2$, given a message W_i, the tth output of Transmitter i is denoted by $X_{i,t} = e_{i,t}(W_i, Y^{t-1})$.

Let $(\hat{W}_{1,t}, \hat{W}_{2,t}) = d_t(Y^t)$. Then, the decoded messages at the decoder are denoted by $\hat{W}_1 = \hat{W}_{1,T}$, and $\hat{W}_2 = \hat{W}_{2,T}$. In what follows, for any (M_1, M_2, N) VLC, we define average rate-pair, error probability, and error exponent. Average rates for an (M_1, M_2, N) VLC are defined as

$$R_i \triangleq \frac{\log_2 M_i}{\mathbb{E}[T]}, \quad i = 1, 2.$$

The probability of error is defined as

$$P_e = P\left((\hat{W}_1, \hat{W}_2) \neq (W_1, W_2)\right).$$

The error exponent of a VLC with probability of error P_e and stopping time T is defined as $E \triangleq -\frac{\log_2 P_e}{\mathbb{E}[T]}$.

Definition 40. *A reliability function $E(R_1, R_2)$ is said to be achievable for a given MAC, if for any $R_1, R_2 > 0$ and $\epsilon > 0$ there exists an (M_1, M_2, N)-VLC such that*

$$-\frac{\log_2 P_e}{N} \geq E(R_1, R_2) - \epsilon, \text{ and } \frac{\log_2 M_i}{N} \geq R_i - \epsilon,$$

where $i = 1, 2$, and P_e is the error probability of the VLC.

Definition 41. *The reliability function of a MAC with feedback is defined as the supremum of all achievable reliability functions* $E(R_1, R_2)$.

6.1.1 The Feedback-Capacity Region of MAC

We summarize Kramer's results presented in [47] for the feedback capacity of MAC. We use *directed information* and *conditional directed information* as defined in [47]. The normalized directed information from a sequence $\mathbf{X}^n$ to a sequence $\mathbf{Y}^n$ when causally conditioned on $\mathbf{Z}^n$ is denoted by

$$I_n(X \to Y||Z) = \frac{1}{n} I(\mathbf{X}^n \to \mathbf{Y}^n||\mathbf{Z}^n). \tag{6.3}$$

The feedback-capacity region of a discrete memoryless MAC with feedback $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}, Q)$ is denoted by $\mathcal{C}$, and is the closure of the set of all rate-pairs (R_1, R_2) such that

$$\begin{aligned} R_1 &\leq I_L(X_1 \to Y||X_2) \\ R_2 &\leq I_L(X_2 \to Y||X_1) \\ R_1 + R_2 &\leq I_L(X_1 X_2 \to Y), \end{aligned}$$

where L is a positive integer, and $P_{X_1^L X_2^L Y^L}$ factors as

$$\prod_{l=1}^{L} P_{1,l}(x_{1l}|x_1^{l-1}y^{l-1}) P_{2,l}(x_{2l}|x_2^{l-1}y^{l-1}) Q(y_l|x_{1,l}x_{2,l}). \tag{6.4}$$

Definition 42. *Let* $\lambda_1, \lambda_2, \lambda_3 \geq 0$, *and* $\lambda_1 + \lambda_2 + \lambda_3 = 1$. *Define*

$$\begin{aligned} C_{\underline{\lambda}} = \sup_{L \in \mathbb{N}} \sup_{P_{X_1^L X_2^L Y^L}} & \lambda_1 I_L(X_1 \to Y|X_2) + \lambda_2 I_L(X_2 \to Y|X_1) \\ & + \lambda_3 I_L(X_1 X_2 \to Y), \end{aligned}$$

where $P_{X_1^L X_2^L Y^L}$ *factors as in* (6.4).

Fact 1. The feedback-capacity of a discrete memoryless MAC with feedback is the same as the closure of the set of rate-pairs (R_1, R_2) such that the inequality

$$\lambda_1 R_1 + \lambda_2 R_2 + \lambda_3 (R_1 + R_2) \leq C_{\underline{\lambda}}$$

holds for all $\lambda_1, \lambda_2, \lambda_3 \geq 0$, with $\lambda_1 + \lambda_2 + \lambda_3 = 1$.

6.1.2 Notational Conventions

For more convenience, we denote a rate-pair (R_1, R_2) by (R_1, R_2, R_3), where $R_3 = R_1 + R_2$. For a $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}, Q)$ MAC we use the following notational convenience

$$I_L^1 \triangleq I_L(X_1 \to Y || X_2), \tag{6.5}$$

$$I_L^2 \triangleq I_L(X_2 \to Y || X_1), \tag{6.6}$$

$$I_L^3 \triangleq I_L(X_1 X_2 \to Y). \tag{6.7}$$

The Kullback–Leibler divergence for the MAC with transition probability matrix Q is defined as

$$D_Q(x_1, x_2 || z_1, z_2) = \sum_{y \in \mathcal{Y}} Q(y|x_1, x_2) \log_2 \frac{Q(y|x_1, x_2)}{Q(y|z_1, z_2)},$$

where $(x_1, x_2), (z_1, z_2) \in \mathcal{X}_1 \times \mathcal{X}_2$. For notational convenience we denote

$$D_1(x_1, x_2 || z_1, z_2) = D_Q(x_1, x_2 || z_1, x_2)$$

$$D_2(x_1, x_2 || z_1, z_2) = D_Q(x_1, x_2 || x_1, z_2)$$

$$D_3(x_1, x_2 || z_1, z_2) = D_Q(x_1, x_2 || z_1, z_2).$$

6.2 A Lower-Bound for the Reliability Function

We build upon Yamamoto-Itoh transmission scheme for PtP channel coding with feedback [88]. The scheme sends the messages W_1, W_2 through blocks of length n. The transmission process is performed in two stages: 1) The "data transmission" stage taking up to $n(1-\gamma)$ channel uses, 2) The "confirmation" stage taking up to $n\gamma$ channel uses, where γ is a design parameter taking values from $[0, 1]$.

Stage 1 For the first stage, we use any coding scheme that achieves the feedback-capacity of the MAC. The length of this coding scheme is at most $n(1-\gamma)$. Let $\hat{W}_1, \hat{W}_2$ denote the decoder's estimation of the messages at the end of the first stage. Define the following random variables:

$$H_i = 1\{\hat{W}_i \neq W_i\}, \quad i = 1, 2.$$

Because of the feedback, $\hat{W}_1$ and $\hat{W}_2$ are known at each transmitter. Therefore, at the end of the first stage, transmitter i has access to $W_i, \hat{W}_1, \hat{W}_2$, and H_i, where $i = 1, 2$.

Stage 2 The objective of the second stage is to inform the receiver whether the hypothesis $\Theta_0 : (\hat{W}_1, \hat{W}_2) = (W_1, W_2)$ or $\Theta_1 : (\hat{W}_1, \hat{W}_2) \neq (W_1, W_2)$ is correct. For that, each transmitter employs a code of size two and length γn. The codewords of such codebooks are denoted by two pairs of sequences $(\underline{x_1}(0), \underline{x_2}(0))$ and $(\underline{x_1}(1), \underline{x_2}(1))$ each with elements belonging to $\mathcal{X}_1 \times \mathcal{X}_2$. Fix a joint-type P_n defined over the set $\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_1 \times \mathcal{X}_2$ and for sequences of length γn. The sequences $(\underline{x_1}(0), \underline{x_2}(0), \underline{x_1}(1), \underline{x_2}(1))$ are selected randomly among all the sequences with joint-type P_n. During this stage and given H_1, Transmitter 1 sends $\underline{x_1}(H_1)$. Similarly, Transmitter 2 sends $\underline{x_2}(H_2)$.

Decoding Upon receiving the channel output, the receiver estimates H_1, H_2. Denote this estimation by $\hat{H}_1, \hat{H}_2$. If $(\hat{H}_1, \hat{H}_2) = (0, 0)$, then the hypothesis $\hat{\Theta} = \Theta_0$ is

declared. Otherwise, $\hat{\Theta} = \Theta_1$ is declared. Because of the feedback, $\hat{\Theta}$ is also available at each encoders. If $\hat{\Theta} = \Theta_0$, then transmission stops and a new data packet is transmitted at the next block. Otherwise, the message is transmitted again at the next block. The process continues until $\hat{\Theta} = \Theta_0$ occurs.

The confirmation stage in the proposed scheme can be viewed as a decentralized binary hypothesis problem in which a binary hypothesis $\{\Theta_0, \Theta_1\}$ is observed partially by two distributed agents and the objective is to convey the true hypothesis to a central receiver. This problem is qualitatively different from the sequential binary hypothesis testing problem as identified in [89] for PtP channel. Note also that in the confirmation stage we use a different coding strategy than the one used in Yamamoto-Itoh scheme [88]. Here, all four codewords have a joint-type P_n. It can be shown that repetition codes, and more generally, constant composition codes are strictly suboptimal in this problem.

Theorem VI.1. *The following is a lower-bound for the reliability function of any discrete memoryless MAC:*

$$E_l(R_1, R_2) = \min_{\substack{\lambda_1,\lambda_2,\lambda_3 \geq 0 \\ \lambda_1+\lambda_2+\lambda_3=1}} D_l(1 - \frac{\sum_i \lambda_i R_i}{C_{\underline{\lambda}}}), \tag{6.8}$$

where,

$$D_l \triangleq \sup_{P_{X_1X_2Z_1Z_2}} \min_{i=1,2,3} \mathbb{E}\left[D_i(X_1, X_2||Z_1, Z_2)\right], \tag{6.9}$$

and the supremum is taken over all probability distributions $P_{X_1X_2Z_1Z_2}$ defined over $\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_1 \times \mathcal{X}_2$.

Proof. The proof is given in Appendix E.1. □

6.3 An Upper-bound for the Reliability Function

In this part of the paper, we establish an upper-bound for the reliability function of any discrete memoryless MAC. Define

$$D_i \triangleq \max_{\substack{x_1,z_1\in\mathcal{X}_1,\\ x_2,z_2\in\mathcal{X}_2}} D_i(x_1,x_2||z_1,z_2), \quad i=1,2,3. \tag{6.10}$$

Theorem VI.2 (Upper-bound)**.** *For any (N,M_1,M_2) VLC with probability of error P_e, and any $\epsilon>0$, there exists a function δ such that the following is an upper-bound for the reliability function of the VLC*

$$\begin{aligned} E(R_1,R_2) \leq & \min_{\substack{\lambda_1,\lambda_2,\lambda_3\geq 0\\ \lambda_1+\lambda_2+\lambda_3=1}} \min_{j\in\{1,2,3\}} D_j\left(1-\frac{\lambda_j R_j}{C_\lambda}\right) \\ & + \delta(P_e, M_1M_2, \epsilon), \end{aligned} \tag{6.11}$$

where (R_1,R_2) is the rate pair of the VLC and δ satisfies

$$\lim_{\epsilon\to 0}\lim_{P_e\to 0}\lim_{M_1M_2\to\infty} \delta(P_e,M_1M_2,\epsilon)=0.$$

Corollary 3. *From Theorem VI.2, the following is an upper-bound for the error exponent of a MAC:*

$$E_u(R_1,R_2) = \min_{\substack{\lambda_1,\lambda_2,\lambda_3\geq 0\\ \lambda_1+\lambda_2+\lambda_3=1}} D_u\left(1-\frac{\sum_{i=1}^3\lambda_i R_i}{C_\lambda}\right)+\delta,$$

where $D_u=\max\{D_1,D_2,D_3\}$, and δ is as in Theorem VI.2.

Proof. The proof is given in Appendix E.5. □

6.3.1 Proof of the Upper-Bound

Consider any (N, M_1, M_2) VLC with probability of error P_e, and stopping time T. Suppose the message at Encoder 2, W_2, is made available to all terminals. For the new setup, as W_2 is available at the Decoder, the average probability of error is $P_e^1 \triangleq P\{\hat{W}_1 \neq W_1\}$. Note that $P_e \geq P_e^1$. We refer to such setup as W_2-assisted MAC. For a maximum a *posteriori* decoder, after n uses of the channel and assuming the realization $Y^n = y^n$ and $W_2 = w_2$, define

$$T_1^\delta \triangleq \inf\big\{n : \max_{1\leq i\leq M_1} P(W_1 = i|y^n, w_2) \geq 1-\delta\big\},$$

where $\delta > 0$ is a fixed real number. Also, let $\tau_1 \triangleq \min\{T, T_1^\delta\}$. Note that τ_1 is a stopping time w.r.t the filtration $\{\mathcal{F}_{W_2} \times \mathcal{F}_t\}_{t>0}$. The following lemma provides a lower-bound on the probability of error for such setup.

Lemma 17. *The probability of error, P_e, for a hypothesis testing over a W_2-assisted MAC and variable length codes satisfies the following inequality*

$$P_e \geq \frac{\min\{P(H), P(H^c)\}}{4} e^{-D_1 \mathbb{E}[T]},$$

where $\{H, H^c\}$ are the two hypothesizes and T is the stopping time of the variable length code.

Lemma 18. *For a given MAC with finite D_3 the following holds*

$$\zeta p(w_1, w_2|y^{n-1}) \leq p(w_1, w_2|y^n) \leq \frac{p(w_1, w_2|y^{n-1})}{\zeta},$$

where $\zeta \triangleq \min_{x_1,x_2,y} Q(y|x_1, x_2)$.

The above lemmas are extensions of Lemma 1 and Proposition 2 in [89] for MAC. The proofs follow from similar arguments and are omitted.

Lemma 19. *Given a MAC with $D_3 < \infty$, and for any (N, M_1, M_2) VLC with probability of error P_e the following holds*

$$P_e \geq \frac{\zeta\delta}{4} e^{-D_1 \mathbb{E}[T-\tau_1]}, \tag{6.12}$$

where $\zeta \triangleq \min_{x_1,x_2,y} Q(y|x_1, x_2)$.

Proof. Suppose the VLC is used for a W_2-assisted MAC. As discussed before, $P_e \geq P_e^1$. We modify the encoding and the decoding functions of the VLC used for the MAC. Let $\mathcal{H}_1 \subseteq \mathcal{M}_1$ be a subset of the message set $\mathcal{M}_1$. The subset $\mathcal{H}_1$ is to be determined at time τ_1. The new decoding function, at time T, decides whether the message belongs to $\mathcal{H}_1$. The new encoding functions are the same as the original one until the time τ_1. Then, after τ_1, the transmitters perform a VLC to resolve the binary hypothesis $\{W_1 \in \mathcal{H}_1\}$ and $\{W_1 \notin \mathcal{H}_1\}$. This hypothesis problem is performed from τ_1 to T. With these modifications, the error probability of this binary hypothesis problem is a lower-bound on P_e. In what follows, we present a construction for $\mathcal{H}_1$. Then, we apply Lemma 17 to complete the proof.

Let $P_e^1(y^n, w_2) \triangleq 1 - \max_{1\leq i\leq M_1} P(W_1 = i|y^n, w_2)$. The quantity $P_e^1(y^{\tau_1}, w_2)$ can be calculated at all terminals. By definition, at time $\tau_1 - 1$, the inequality $P(W_1 = i|Y^{\tau_1-1}, W_2) < 1 - \delta$ holds almost surely for all $i \in [1 : M_1]$. This implies that $P_e^1(Y^{\tau_1-1}, W_2) > \delta$. Hence, by Lemma 18 at time τ_1 the inequality $P_e^1(Y^{\tau_1}, W_2) \geq \zeta\delta$ holds almost surely. We consider two cases $P_e^1(y^{\tau_1}, w_2) \leq \delta$ and $P_e^1(y^{\tau_1}, w_2) > \delta$, where δ is the constant used in the definition of T_1^δ. For the first case, $\mathcal{H}_1$ is the set consisting of the message with the highest a *posteriori* probability. Since $P_e^1(y^{\tau_1}, w_2) \leq \delta$, then $P(\mathcal{H}_1) \geq 1 - \delta$. In addition, as $P_e^1(y^{\tau_1}, w_2) \geq \zeta\delta$, then $P(\mathcal{H}_1^c) > \zeta\delta$. For the second case, set $\mathcal{H}_1$ to be a set of messages such that $P(\mathcal{H}_1) > \delta/2$ and $P(\mathcal{H}_1) < 1-\delta$. Such set exists, since $P(W_1 = i|Y^{\tau-1}, W_2) < 1 - \delta$ holds for all messages $i \in [1 : M_1]$.

Note that by the above construction, for each case, $P(\mathcal{H}_1) \in [\zeta\delta, 1 - \zeta\delta]$. Thus,

from Lemma 17 and the argument above, the inequality

$$P\{\hat{W}_1 \neq W_1 | Y^\tau, W_2\} \geq \frac{\zeta\delta}{4} e^{-D_1 \mathbb{E}[T-\tau|Y^\tau, W_2]}$$

holds almost surely. Next, we take the expectation of the above expression. The lemma follows by the convexity of e^{-x} and Jensen's inequality.

□

Next, we apply the same argument for the case where W_1 is available at all the terminals. For that define

$$T_2^\delta \triangleq \inf\big\{n : \max_{1\leq j\leq M_2} P(W_2 = j|y^n, w_1) \geq 1-\delta\big\},$$

and let $\tau_2 \triangleq \min\{T, T_2^\delta\}$. By symmetry, Lemma 19 holds for this case and we obtain

$$P_e \geq \frac{\zeta\delta}{4} e^{-D_2 \mathbb{E}[T-\tau_2]}. \tag{6.13}$$

Next, define the following stopping times:

$$T_3^\delta \triangleq \inf\big\{n : \max_{i,j} P(W_1 = i, W_2 = j|y^n) \geq 1-\delta\big\}.$$

Also, let $\tau_3 = \min\{T, T_3^\delta\}$. using a similar argument as in the above, we can show that

$$P_e \geq \frac{\zeta\delta}{4} e^{-D_3 \mathbb{E}[T-\tau_3]}. \tag{6.14}$$

For that, after time τ_3, we formulate a binary hypothesis problem in which the transmitters determine whether $(W_1, W_2) \in \mathcal{H}_3$ or not. Here, $\mathcal{H}_3$ is a subset which is constructed using a similar method as for $\mathcal{H}_1$ in the proof of Lemma 19. We further allow the transmitters to communicate with each other after τ_3. The maximum of

the right-hand sides of (6.12), (6.13) and (6.14) gives a lower-bound on P_e. The lower-bound depends on the expectation of the stopping times $\tau_i, i = 1, 2, 3$. In what follows, we provide a lower-bound on $\mathbb{E}[\tau_i]$. Define the following random processes.

$$H_t^1 \triangleq H(W_1|\ \mathcal{F}_{W_2} \times \mathcal{F}_t),$$
$$H_t^2 \triangleq H(W_2|\ \mathcal{F}_{W_1} \times \mathcal{F}_t),$$
$$H_t^3 \triangleq H(W_1, W_2|\ \mathcal{F}_t),$$

Lemma 20. *Given a (M_1, M_2, N)-VLC, for any $\epsilon > 0$ there exist L and a probability distribution $P_{X_1^L X_2^L Y^L}$ that factors as in (6.4) such that the following inequalities hold almost surely for $1 \leq t \leq N$*

$$\mathbb{E}[H_{t+1}^1 - H_t^1 | \mathcal{F}_{W_2} \times \mathcal{F}_t] \geq -(I_L^1 + \epsilon),$$
$$\mathbb{E}[H_{t+1}^2 - H_t^2 | \mathcal{F}_{W_1} \times \mathcal{F}_t] \geq -(I_L^2 + \epsilon),$$
$$\mathbb{E}[H_{t+1}^3 - H_t^3 | \mathcal{F}_t] \geq -(I_L^3 + \epsilon).$$

where $i = 1, 2, 3$, and I_L^i is defined as in (6.5)-(6.7).

Proof. The proof is provided in Appendix E.2. □

We need the following lemma to proceed. The lemma is a result of Lemma 4 in [42], and we omit its proof.

Lemma 21. *For any $t \geq 1$ and $i = 1, 2, 3$, the following inequality holds almost surely w.r.t $\mathcal{F}_{W_1} \times \mathcal{F}_{W_2} \times \mathcal{F}_t$*

$$\log H_t^i - \log H_{t+1}^i \leq \max_{\substack{j,l \in [1:M_1] \\ k,m \in [1:M_2]}} \max_{y \in \mathcal{Y}} \frac{\hat{Q}_{j,k}(y)}{\hat{Q}_{l,m}(y)}.$$

From Lemma 20 and the fact that $H_t^i \leq \log_2 M_i < \infty$, the processes $\{H_t^i + (I_L^1 + \epsilon)t\}_{t>0}$ are submartingales for $i = 1, 2, 3$. In addition, from Lemma 21 and the

inequalities $\mathbb{E}[\tau_i] \leq \mathbb{E}[T] \leq N < \infty$, we can apply Doob's Optional Stopping Theorem for each submartingale $\{H_t^i + (I_L^1 + \epsilon)t\}_{t>0}$. Then, we get:

$$\log M_i \leq \mathbb{E}[H_{\tau_i}^i] + \mathbb{E}[\tau_i](I_L^i + \epsilon) \tag{6.15}$$

where $M_3 = M_1 M_2$.

Lemma 22. *The following inequality holds for each* $i = 1, 2, 3$

$$\mathbb{E}[H_{\tau_i}^i] \leq h_b(\delta) + (\delta + \frac{P_e}{\delta}) \log_2 M_i.$$

Proof. We prove the lemma for the case $i = 1$. The proof for $i = 2, 3$ follows from a similar argument. For $i = 1$, we obtain

$$\begin{aligned}
\mathbb{E}[H_{\tau_1}^1] &= P\{P_e(Y^{\tau_1}, W_2) > \delta\}\mathbb{E}[H_{\tau_1}^i | P_e(Y^{\tau_1}, W_2) > \delta] \\
&\quad + P\{P_e(Y^{\tau_1}, W_2) \leq \delta\}\mathbb{E}[H_{\tau_1}^1 | P_e(Y^{\tau_1}, W_2) \leq \delta] \qquad (6.16) \\
&\leq P\{P_e(Y^{\tau_1}, W_2) > \delta\} \log_2 M_1 + P\{P_e(Y^{\tau_1}, W_2) \leq \delta\}\mathbb{E}[H_{\tau_1}^i | P_e(Y^{\tau_1}, W_2) \leq \delta]. \qquad (6.17)
\end{aligned}$$

Note that the event $\{P_e(Y^{\tau_1}, W_2) > \delta\}$ implies that $\tau_1 = T$, and $P_e(y^{\tau_1}, W_2) > \delta$ for all $0 \leq n \leq T$. Hence, this event is included in the event $\{P_e(Y^T, W_2) > \delta\}$. Thus, applying Markov inequality gives

$$P\{P_e(Y^{\tau_1}, W_2) > \delta\} \leq P\{P_e(Y^T, W_2) > \delta\} \leq \frac{P_e}{\delta}.$$

As a result of the above argument, the right-hand side of (6.17) does not exceed the following

$$\frac{P_e}{\delta} \log_2 M_1 + \mathbb{E}[H_{\tau_1}^1 | P_e(Y^{\tau_1}, W_2) \leq \delta].$$

From Fano's inequality we obtain

$$\mathbb{E}[H^1_{\tau_1}|P_e(Y^{\tau_1}, W_2) \leq \delta] \leq h_b(\delta) + \delta \log_2 M_1.$$

The proof is complete from the above inequality. □

As a result of the above lemma and (6.15), the inequality $\mathbb{E}[\tau_i] \geq \frac{\log M_i}{I^i_L + \epsilon} - \frac{h_b(\delta)}{I^i_L + \epsilon}$ holds. Finally, combining this inequality with (6.12)-(6.14) completes the proof of the theorem.

6.3.2 An Alternative Proof for the Upper-Bound

In this part of the paper, we provide a series of Lemmas that are used to prove the Theorem. Define the following random processes.

Lemma 23. *For an (M_1, M_2, N)-VLC with probability of error P_e the following inequality holds*

$$\mathbb{E}[H^i_T] \leq h_b(P_e) + P_e \log_2(M_1 M_2 - 1), \quad \textit{for} \quad i = 1, 2, 3.$$

Proof. The proof follows from Fano's Lemma as in [42]. □

Lemma 24. *There exists $\epsilon > 0$ such that, if $H^i_t \leq \epsilon$, then*

$$\mathbb{E}[\log H^1_{t+1} - \log H^1_t | \mathcal{F}_{W_2} \times \mathcal{F}_t] \geq -(D_1 + \epsilon),$$
$$\mathbb{E}[\log H^2_{t+1} - \log H^2_t | \mathcal{F}_{W_1} \times \mathcal{F}_t] \geq -(D_2 + \epsilon),$$
$$\mathbb{E}[\log H^3_{t+1} - \log H^3_t | \mathcal{F}_t] \geq -(D_3 + \epsilon)$$

holds almost surely, where $D_i, i = 1, 2, 3$ are defined in (6.10).

Proof. The proof is given in Appendix E.3. □

Lemma 25. *For $i = 1, 2, 3$, define random process $\{Z_t^{(i)}\}_{t\geq 1}$ as*

$$Z_t^{(i)} = \left(\frac{\log H_t^i - \log \epsilon}{D_i} + t + f_i(\log \frac{H_t^i}{\epsilon})\right) \mathbb{1}\{H_t^i \leq \epsilon\} + \left(\frac{H_t^i - \epsilon}{I_L^i} + t\right) \mathbb{1}\{H_t^i \geq \epsilon\} \tag{6.18}$$

where the function f_i is defined as $f_i(y) = \frac{1-e^{-\mu_i y}}{D_i \mu_i}$. Then, there exists $\mu_i > 0$ such that $Z_t^{(i)}$ is a submartingale w.r.t $\mathcal{F}_{W_1} \times \mathcal{F}_{W_2} \times \mathcal{F}_t$.

Outline of the proof. Suppose $W_2 = m$ for some $m \in [1 : M_2]$. Given this event and using the same argument as in the proof of Theorem 1 in [42] we can show that $Z_t^{(i)}|W_2 = m$ is a submartingale for all m. More precisely, the inequality

$$\mathbb{E}\{Z_t^{(i)} - Z_{t+1}^{(i)}|\mathcal{F}_{W_1} \times \mathcal{F}_{W_2}\} \leq 0,$$

holds almost surely w.r.t $\mathcal{F}_{W_1} \times \mathcal{F}_{W_2}$. Taking the expectation of the both sides in the above inequality gives

$$\mathbb{E}\{Z_t^{(i)} - Z_{t+1}^{(i)}\} \leq 0, \quad \forall t \geq 0,\ i = 1, 2, 3.$$

Thus, $Z_t^{(i)}$ is a submartingale for $i = 1, 2, 3$ and w.r.t $\mathcal{F}_{W_1} \times \mathcal{F}_{W_2} \times \mathcal{F}_t$. □

Corollary 4. *Suppose $\alpha_1, \alpha_2, \alpha_3$ are non-negative numbers such that $\alpha_1 + \alpha_2 + \alpha_3 = 1$. Define $Z_t = \alpha_1 Z_t^{(1)} + \alpha_2 Z_t^{(2)} + \alpha_3 Z_t^{(3)}$. Then, Z_t is a submartingale w.r.t $\mathcal{F}_{W_1} \times \mathcal{F}_{W_2} \times \mathcal{F}_t$.*

The Theorem follows from the above lemma, and the proof is given in Appendix E.4.

6.4 The Shape of the Lower and Upper Bounds

In this Section, we point out a few remarks on $E_u(R_1, R_2)$ and the lower-bound $E_l(R_1, R_2)$ defined in Theorem VI.1. Furthermore, we provide an alternative representation for the bounds and show that the lower and upper-bounds match for a class of MACs.

We first compare the lower bound in (6.8) and the upper-bound in Corollary 3. For a given arbitrary rate pair (R_1, R_2) inside the feedback-capacity of a given MAC, consider a sequence of VLCs with rates (R_1, R_2) and with average probability of error approaching zero. Then, the following holds:

$$\lim_{\epsilon\to 0}\lim_{P_e\to 0}\lim_{M_1 M_2\to\infty}\frac{E_u(R_1,R_2)}{E_l(R_1,R_2)}=\frac{D_u}{D_l}$$

As a result of the above remark, it is concluded that for small enough probability of error, the bounds are different only in the constants D_u and D_l.

Next, provide an alternative representation for the lower/upper-bound. For that, suppose (R_1, R_2) is a point inside the capacity region $\mathcal{C}$. By $(||\underline{R}||, \theta_R)$ denote the polar coordinate of (R_1, R_2) in $\mathbb{R}^2$. It is shown in the following Remark that the optimum $\underline{\lambda}$ in E_u and E_l is independent of the Euclidean norm of (R_1, R_2), i.e., $\|\underline{R}\|$.

Remark 20. Given an arbitrary $\alpha > 0$ and a rate pair (R_1, R_2) in the capacity region, the optimum $\underline{\lambda}$ for $E_l(R_1, R_2)$ is the same as the one for $E_l(\alpha R_1, \alpha R_2)$.

Proof. Note that one can write $E_l(R_1, R_2)$ as

$$\begin{aligned}E_l(R_1,R_2)&=D_l\left(1-\max_{\substack{\lambda_1,\lambda_2,\lambda_3\geq 0\\ \lambda_1+\lambda_2+\lambda_3=1}}\frac{\sum_{i=1}^3\lambda_i R_i}{C_\lambda}\right),\\&=D_l\left(1-\frac{\sum_{i=1}^3\lambda_i^* R_i}{C_{\lambda^*}}\right),\end{aligned}$$

where $\underline{\lambda}^*$ is the optimum $\underline{\lambda}$ for E_l. Next, replace (R_1, R_2) with $(\alpha R_1, \alpha R_2)$ for some

constant $\alpha > 0$. Then, we obtain

$$E_l(\alpha R_1, \alpha R_2) = D_l \left(1 - \alpha \max_{\substack{\lambda_1,\lambda_2,\lambda_3 \geq 0 \\ \lambda_1+\lambda_2+\lambda_3=1}} \frac{\sum_{i=1}^{3} \lambda_i R_i}{C_\lambda} \right),$$
$$\stackrel{(a)}{=} D_l \left(1 - \alpha \frac{\sum_{i=1}^{3} \lambda_i^* R_i}{C_{\lambda^*}} \right),$$

where (a) follows as the objective function for the maximization is the same as the one in $E_l(R_1, R_2)$. This implies that there is an identical $\underline{\lambda}^*$ which optimizes the expression in $E_l(R_1, R_2)$ and $E_l(\alpha R_1, \alpha R_2)$. □

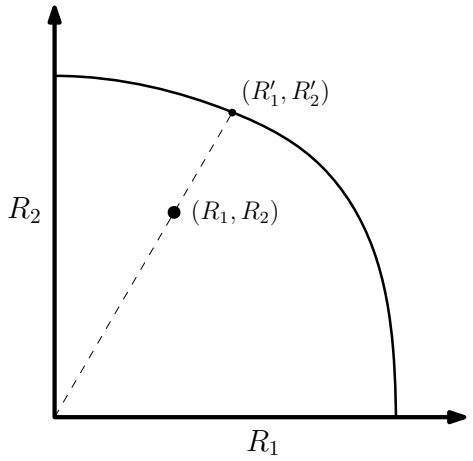

Figure 6.1: Given a rate pair (R_1, R_2) which is inside the capacity region, consider the line passing (R_1, R_2) and the origin. Then, (R_1', R_2') is the point of intersection of this line with the boundary of the capacity region.

Now, consider the line passing (R_1, R_2) and the origin. Let (R_1', R_2') denote the point of intersection of this line with the boundary of the capacity region. Fig. 6.1 shows how (R_1', R_2') is determined. Since, $R_i' = \alpha R_i, i = 1, 2$ for some $\alpha > 0$, then the optimum $\underline{\lambda}$ in $E_l(R_1', R_2')$ is the same as the one in $E_l(R_1, R_2)$. Therefore, from this

argument and the fact that $R_i = \frac{R'_i}{\alpha}, i = 1, 2$, we can rewrite $E_l(R_1, R_2)$ as

$$E_l(R_1, R_2) = \min_{\substack{\lambda_1,\lambda_2,\lambda_3 \geq 0 \\ \lambda_1+\lambda_2+\lambda_3=1}} D_l\left(1 - \frac{1}{\alpha}\frac{\sum_{i=1}^{3}\lambda_i R'_i}{C_\lambda}\right),$$
$$\stackrel{(a)}{=} D_l\left(1 - \frac{1}{\alpha}\right),$$

where (a) follows, since (R'_1, R'_2) is on the capacity boundary. Note that $\alpha = \frac{\|\underline{R}\|}{\|\underline{R'}\|}$. Therefore, $E_l(R_1, R_2) = D_l\left(1 - \frac{\|\underline{R}\|}{\|\underline{R'}\|}\right)$. Moreover, note that $\|\underline{R'}\|$ depends on (R_1, R_2) only through θ_R; in particular, it equals to $C(\theta_R)$ which is a function of θ_R. With this notation, we can rewrite E_l as

$$E_l(R_1, R_2) = D_l\left(1 - \frac{\|\underline{R}\|}{C(\theta_R)}\right)$$

Using a similar argument for E_u, we have

$$E_u(R_1, R_2) = D_u\left(1 - \frac{\|\underline{R}\|}{C(\theta_R)}\right) + \delta.$$

As a conclusion of the above argument, the lower (upper) bound increases linearly with respect to a specific Euclidean distance measure defined between the transmission rate pair and the capacity boundary. Fig. 6.2 shows the shape of a typical upper (lower) bound as a function of the transmission rate pairs.

6.4.1 On the Tightness of the Bounds on the Error Exponent

In what follows, we provide examples of classes of channels for which the lower and upper bound coincide.

Example 11. Consider a MAC in which the output is (Y_1, Y_2) and the transition probability matrix is described by the product $Q_{Y_1|X_1}Q_{Y_2|X_2}$. This MAC consists of two parallel (independent) point-to-point channels. Suppose, C_1 and C_2 are the

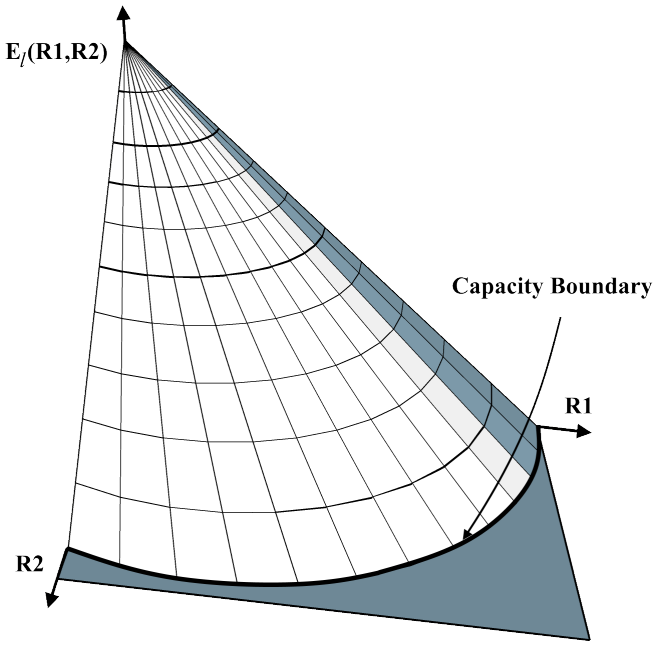

Figure 6.2: The conceptual shape of the lower/upper-bound on the error exponent of a given MAC with respect to the transmission rate pair (R_1, R_2).

capacity of the first and the second parallel channel, respectively. For this MAC, one can use two parallel Yamamoto-Itoh schemes, one for each channel. Based on the results for the point-to-point case, it is not difficult to show that the error exponent for such MAC satisfies

$$E(R_1, R_2) \geq \min\{D_1(1 - \frac{R_1}{C_1}), D_2(1 - \frac{R_2}{C_2})\}, \tag{6.19}$$

where C_1 and C_2 are the point-to-point capacity of the channel corresponding to $Q_{Y_1|X_1}$ and $Q_{Y_2|X_2}$, respectively. Note that this lower-bound is not covered by the proposed coding strategy given in Section 6.2. For such MAC, the upper-bound given in (6.11) is simplified to

$$E(R_1, R_2) \leq \min_{\lambda_1, \lambda_2 \geq 0} \min_{j \in \{1,2\}} D_j \left(1 - \frac{\lambda_j R_j}{\lambda_1 C_1 + \lambda_2 C_2}\right) + \delta.$$

The right-hand side of the above inequality is further upper-bounded by substituting

$(\lambda_1, \lambda_2) = (0, 1)$ or $(\lambda_1, \lambda_2) = (1, 0)$. Therefore, we obtain

$$E(R_1, R_2) \leq \min_{j \in \{1,2\}} D_j \left(1 - \frac{R_j}{C_j}\right) + \delta$$

By letting $\delta \to 0$ as in Theorem VI.2, the above bound can be made arbitrary close to the lower-bound given in (6.19).

Example 12. Consider a MAC with input alphabets $\mathcal{X}_1 = \mathcal{X}_2 = \{0, 1, 2\}$, and output alphabet $\mathcal{Y} = \{0, 1, 2\}$. The transition probability of the channel is described by the following relation:

$$Y = X_1 \oplus_3 X_2 \oplus_3 N_p,$$

where the additions are modulo-3 addition, and N_p is a random variable with $P(N_p = 1) = P(N_p = 2) = p$, and $P(N_p = 0) = 1 - 2p$, where $0 \leq p \leq 1/2$. It can be shown that for this channel $D_l = D_u = (1-3p) \log \frac{1-2p}{p}$. Hence, the upper-bound in Corollary 3 can be made arbitrary close to the lower-bound in Theorem VI.1.

The argument in the above example can be extended to m-ary additive MACs for $m > 2$, where the transition probability of the channel is described by

$$Y = X_1 \oplus_m X_2 \oplus_m N_p,$$

where all the random variables take values from $\mathbb{Z}_m$, and N_p is a random variable with $P(N_p = i) = p$ for any $i \in \mathbb{Z}_m,\ i \neq 0$ and $P(N_p = 0) = 1 - (m-1)p$. It can be shown that for this channel

$$D_l = D_u = (1 - mp) \log \frac{1 - (m-1)p}{p}.$$

APPENDICES

APPENDIX A

Proofs for Chapter II

A.1 Proof of Lemma 1

Proof. Using (2.3) we get $\mathcal{U}_n = \bigotimes_{q\in\mathcal{Q}} A_\epsilon^{(k_{q,n})}(U_q)$, where $k_{q,n} = P_Q(q)k_n$, and the distribution of U_q is the same as the conditional distribution of U given $Q = q$. Using well-known results on the size of ϵ-typical sets we can provide a bound on $|A_\epsilon^{(k_{q,n})}(U_q)|$. More precisely, there exists N_q such that for all $k_{q,n} > cN_q$, we have $|\frac{1}{k_{q,n}} \log_2 |A_\epsilon^{(k_{q,n})}(U_q)| - H(U_q)| \leq 2\epsilon'_q$, where using the same argument as in [3]

$$\epsilon'_q = -\frac{\epsilon}{p^r} \sum_{a\in\mathbb{Z}_{p^r}, P(U_q=a)>0} \log_2 P(U_q = a).$$

Therefore,

$$\begin{aligned}
\frac{1}{k_n} \log_2 |\mathcal{U}_n| &= \frac{1}{k_n} \sum_{q\in\mathcal{Q}} \log_2 |A_\epsilon^{(k_{q,n})}(U_q)| \\
&\leq \sum_{q\in\mathcal{Q}} \frac{k_{q,n}}{k_n} (H(U_q) + 2\epsilon'_q) \\
&\overset{(a)}{=} H(U|Q) + \sum_{q\in\mathcal{Q}} P_Q(q) 2\epsilon'_q \leq H(U|Q) + \epsilon',
\end{aligned}$$

where $\epsilon' \triangleq 2\max_{q\in Q}\epsilon'_q$. Note that (a) holds as $P_Q(q) = k_{q,n}/k_n$. Using a similar argument we can show that $\frac{1}{k_n}\log_2|\mathcal{U}_n| \geq H(U|Q) - \epsilon'$. Finally, by setting $N = \max_q N_q$, and combining the bounds on $\frac{1}{k_n}\log_2|\mathcal{U}_n|$ the proof is completed. □

A.2 Proof of Lemma 2

Proof. For any $\mathbf{u} \in \mathcal{U}_n$, define

$$\theta(\mathbf{u}) \triangleq \sum_{\substack{\mathbf{u}'\in\mathcal{U}_n \\ \mathbf{u}'\neq\mathbf{u}}} \mathbb{1}\{\Phi_n(\mathbf{u}') = \Phi_n(\mathbf{u})\}.$$

Note that $\theta(\mathbf{u})$ is the number of vectors $\mathbf{u}' \in \mathcal{U}_n$ that have the same output as for $\mathbf{u}$, i.e., $\Phi_n(\mathbf{u}') = \Phi_n(\mathbf{u})$. Let $\mathcal{A} \triangleq \{\mathbf{u} \in \mathcal{U}_n : \ \theta(\mathbf{u}) = 0\}$. Note that $\mathcal{A}$ is a subset over which Φ_n is injective. We show that $|\mathcal{A}^c| \leq \delta\ |\mathcal{U}_n|$ with high probability. Using Markov inequality:

$$\mathbb{P}\{|\mathcal{A}^c| \geq \delta|\mathcal{U}_n|\} \leq \frac{\mathbb{E}[|\mathcal{A}^c|]}{\delta|\mathcal{U}_n|},$$

where the expectation is taken with respect to the distribution on random mapping Φ_n. Note that

$$|\mathcal{A}^c| = \sum_{u\in\mathcal{U}_n} \mathbb{1}\{\theta(u) > 0\} \leq \sum_{u\in\mathcal{U}_n} \theta(u)$$

Hence,

$$\mathbb{P}\{|\mathcal{A}^c| \geq \delta|\mathcal{U}_n|\} \leq \frac{1}{\delta|\mathcal{U}_n|}\sum_{u\in\mathcal{U}_n}\mathbb{E}[\theta(u)]. \tag{A.1}$$

By definition, $\mathbb{E}[\theta(u)] = \sum_{\mathbf{u}'\neq\mathbf{u}}\mathbb{P}\{\Phi_n(\mathbf{u}') = \Phi_n(\mathbf{u})\}$. We provide an upper bound on $\mathbb{E}[\theta(u)]$.

Let $H_s = p^s\mathbb{Z}_{p^r}$ be a subgroup of $\mathbb{Z}_{p^r}$, where $s \in [0 : r-1]$. If $a \in \mathbb{Z}_{p^r} - \{0\}$, then there exits a maximum $s \in [0 : r-1]$ such that $a \in H_s$. That is $a \in H_s$ and $a \notin H_t$ for all $t > s$. As a result, for any $\mathbf{u}' \in \mathcal{U}_n$ there are r cases for the maximum s such that $u - u' \in H_s^{k_n}$. Considering these cases, we obtain

$$\sum_{\substack{\mathbf{u}' \in \mathcal{U}_n \\ \mathbf{u}' \neq \mathbf{u}}} P\{\Phi_n(\mathbf{u}') = \Phi_n(\mathbf{u})\} = \sum_{s=0}^{r-1} \sum_{\substack{\mathbf{u}' \in \mathcal{U}_n \\ \mathbf{u}' - \mathbf{u} \in H_s^{k_n} \setminus H_{s+1}^{k_n}}} P\{\Phi_n(\mathbf{u}') = \Phi_n(\mathbf{u})\} \tag{A.2}$$

Since Φ_n is a linear map, we have $P\{\Phi_n(\mathbf{u}') = \Phi_n(\mathbf{u})\} = P\{\Phi_n(\mathbf{u}' - \mathbf{u}) = 0\}$. Next, we use Lemma 29 (see Appendix A.9). Since $\mathbf{u}' - \mathbf{u} \in H_s^{k_n} \setminus H_{s+1}^{k_n}$, then $P\{\Phi_n(\mathbf{u}' - \mathbf{u}) = 0\} = p^{-n(r-s)}$. Therefore, using (A.2) and the expression for $\mathbb{E}[\theta(u)]$, we get

$$\mathbb{E}[\theta(u)] \leq \sum_{s=0}^{r-1} \sum_{\substack{\mathbf{u}' \in \mathcal{U}_n \\ \mathbf{u}' - \mathbf{u} \in H_s^{k_n}}} p^{-n(r-s)} \tag{A.3}$$

Next, we replace the summation over $\mathbf{u}'$ with the size of the set $\mathcal{U}_n \bigcap (\mathbf{u} + H_s^{k_n})$. Since $\mathcal{U}_n$ is a Cartesian product of typical sets, we use Lemma 30 (see Appendix A.9) to obtain the following bound

$$|\mathcal{U}_n \bigcap (\mathbf{u} + H_s^{k_n})| \leq \prod_q 2^{k_{q,n}\left(H(U_q|[U_q]_s) + \epsilon_q'\right)},$$

where $k_{q,n} = P_Q(q)k_n$. Therefore, the following bound holds:

$$\mathbb{E}[\theta(u)] \leq \sum_{s=0}^{r-1} 2^{k_n(H(U|Q[U]_s) + \epsilon')} p^{-n(r-s)} \tag{A.4}$$

By assumption, $H(U|[U]_s, Q) \leq \frac{1}{c}(r-s)\log_2 p - \epsilon, \forall s \in [0 : r-1]$. Therefore, for appropriate choice of ϵ and for sufficiently large n, the right-hand side of (A.4) can be made arbitrary small (say smaller than $\delta\gamma$). Therefore, from Markov inequality

given in (A.1), we obtain

$$\mathbb{P}\{|\mathcal{A}^c| \geq \delta|\mathcal{U}_n|\} \leq \frac{1}{\delta|\mathcal{U}_n|} \sum_{u \in \mathcal{U}_n} \gamma\delta = \gamma.$$

□

A.3 Proof of Lemma 4

Proof. Let $\mathcal{C}_n$ be the random (n, k_n)-QGC as in Lemma 4. For shorthand, for any $\mathbf{u} \in \mathcal{U}_n$, denote $\Phi_n(\mathbf{u}) = \mathbf{u}\mathbf{G}_n$, where $\mathbf{G}_n$ is the random matrix corresponding to $\mathcal{C}_n$. Fix $\mathbf{u}_0 \in \mathcal{U}_n$. Without loss of generality assume $\mathbf{c}(\theta) = \Phi_n(\mathbf{u}_0) + B$, where B is the translation associated with $\mathcal{C}_n$. Define the event $\mathcal{E}_n(\mathbf{u}) := \{(\Phi_n(\mathbf{u}) + B, \tilde{\mathbf{Y}}) \in A_\epsilon^{(n)}(X, Y)\}$, and let $\mathcal{E}_n$ be the event of interest as given in the lemma. Then $\mathcal{E}_n$ is the union of $\mathcal{E}_n(\mathbf{u})$ for all $\mathbf{u} \in \mathcal{U}_n \backslash \{\mathbf{u}_0\}$. By the union bound, the probability of $\mathcal{E}_n$ is bounded as

$$P(\mathcal{E}_n) \leq \sum_{\substack{\mathbf{u} \in \mathcal{U}_n \\ \mathbf{u} \neq \mathbf{u}_0}} P(\mathcal{E}_n(\mathbf{u})) \tag{A.5}$$

For any $\mathbf{u} \in \mathcal{U}_n$, the probability of $\mathcal{E}_n(\mathbf{u})$, can be calculated as,

$$P(\mathcal{E}_n(\mathbf{u})) = \sum_{\mathbf{x}_0 \in \mathbb{Z}_{p^r}^n} \sum_{\mathbf{y} \in \mathcal{Y}^n} P(\Phi_n(\mathbf{u}_0) + B = \mathbf{x}_0, \tilde{\mathbf{Y}} = \mathbf{y}, \mathcal{E}_n(\mathbf{u})) \tag{A.6}$$

$$= \sum_{\mathbf{x}_0 \in \mathbb{Z}_{p^r}^n} \sum_{\mathbf{y} \in A_\epsilon^{(n)}(Y)} \sum_{\mathbf{x}:(\mathbf{x},\mathbf{y}) \in A_\epsilon^{(n)}(X,Y)} P(\Phi_n(\mathbf{u}_0) + B = \mathbf{x}_0, \tilde{\mathbf{Y}} = \mathbf{y}, \Phi_n(\mathbf{u}) + B = \mathbf{x}) \tag{A.7}$$

By assumption, conditioned on $\Phi_n(\mathbf{u}_0) + B$, the random variable $\tilde{\mathbf{Y}}$ is independent of

$\Phi_n(\mathbf{u}) + B$. Therefore, the summand in (A.7) is simplified to

$$P(\Phi_n(\mathbf{u}_0) + B = \mathbf{x}_0, \Phi_n(\mathbf{u}) + B = \mathbf{x}) P^n_{Y|X}(\mathbf{y}|\mathbf{x}_0). \tag{A.8}$$

Since B is uniform over $\mathbb{Z}^n_{p^r}$, and is independent of other random variables,

$$P(\Phi_n(\mathbf{u}_0) + B = \mathbf{x}_0, \Phi_n(\mathbf{u}) + B = \mathbf{x}) = p^{-nr} P(\Phi_n(\mathbf{u} - \mathbf{u}_0) = \mathbf{x} - \mathbf{x}_0). \tag{A.9}$$

Using Lemma 29 (in Appendix A.9), if $\mathbf{u} - \mathbf{u}_0 \in H^{k_n}_s \backslash H^{k_n}_{s+1}$, then $P(\Phi_n(\mathbf{u} - \mathbf{u}_0) = \mathbf{x} - \mathbf{x}_0) = p^{-n(r-s)} \mathbb{1}\{\mathbf{x} - \mathbf{x}_0 \in H^{k_n}_s\}$. Therefore, using (A.7), and for $\mathbf{u} - \mathbf{u}_0 \in H^{k_n}_s \backslash H^{k_n}_{s+1}$ we obtain

$$P(\mathcal{E}_n(\mathbf{u})) = \sum_{\mathbf{x}_0 \in \mathbb{Z}^n_{p^r}} \sum_{\mathbf{y} \in A^{(n)}_\epsilon(Y)} \sum_{\substack{\mathbf{x}: \\ (\mathbf{x},\mathbf{y}) \in A^{(n)}_\epsilon(X,Y) \\ \mathbf{x} - \mathbf{x}_0 \in H^n_s}} p^{-nr} P^n_{Y|X}(\mathbf{y}|\mathbf{x}_0) p^{-n(r-s)}$$

Denote $\mathcal{A} \triangleq \{\mathbf{x} : (\mathbf{x}, \mathbf{y}) \in A^{(n)}_\epsilon(X, Y),\ \mathbf{x} - \mathbf{x}_0 \in H^n_s\}$. Note that if $([\mathbf{x}_0]_s, \mathbf{y}) \notin A^{(n)}_\epsilon([X]_s Y)$, then $\mathcal{A} = \emptyset$. Therefore,

$$P(\mathcal{E}_n(\mathbf{u})) = \sum_{\substack{(\mathbf{x}_0,\mathbf{y}): \\ ([\mathbf{x}_0]_s,\mathbf{y}) \in A^{(n)}_\epsilon([X]_s Y)}} \sum_{\mathbf{x} \in \mathcal{A}} p^{-nr} P^n_{Y|X}(\mathbf{y}|\mathbf{x}_0) p^{-n(r-s)} \tag{A.10}$$

Next, we replace the summation over $\mathbf{x}$ with the size of the set $\mathcal{A}$. We bound the

size of $\mathcal{A}$ using Lemma 30. Therefore, an upper-bound on (A.10) is

$$P(\mathcal{E}_n(\mathbf{u})) \leq \left(\sum_{\substack{(\mathbf{x}_0,\mathbf{y}):\\ ([\mathbf{x}_0]_s,\mathbf{y})\in A_\epsilon^{(n)}([X]_sY)}} p^{-nr} P_{Y|X}^n(\mathbf{y}|\mathbf{x}_0) \right) p^{-n(r-s)} 2^{n(H(X|Y,[X]_s)+\delta(4\epsilon))}$$

$$\leq \left(\sum_{\mathbf{x}_0\in\mathbb{Z}_{p^r}^n} \sum_{\mathbf{y}\in\mathcal{Y}^n} p^{-nr} P_{Y|X}^n(\mathbf{y}|\mathbf{x}_0) \right) p^{-n(r-s)} 2^{n(H(X|Y,[X]_s)+\delta(4\epsilon))} \tag{A.11}$$

$$\leq p^{-n(r-s)} 2^{n(H(X|Y,[X]_s)+\delta(4\epsilon))}. \tag{A.12}$$

Note that if $\mathbf{a}\in\mathbb{Z}_{p^r}^k, \mathbf{a}\neq\mathbf{0}$ then there exists $s\in[0:r-1]$ such that $\mathbf{a}\in H_s^k\backslash H_{s+1}^k$. Therefore, there are r different cases for each value of s. Using (A.12), and considering these cases, we obtain

$$P(\mathcal{E}_n) \leq \sum_{s=0}^{r-1} \sum_{\substack{\mathbf{u}\in\mathcal{U}_n\\ \mathbf{u}-\mathbf{u}_0\in H_s^{k_n}\backslash H_{s+1}^{k_n}}} P(\mathcal{E}_n(\mathbf{u})) \leq \sum_{s=0}^{r-1} \sum_{\substack{\mathbf{u}\in\mathcal{U}_n\\ \mathbf{u}-\mathbf{u}_0\in H_s^{k_n}\backslash H_{s+1}^{k_n}}} 2^{n(H(X|Y[X]_s)+\delta(4\epsilon))} p^{-n(r-s)}$$

$$\leq \sum_{s=0}^{r-1} |\mathcal{U}_n \bigcap (\mathbf{u}_0 + H_s^k)| 2^{n(H(X|Y[X]_s)+\delta(4\epsilon))} p^{-n(r-s)}$$

Note that $\mathcal{U}_n$ is the Cartesian product of ϵ-typical sets $A_\epsilon^{(p(q)k_n)}(U_q)$, $q\in\mathcal{Q}$. For each component q of $\mathcal{U}_n$, we can apply Lemma 30. Therefore,

$$|\mathcal{U}_n \cap (\mathbf{u}_0 + H_s^k)| \leq 2^{\sum_q p(q)k_n(H(U_q|[U_q]_s)+\delta(2\epsilon))} = 2^{k_n(H(U|[U]_s,Q)+\delta(2\epsilon))}.$$

Finally,

$$P(\mathcal{E}_n) \leq \sum_{s=0}^{r-1} 2^{n\left(\frac{k_n}{n}(H(U|[U]_s,Q)+H(X|Y,[X]_s)+\frac{k_n}{n}\delta(2\epsilon)+\delta(4\epsilon)\right)} p^{-n(r-s)}$$

As a result $\lim_{n\to\infty} P(\mathcal{E}_n) = 0$, if the inequality

$$cH(U|[U]_s, Q) \leq \log_2 p^{r-s} - H(X|Y, [X]_s) - 2(2+c)\delta(\epsilon),$$

holds for all $0 \leq s \leq r-1$. Multiply each side of this inequality by $\frac{H(U|Q)}{H(U|Q,[U]_s)}$. This gives the following bound

$$cH(U|Q) \leq \frac{H(U|Q)}{H(U|Q,[U]_s)}(\log_2 p^{r-s} - H(X|Y, [X]_s) - 2(2+c)\delta(\epsilon))$$

By definition $R_n = \frac{1}{n}\log_2 |\mathcal{C}_n| \leq cH(U|Q) + \epsilon'$. Therefore,

$$R_n \leq \frac{H(U|Q)}{H(U|Q,[U]_s)}(\log_2 p^{r-s} - H(X|Y, [X]_s) - 2(2+c)\delta(\epsilon)),$$

and the proof is completed. □

A.4 Proof of Lemma 5

Proof. We use the same notation as in the proof of Lemma 4. For any typical sequence $\mathbf{x}$ define

$$\lambda_n(\mathbf{x}) = \sum_{\hat{\mathbf{x}} \in A_\epsilon^{(n)}(\hat{X}|\mathbf{x})} \sum_{\mathbf{u}\in\mathcal{U}_n} \mathbb{1}\{\Phi_n(\mathbf{u}) + B = \hat{x}\}.$$

Note $\lambda_n(\mathbf{x})$ counts the number of codewords that are conditionally typical with $\mathbf{x}$ with respect to $p(\hat{\mathbf{x}}|\mathbf{x})$. We show that $\lim_{n\to\infty} P(\lambda_n(\mathbf{x}) = 0) = 0$ for any ϵ-typical sequence $\mathbf{x}$. This implies that $\lim_{n\to\infty} P(\lambda_n(\mathbf{X}^n) = 0) = 0$, where $\mathbf{X}^n \sim \prod_{i=1}^n p(x)$. This proves the statements of the Lemma. Hence, it suffices to show that $\lim_{n\to\infty} P(\lambda_n(\mathbf{x}) = 0) =$

0. We have,

$$P\{\lambda_n(\mathbf{x}) = 0\} \leq P\Big\{\lambda_n(\mathbf{x}) \leq \frac{1}{2}E(\lambda_n(x))\Big\} \leq P\Big\{|\lambda_n(x) - E(\lambda_n(x))| \geq \frac{1}{2}E(\lambda_n(x))\Big\} \tag{A.13}$$

Hence, by Chebyshev's inequality, $P\{\lambda_n(\mathbf{x}) = 0\} \leq \frac{4Var(\lambda_n(x))}{E(\lambda_n(x))^2}$. Note that

$$E(\lambda_n(x)) = \sum_{\hat{\mathbf{x}} \in A_\epsilon^{(n)}(\hat{X}|\mathbf{x})} \sum_{\mathbf{u} \in \mathcal{U}_n} P\{\Phi(\mathbf{u}) + B = \hat{\mathbf{x}}\} \tag{A.14}$$

Since B is uniform over $\mathbb{Z}_{p^r}^n$, we get

$$E(\lambda_n(x)) = |A_\epsilon^{(n)}(X|\hat{\mathbf{x}})||\mathcal{U}_n|p^{-rn}. \tag{A.15}$$

Note $2^{k_n(H(U|Q)-2\epsilon')} \leq |\mathcal{U}_n| \leq 2^{k_n(H(U|Q)+2\epsilon')}$, where

$$\epsilon' = -\frac{\epsilon}{p^r} \sum_{q \in \mathcal{Q}} P_Q(q) \sum_{a \in \mathbb{Z}_{p^r} : P_{U|Q}(a|q) > 0} \log P_{U|Q}(a|q).$$

Therefore,

$$2^{n(H(\hat{X}|X)-2\tilde{\epsilon})}2^{k_n(H(U|Q)-2\epsilon')}p^{-rn} \leq E(\lambda_n(x)) \leq 2^{n(H(\hat{X}|X)+2\tilde{\epsilon})}2^{k_n(H(U|Q)+2\epsilon')}p^{-rn}, \tag{A.16}$$

To calculate the variance, we start with

$$E(\lambda_n(x)^2) = \sum_{\hat{\mathbf{x}}, \hat{\mathbf{x}}' \in A_\epsilon^{(n)}(\hat{X}|\mathbf{x})} \sum_{\mathbf{u}, \mathbf{u}' \in \mathcal{U}_n} P\{\Phi(\mathbf{u}) + B = \hat{\mathbf{x}}, \Phi(\mathbf{u}') + B = \hat{\mathbf{x}}'\}.$$

Since B is independent of other random variables, the most inner term in the above summations is simplified to $p^{-nr}P\{\Phi(\mathbf{u} - \mathbf{u}') = \hat{\mathbf{x}} - \hat{\mathbf{x}}'\}$. Using Lemma 29 (in

Appendix A.9), if $\mathbf{u}-\mathbf{u}' \in H_s^{k_n} \backslash H_{s+1}^{k_n}$, then

$$P\{\Phi(\mathbf{u}-\mathbf{u}') = \hat{\mathbf{x}}-\hat{\mathbf{x}}'\} = p^{-n(r-s)} \mathbb{1}\{\hat{\mathbf{x}}-\hat{\mathbf{x}}' \in H_s^n\}$$

Considering all the cases for the values of s, we get

$$E(\lambda_n(x)^2) = \sum_{s=0}^{r} \sum_{\substack{\mathbf{u},\mathbf{u}' \in \mathcal{U}_n \\ \mathbf{u}-\mathbf{u}' \in H_s^{k_n} \backslash H_{s+1}^{k_n}}} \sum_{\substack{\hat{\mathbf{x}},\hat{\mathbf{x}}' \in A_\epsilon^{(n)}(\hat{X}|\mathbf{x}) \\ \hat{\mathbf{x}}-\hat{\mathbf{x}}' \in H_s^n}} p^{-nr} p^{-n(r-s)}$$

Since the innermost terms in the above summations do not depend on the individual values of $\mathbf{x}, \hat{\mathbf{x}}, \mathbf{u}, \mathbf{u}'$, the corresponding summations can be replaced by the size of the associated sets. Moreover, we provide an upper bound on the summation over $\mathbf{u}, \mathbf{u}'$ by replacing $H_s^{k_n} \backslash H_{s+1}^{k_n}$ with $H_s^{k_n}$. Using Lemma 30 for $\mathbf{x}, \hat{\mathbf{x}}$, we get

$$E(\lambda_n(x)^2) \leq \sum_{s=0}^{r} \sum_{\mathbf{u} \in \mathcal{U}_n} \sum_{\substack{\mathbf{u}' \in \mathcal{U}_n \\ \mathbf{u}-\mathbf{u}' \in H_s^{k_n}}} 2^{n(H(\hat{X}|X)+\tilde{\epsilon}+H(\hat{X}|X,[\hat{X}]_s)+\delta(4\epsilon))} p^{-nr} p^{-n(r-s)}$$

For any $\mathbf{u} \in \mathcal{U}_n$, by applying Lemma 30 we get $|\mathcal{U}_n \bigcap (\mathbf{u}+H_s^{k_n})| \leq 2^{k_n(H(U|Q,[U]_s)+\delta(4\epsilon))}$. As a result,

$$E(\lambda_n(x)^2) \leq \sum_{s=0}^{r} 2^{k_n(H(U|Q,[U]_s)+\delta(4\epsilon))} 2^{k_n(H(U|Q)+\epsilon')} 2^{n(H(\hat{X}|X)+\tilde{\epsilon}+H(\hat{X}|X,[\hat{X}]_s)+\delta(4\epsilon))} p^{-nr} p^{-n(r-s)}.$$

Note that the case $s=0$ gives $E^2(\lambda_n(x))$. Therefore,

$$Var(\lambda_n(x)^2) \leq p^{-nr} \sum_{s=1}^{r} 2^{k_n(H(U|Q)+H(U|Q,[U]_s))} 2^{n(H(\hat{X}|X)+H(\hat{X}|X,[\hat{X}]_s))} 2^{n(1+c)(\epsilon+\delta(4\epsilon))} p^{-n(r-s)} \tag{A.17}$$

Finally, using (A.16), (A.17) and the Chebyshev's inequality as argued before, we

get

$$P\{\lambda_n(\mathbf{x}) = 0\} \leq 4 \sum_{s=1}^{r} 2^{k_n(-H(U|Q)+H(U|Q,[U]_s))} 2^{n(-H(\hat{X}|X)+H(\hat{X}|X,[\hat{X}]_s))} 2^{n(1+c)(\epsilon+\delta(4\epsilon))} p^{nr} p^{-n(r-s)}$$

$$= 4\ 2^{n(1+c)(\epsilon+\delta(4\epsilon))} \sum_{s=1}^{r} 2^{-k_n H([U]_s|Q)} 2^{-nH([\hat{X}]_s|X)} p^{ns}.$$

The second equality follows, because $H(V|W) - H(V|[V]_s, W) = H([V]_s|W)$ holds for any random variables V and W. Therefore, $P\{\lambda_n(\mathbf{x})\}$ approaches zero, as $n \to \infty$, if

$$cH([U]_s|Q) \geq \log_2 p^s - H([\hat{X}]_s|X) + (1+c)(\epsilon + \delta(4\epsilon)), \quad \text{for} \ \ 1 \leq s \leq r.$$

By the definition of rate and the above inequalities the proof is completed.

□

A.5 Proof of Theorem II.2

Fix a positive integer n, and define $l_1 \triangleq c_1 n, l_2 \triangleq c_2 n$, and $k \triangleq \tilde{c} n$, where $\tilde{c}, c_1$ and c_2 are positive real numbers such that l_1, l_2 and k are integers.

Codebook Generation We use two nested QGC's, one for each encoder. The codebook for Encoder 1 is an (n, k, l_1) nested QGC (as in Definition 5) with random variables (W_1, V_1, Q). Let $\mathcal{C}_{I,1}, \bar{\mathcal{C}}_1$, and $\mathcal{C}_{O,1}$ denote the corresponding inner code, shift code and the outer code (as in Definition 5), respectively. The codebook for Encoder 2 is an (n, k, l_2) nested QGC with random variables (W_2, V_2, Q), inner code $\mathcal{C}_{I,2}$, shift code $\bar{\mathcal{C}}_2$, and outer code $\mathcal{C}_{O,2}$. The codebook at the decoder is denoted by $\mathcal{C}_d$ which is an (n, k) QGC with random variables $(W_1 + W_2, Q)$.

Conditioned on Q, the random variables (W_1, W_2, V_1, V_2) are mutually indepen-

dent. The random variable V_i is uniform over $\{0,1\}$, and is independent of Q.

The nested QGCs and $\mathcal{C}_d$ have identical generator matrices but different translations and index random variables. Note that each nested QGC has two generator matrices/translations, one for the inner code and one for the shift code as in Definition 5. The generator matrix and the translation for the inner codes $\mathcal{C}_{I,i}, i = 1, 2$, are denoted by $\mathbf{G}$ and $\mathbf{b}$, respectively. The generator matrix and the translation used for shift code $\mathcal{C}_{I,i}$, are denoted by $\bar{\mathbf{G}}$ and $\bar{\mathbf{b}}_i$, respectively, where $i = 1, 2$. The elements of $\mathbf{G}, \bar{\mathbf{G}}, \mathbf{b}$, and $\bar{\mathbf{b}}_i, i = 1, 2$ are generated randomly and independently from $\mathbb{Z}_{p^r}$.

By $R_{O,i}$ and $R_{I,i}$ denote the rate of the inner code and outer code defined for the ith nested QGC. Define $R_i \triangleq R_{O,i} - R_{I,i}, i = 1, 2$.

Encoding Suppose $(\mathbf{x}_1, \mathbf{x}_2)$ is a realization of (X_1^n, X_2^n). The first encoder checks if $\mathbf{x}_1$ is ϵ-typical and $\mathbf{x}_1 \in \mathcal{C}_{O,1}$. If not, an encoding error E_1 is declared. In the case of no encoding error, by Definition 5, $\mathbf{x}_1 = \mathbf{c}_{I,1} + \bar{\mathbf{c}}_1$, where $\mathbf{c}_{I,1} \in \mathcal{C}_{I,1}$ and $\bar{\mathbf{c}}_1 \in \bar{\mathcal{C}}_1$. The first encoder sends the index of $\bar{\mathbf{c}}_1$. Note $\bar{\mathbf{c}}_1$ determines the index of the bin which contains $\mathbf{x}_1$. Similarly, if $\mathbf{x}_2 \in A_\epsilon^{(n)}(X_2)$ and $\mathbf{x}_2 \in \mathcal{C}_{O,2}$, the second encoder sends finds $\mathbf{c}_{I,2} \in \mathcal{C}_{I,2}$ and $\bar{\mathbf{c}}_2 \in \bar{\mathcal{C}}_2$ such that $\mathbf{x}_2 = \mathbf{c}_{I,2} + \bar{\mathbf{c}}_2$. Then it sends the index of $\bar{\mathbf{c}}_2$. If no such $\mathbf{c}_{I,2}$ and $\bar{\mathbf{c}}_2$ are found, an error event E_2 is declared.

Decoding The decoder wishes to reconstruct $\mathbf{x}_1 + \mathbf{x}_2$. Assume there is no encoding error. Upon receiving the bin numbers from the encoders, the decoder calculates $\bar{\mathbf{c}}_1$ and $\bar{\mathbf{c}}_2$. Then, it finds $\tilde{\mathbf{c}} \in \mathcal{C}_d$ such that $\tilde{\mathbf{c}} + \bar{\mathbf{c}}_1 + \bar{\mathbf{c}}_2 \in A_\epsilon^{(n)}(X_1 + X_2)$. If $\tilde{\mathbf{c}}$ is unique, then $\tilde{\mathbf{c}} + \bar{\mathbf{c}}_1 + \bar{\mathbf{c}}_2$ is declared as a reconstruction of $\mathbf{x}_1 + \mathbf{x}_2$. An error event E_d occurs, if no unique $\tilde{\mathbf{c}}$ was found.

We need to find conditions for which the probability of the error events E_1, E_2 and E_d approach zero. By $\mathcal{W}_i$ denote the index set of $\mathcal{C}_{I,i}$, and let $\mathcal{V}_i$ be the index set of $\bar{\mathcal{C}}_i, i = 1, 2$.

Error Let $(f_1(\cdot), f_2(\cdot))$ and $g(\cdot,\cdot)$ denote the encoding and decoding functions corresponding to the above coding scheme. The overall error event is defined as

$$E \triangleq \{\mathbf{X}_1^n + \mathbf{X}_2^n \neq g(f_1(\mathbf{X}_1^n), f_2(\mathbf{X}_2^n))\}$$

For the achievability, we need to show that $P(E)$ can be made arbitrary small for sufficiently large n. For that, using the aforementioned encoding and decoding error events we have

$$P(E) \leq P(E_1 \cup E_2 \cup E_d) + P(E|E_1^c \cap E_2^c \cap E_d^c)$$

Using standard arguments for typical sequences, we can show that when there is no encoding and decoding error (i.e., $E_1^c \cap E_2^c \cap E_d^c$) the error probability $P(E|E_1^c \cap E_2^c \cap E_d^c)$ approaches 0 as $n \to \infty$. As a result, the second term above is sufficiently small for large enough n. Therefore, for sufficiently large n and from the union bound on the first term we obtain,

$$P(E) \leq P(E_1) + P(E_2) + P(E_d) + \epsilon$$

A.5.1 Analysis of E_1, E_2

In what follows, we apply the covering lemma (Lemma 5) to bound the probability of the encoding errors. For that the outer code $\mathcal{C}_{O,i}$ is used to "cover" the source X_i. Note that $\mathcal{C}_{O,i}$ is the outer code for the (n,k,l) nested QGC used at Encoder i, $i = 1, 2$. Therefore, $\mathcal{C}_{O,i}$ is a $(n, k+l)$ QGC with appropriately defined index random variables (as is defined in Lemma 3). The random variables defined for $\mathcal{C}_{O,i}$ are $(U_i, (Q, J_i))$, where given $J_i = 1$ we have $U_i = W_i$, and given $J_i = 2$ we get

$U_i = V_i$. In addition, $P(J_i = 0) = \frac{k}{l_i+k}$, and $P(J_i = 1) = \frac{l_i}{l_i+k}$. We apply Lemma 5 to bound the probability of E_i. In this lemma set $\hat{X} = X = X_i$ with probability one, $\mathcal{C}_n = \mathcal{C}_{O,i}$, and $R_n = R_{O,i}, i = 1, 2$. Using Lemma 5, $P(E_i)$ is sufficiently small for large blocklength n if

$$R_{O,i} \geq \max_{1 \leq s \leq r} \frac{H(U_i|Q, J_i)}{H([U_i]_s|Q, J_i)}(\log_2 p^s + o(\epsilon)).$$

Using Remark 3, and the above bound we get $\frac{k+l_i}{n} H([U_i]_s|Q, J_i) \geq \log_2 p^s + o(\epsilon)$ for $s \in [1 : r]$. Therefore, by the definition of U_i and J_i, we get

$$\frac{k}{n} H([W_i]_s|Q) + \frac{l_i}{n} H(V_i|Q) \geq \log_2 p^s + o(\epsilon),\ 1 \leq s \leq r.$$

Note that in this bound we use the equality $H([V_i]_s) = H(V_i)$. This equality holds because V_i takes values from $\{0, 1\}$. Again using Remark 3, we get $|R_i - \frac{l_i}{n} H(V_i|Q)| \leq o(\epsilon)$. Hence, if the following holds

$$\frac{k}{n} H([W_i]_s|Q) + R_i \geq \log_2 p^s + o(\epsilon),\ 1 \leq s \leq r,\ i = 1, 2, \tag{A.18}$$

then $P(E_i) \to 0$ as $n \to \infty$.

A.5.2 Analysis of E_d

Upon receiving the bin numbers, the decoder calculates $\bar{\mathbf{c}}_1$ and $\bar{\mathbf{c}}_2$. The decoding error consists of two events: 1) no typical sequence $\tilde{\mathbf{z}}$ was found, and 2) multiple typical sequences $\tilde{\mathbf{z}}$ were found. Using standard arguments, one can show that the probability of the first event is sufficiently small for large enough n. In what follows, we bound the probability of the second event, i.e., $E_{d,2}$. This event occurs, if there exist more than one $\tilde{\mathbf{c}} \in \mathcal{C}_{I,1} + \mathcal{C}_{I,2}$ such that $\tilde{\mathbf{c}} + \bar{\mathbf{c}}_1 + \bar{\mathbf{c}}_2$ is ϵ-typical with respect to $P_{X_1+X_2}$.

To bound $P(E_{d,2})$ we need to take into account whether there is an encoding error or not. For that, first we provide an alternative representation for the encoding errors. For any sequence $\mathbf{x}_i \in \mathbb{Z}_{p^r}^n$ define

$$\lambda_i(\mathbf{x}_i) = \sum_{\mathbf{w}_i \in \mathcal{W}_i} \sum_{\mathbf{v}_i \in \mathcal{V}_i} \mathbb{1}\{\mathbf{x}_i = \mathbf{w}_i\mathbf{G} + \mathbf{v}_i\bar{\mathbf{G}} + \mathbf{b} + \bar{\mathbf{b}}_i\},$$

where $i = 1, 2$ and $(\mathbf{G}, \bar{\mathbf{G}}, \mathbf{b}, \bar{\mathbf{b}}_i)$ are the generator matrices and translations defined for the ith nested QGC. With this notation, E_i occurs if $\lambda_i(\mathbf{x}_i) = 0$, where $(\mathbf{x}_1, \mathbf{x}_2)$ is a realization of the sources. Next, we define a super-set of the encoding error events as

$$E_i' \triangleq \{\lambda_i(\mathbf{x}_i) < \frac{1}{2}E(\lambda_i(\mathbf{x}_i))\}, \qquad i = 1, 2, \tag{A.19}$$

where $E(\lambda_i(\mathbf{x}_i))$ is the expected value of $\lambda_i(\mathbf{x}_i)$. Note that $E_i \subseteq E_i', i = 1, 2$.

For the modified encoding error events (E_1', E_2') given in (A.19) we have

$$\begin{aligned} P(E_{d,2}) &\leq P(E_1' \cup E_2') + P(E_{d,2} \cap E_1'^c \cap E_2'^c) \\ &\leq P(E_1') + P(E_2') + P(E_{d,2} \cap E_1'^c \cap E_2'^c) \end{aligned}$$

For the first two terms above, based on the proof of Lemma 5, we can showed that $P(E_i') \to 0$ as $n \to \infty$. Note that $P(E_i')$ is the same as the second term in (A.13) in the proof of the covering. In fact, for the proof of the covering bound, we showed that such probability approaches 0 as $n \to \infty$.

In what follow, we show that the second probability in the above approaches 0 as $n \to \infty$.

Analysis of $\mathbf{P(E_{d,2}|E_1'^c \cap E_2'^c)}$ Note that $E_1'^c \cap E_2'^c$ implies that there is no encoding error; because $\lambda_i(\mathbf{x}_i) > 1/2E(\lambda_i(\mathbf{x}_i))$. Since there is no error at the encoding stage,

$\mathbf{x}_i \in \mathcal{C}_{O,i}, i = 1, 2$. By Definition 5, every codeword in $\mathcal{C}_{O,i}$ is characterized by a pair $(\mathbf{v}_i, \mathbf{w}_i)$, where $\mathbf{v}_i \in \mathcal{V}_i, \mathbf{w}_i \in \mathcal{W}_i, i = 1, 2$. Given $\mathbf{x}_i$, if more than one pair was found at the ith encoder, select one randomly and uniformly. By $P(\mathbf{v}_i, \mathbf{w}_i|\mathbf{x}_i)$ denote the probability that $(\mathbf{v}_i, \mathbf{w}_i)$ is selected at the ith encoder. Then, $P(\mathbf{v}_i, \mathbf{w}_i|\mathbf{x}_i) = \frac{1}{\lambda_i(\mathbf{x}_i)} \mathbb{1}\{\mathbf{w}_i\mathbf{G} + \mathbf{v}_i\bar{\mathbf{G}} + \mathbf{b} + \bar{\mathbf{b}}_i = \mathbf{x}_i\}$. Fix $\mathbf{G}$, $\tilde{\mathbf{G}}_i, \mathbf{b}$ and $\bar{\mathbf{b}}_i, i = 1, 2$. Suppose $\mathbf{x}_1$ and $\mathbf{x}_2$ are the realizations of the sources X_1 and X_2, respectively. Moreover, suppose $(\mathbf{x}_1, \mathbf{x}_2) \in A_\epsilon^{(n)}(X_1, X_2)$. Therefore,

$$P(E_{d,2} \cap E_1'^c \cap E_2'^c|\mathbf{x}_1, \mathbf{x}_2) = \mathbb{1}\Big\{\lambda_i(\mathbf{x_i}) \geq \frac{1}{2}E(\lambda_i(\mathbf{x}_i)), i = 1, 2\Big\} \left[\prod_{j=1}^{2} \sum_{\mathbf{v}_j \in \mathcal{V}_j} \sum_{\mathbf{w}_j \in \mathcal{W}_j} P(\mathbf{v}_j, \mathbf{w}_j|\mathbf{x}_j)\right]$$
$$P(E_{d,2}|\mathbf{x}_i, \mathbf{v}_i, \mathbf{w}_i, i = 1, 2)$$

In what follows, we bound $P(E_{d,2}|\mathbf{x}_i, \mathbf{v}_i, \mathbf{w}_i, i = 1, 2)$, $P(\mathbf{v}_1, \mathbf{w}_1|\mathbf{x}_1)$, and $P(\mathbf{v}_2, \mathbf{w}_2|\mathbf{x}_2)$. For the first conditional probability we have

$$P(E_{d,2}|\mathbf{x}_i, \mathbf{v}_i, \mathbf{w}_i, i = 1, 2) = \mathbb{1}\{\exists \tilde{\mathbf{z}} \in A_\epsilon^{(n)}(X_1 + X_2) : \tilde{\mathbf{z}} \neq \mathbf{x}_1 + \mathbf{x}_2, \tilde{\mathbf{z}} \in \mathcal{C}_{I,1} + \mathcal{C}_{I,2} + \bar{\mathbf{c}}_1 + \bar{\mathbf{c}}_2\}$$

where, $\bar{\mathbf{c}}_i = \mathbf{v}_i\bar{\mathbf{G}} + \bar{\mathbf{b}}_i, i = 1, 2$. Let $\mathcal{W} = \mathcal{W}_1 + \mathcal{W}_2$, and define $Z \triangleq X_1 + X_2$. Using the union bound, we have

$$\begin{aligned}
&P(E_{d,2}|\mathbf{x}_i, \mathbf{v}_i, \mathbf{w}_i, i = 1, 2) \\
&\quad \leq \sum_{\tilde{\mathbf{w}} \in \mathcal{W}} \sum_{\substack{\tilde{\mathbf{z}} \in A_\epsilon^{(n)}(Z) \\ \tilde{\mathbf{z}} \neq \mathbf{x}_1 + \mathbf{x}_2}} \mathbb{1}\{\tilde{\mathbf{w}}\mathbf{G} + (\mathbf{v}_1 + \mathbf{v}_2)\bar{\mathbf{G}} + 2\mathbf{b} + \bar{\mathbf{b}}_1 + \bar{\mathbf{b}}_2 = \tilde{\mathbf{z}}\} \\
&\quad \leq \sum_{\substack{\tilde{\mathbf{w}} \in \mathcal{W} \\ \tilde{\mathbf{w}} \neq \mathbf{w}_1 + \mathbf{w}_2}} \sum_{\tilde{\mathbf{z}} \in A_\epsilon^{(n)}(Z)} \mathbb{1}\{\tilde{\mathbf{w}}\mathbf{G} + (\mathbf{v}_1 + \mathbf{v}_2)\bar{\mathbf{G}} + 2\mathbf{b} + \bar{\mathbf{b}}_1 + \bar{\mathbf{b}}_2 = \tilde{\mathbf{z}}\} \qquad \text{(A.20)}
\end{aligned}$$

The second inequality follows, because the condition $\tilde{\mathbf{w}} \neq \mathbf{w}_1 + \mathbf{w}_2$ is less restrictive than $\tilde{\mathbf{z}} \neq \mathbf{x}_1 + \mathbf{x}_2$. This is due to the fact that $\mathbf{G}$ is not injective necessarily.

Next, we provide an upper-bound on $P(\mathbf{v}_i, \mathbf{w}_i|\mathbf{x}_i), i = 1, 2$. Since $E_1'^c \cap E_2'^c$ is in

the conditioning, $\lambda_i(\mathbf{x}_i) \geq \frac{1}{2}E(\lambda_i(\mathbf{x}_i))$. As a result,

$$P(\mathbf{v}_i, \mathbf{w}_i|\mathbf{x}_i) \leq \frac{2}{E(\lambda_i(\mathbf{x}_i))}\mathbb{1}\{\mathbf{w}_i\mathbf{G} + \mathbf{v}_i\bar{\mathbf{G}} + \mathbf{b} + \bar{\mathbf{b}}_i = \mathbf{x}_i\} \tag{A.21}$$

Using the bounds given in (A.20) and (A.21), we get

$$P(E_{d,2} \cap E_1'^c \cap E_2'^c|\mathbf{x}_1, \mathbf{x}_2) \leq \left[\prod_{j=1}^{2} \sum_{\substack{\mathbf{v}_j \in \mathcal{V}_j \\ \mathbf{w}_j \in \mathcal{W}_j}} \frac{2}{E(\lambda_j(\mathbf{x}_j))}\mathbb{1}\{\mathbf{w}_j\mathbf{G} + \mathbf{v}_j\bar{\mathbf{G}} + \mathbf{b} + \bar{\mathbf{b}}_j = \mathbf{x}_j\}\right]$$
$$\sum_{\substack{\tilde{\mathbf{w}} \in \mathcal{W} \\ \tilde{\mathbf{w}} \neq \mathbf{w}_1 + \mathbf{w}_2}} \sum_{\tilde{\mathbf{z}} \in A_\epsilon^{(n)}(Z)} \mathbb{1}\{\tilde{\mathbf{w}}\mathbf{G} + (\mathbf{v}_1 + \mathbf{v}_2)\bar{\mathbf{G}} + 2\mathbf{b} + \bar{\mathbf{b}}_1 + \bar{\mathbf{b}}_2 = \tilde{\mathbf{z}}\}$$

Next, we average $P(E_{d,2} \cap E_1'^c \cap E_2'^c|\mathbf{x}_1, \mathbf{x}_2)$ over all possible choices of $\mathbf{G}, \bar{\mathbf{G}}, \mathbf{b}, \bar{\mathbf{b}}_1$, and $\bar{\mathbf{b}}_2$. We obtain

$$\mathbb{E}\{P(E_{d,2} \cap E_1'^c \cap E_2'^c|\mathbf{x}_1, \mathbf{x}_2)\} \leq \sum_{\substack{\mathbf{v}_1 \in \mathcal{V}_1 \\ \mathbf{w}_1 \in \mathcal{W}_1}} \frac{2}{E(\lambda_1(\mathbf{x}_1))} \sum_{\substack{\mathbf{v}_2 \in \mathcal{V}_2 \\ \mathbf{w}_2 \in \mathcal{W}_2}} \frac{2}{E(\lambda_2(\mathbf{x}_2))} \sum_{\substack{\tilde{\mathbf{w}} \in \mathcal{W} \\ \tilde{\mathbf{w}} \neq \mathbf{w}_1 + \mathbf{w}_2}} \sum_{\tilde{\mathbf{z}} \in A_\epsilon^{(n)}(Z)}$$
$$P\{\tilde{\mathbf{w}}\mathbf{G} + (\mathbf{v}_1 + \mathbf{v}_2)\bar{\mathbf{G}} + 2\mathbf{B} + \bar{\mathbf{B}}_1 + \bar{\mathbf{B}}_2 = \tilde{\mathbf{z}}, \mathbf{w}_i\mathbf{G} + \mathbf{v}_i\bar{\mathbf{G}} + \mathbf{B} + \bar{\mathbf{B}}_i = \mathbf{x}_i, i = 1, 2\}$$

Note $\bar{\mathbf{B}}_1$ and $\bar{\mathbf{B}}_2$ are independent random variables with uniformly distributed over $\mathbb{Z}_{p^r}^n$. Therefore, the innermost term in the above summations equals

$$p^{-2nr}P\{(\tilde{\mathbf{w}} - \mathbf{w}_1 - \mathbf{w}_2)\mathbf{G} = \tilde{\mathbf{z}} - \mathbf{x}_1 - \mathbf{x}_2\}. \tag{A.22}$$

We apply Lemma 29 (in Appendix A.9), to calculate the above probability. If $\tilde{\mathbf{w}} - \mathbf{w}_1 - \mathbf{w}_2 \in H_s^k \backslash H_{s+1}^k$, then (A.22) equals to

$$p^{-2nr}p^{-n(r-s)}\mathbb{1}\{\tilde{\mathbf{z}} - \mathbf{x}_1 - \mathbf{x}_2 \in H_s^k\}. \tag{A.23}$$

As a result, we have

$$\mathbb{E}\{P(E_{d,2}\cap E_1'^{\,c}\cap E_2'^{\,c}|\mathbf{x}_1,\mathbf{x}_2)\}\le \sum_{\substack{\mathbf{v}_1\in\mathcal{V}_1\\ \mathbf{w}_1\in\mathcal{W}_1}}\frac{2}{E(\lambda_1(\mathbf{x}_1))}\sum_{\substack{\mathbf{v}_2\in\mathcal{V}_2\\ \mathbf{w}_2\in\mathcal{W}_2}}\frac{2}{E(\lambda_2(\mathbf{x}_2))} \sum_{s=0}^{r-1}\sum_{\substack{\tilde{\mathbf{w}}\in\mathcal{W}\\ \tilde{\mathbf{w}}-\mathbf{w}_1-\mathbf{w}_2\in H_s^k\setminus H_{s+1}^k}}\sum_{\substack{\tilde{\mathbf{z}}\in A_\epsilon^{(n)}(Z)\\ \tilde{\mathbf{z}}-\mathbf{x}_1-\mathbf{x}_2\in H_s^n}} p^{-2nr}p^{-n(r-s)}$$

Since the innermost terms in the above summations depend only on s, we can replace the summations over $\tilde{\mathbf{w}}$ and $\tilde{\mathbf{z}}$ with the size of the associated sets. We apply Lemma 30 to bound the size of these sets. Also, we can replace the summations over $\mathbf{v}_i$ and $\mathbf{w}_i, i=1,2$ with the size of the related sets. Define $W\triangleq W_1+W_2$, we get,

$$\mathbb{E}\{P(E_{d,2}\cap E_1'^{\,c}\cap E_2'^{\,c}|\mathbf{x}_1,\mathbf{x}_2)\}\le |\mathcal{W}_1||\mathcal{V}_1|\frac{2}{E(\lambda_1(\mathbf{x}_1))}|\mathcal{W}_2||\mathcal{V}_2|\frac{2}{E(\lambda_2(\mathbf{x}_2))} \sum_{s=0}^{r-1}2^{n(H(Z|[Z]_s)+o(\epsilon))}2^{k(H(W|Q,[W]_s)+o(\epsilon))}p^{-2nr}p^{-n(r-s)}.$$

Note that from (A.15) in the proof of Lemma 5, $E(\lambda_i(\mathbf{x}_i))=|\mathcal{W}_i||\mathcal{V}_i|p^{-nr}, i=1,2$. Therefore, we have

$$\mathbb{E}\{P(E_{d,2}\cap E_1'^{\,c}\cap E_2'^{\,c}|\mathbf{x}_1,\mathbf{x}_2)\}\le 4\sum_{s=0}^{r-1}2^{n(H(Z|[Z]_s)+o(\epsilon))}2^{k(H(W|Q,[W]_s)+o(\epsilon))}p^{-n(r-s)}.$$

Note that the above bound does not depend on ϵ-typical sequences $\mathbf{x}_1$ and $\mathbf{x}_2$. Using standard arguments for ϵ-typical sets, the probability that $(\mathbf{X}_1^n,\mathbf{X}_2^n)\notin A_\epsilon^{(n)}(X_1,X_2)$ is upper-bounded by $\frac{c}{n\epsilon^2}$, where $c=\frac{p^{6r}}{4}$. Hence, we have

$$\mathbb{E}\{P(E_{d,2}\cap E_1'^{\,c}\cap E_2'^{\,c})\}\le \frac{c}{n\epsilon^2}+4(1-\frac{c}{n\epsilon^2})\sum_{s=0}^{r-1}2^{n(H(Z|[Z]_s)+o(\epsilon))}2^{k(H(W|Q,[W]_s)+o(\epsilon))}p^{-n(r-s)}.$$

Therefore, $\mathbb{E}\{P(E_{d,2} \cap E_1'^c \cap E_2'^c)\}$ tends to zero as $n \to \infty$, if for all $s \in [0 : r-1]$,

$$\frac{k}{n} H(W|Q, [W]_s) < \log_2 p^{(r-s)} - H(Z|[Z]_s) - o(\epsilon). \tag{A.24}$$

Next, we use (A.24) to show that the bounds in (A.18) are redundant except the following:

$$R_i + \frac{k}{n} H(W_i|Q) = \log_2 p^r. \tag{A.25}$$

For that, we compare (A.25) with the bounds in (A.18) for different values of s. Noting that $H(W_i|Q) = H([W_i]_s|Q) + H(W_i|Q[W_i]_s)$, it is sufficient to show that $\frac{k}{n} H(W_i|Q, [W_i]_s) \le \log_2 p^{r-s}$. To show this inequality, we first prove that

$$H(W_i|Q, [W_i]_s) \le H(W_1 + W_2|Q, [W_1 + W_2]_s), \ i = 1, 2, \ 0 \le s \le r. \tag{A.26}$$

Then, using (A.24), we get $\frac{k}{n} H(W_i|Q, [W_i]_s) \le \log_2 p^{r-s}$. In what follows, we prove (A.26). We have

$$\begin{aligned}
H(W_1 + W_2|Q,[W_1 + W_2]_s) &= H(W_1 + W_2|Q, [[W_1]_s + [W_2]_s]_s) \\
&\ge H(W_1 + W_2|Q, [W_1]_s, [W_2]_s) \\
&= H(W_1, W_2|Q, [W_1]_s, [W_2]_s) - H(W_1|Q, [W_1]_s, [W_2]_s, W_1 + W_2) \\
&\overset{(a)}{=} H(W_2|Q, [W_2]_s) + H(W_1|Q, [W_1]_s) - H(W_1|Q, [W_1]_s, [W_2]_s, W_1 + W_2) \\
&\overset{(b)}{=} H(W_2|Q, [W_2]_s) + I(W_1; W_1 + W_2|Q, [W_1]_s, [W_2]) \\
&\ge H(W_2|Q, [W_2]_s),
\end{aligned}$$

where (a) and (b) hold because of the Markov chain $W_1 \leftrightarrow Q \leftrightarrow W_2$. Similarly, we can show that $H(W_1 + W_2|Q, [W_1 + W_2]_s) \ge H(W_1|Q, [W_1]_s)$.

Finally, using (A.25) and (A.24) the following holds

$$R_i \geq \log_2 p^r - \min_{0\leq s\leq r-1} \frac{H(W_i|Q)}{H(W_1+W_2|Q,[W_1+W_2]_s)}(\log_2 p^{(r-s)} - H(Z|[Z]_s)), \quad \text{(A.27)}$$

where we minimize the above bound over all PMFs of the form

$$P_{QW_1V_1W_2V_2} = P_Q \prod_i \left(P_{V_i|Q} P_{W_i|Q}\right),$$

such that $p(q)$ is a rational number for all $q \in \mathcal{Q}$. Since rational numbers are dense in $\mathbb{R}$, one can consider arbitrary PMF $p(q)$. Lastly, in the next lemma, we show that the cardinality bound $|\mathcal{Q}| \leq r$ is sufficient to optimize (A.27).

Lemma 26. *The cardinality of $\mathcal{Q}$ is bounded by $|\mathcal{Q}| \leq r$.*

Proof. Note that (A.24) and (A.25) give an alternative characterization of the achievable region. Using these equations, observe that this region is convex in $\mathbb{R}^2$. As a result, we can characterize the achievable region by its supporting hyper-planes. Let $\bar{R}_i := \log_2 p^r - R_i, i = 1, 2$. Using (A.27) for any $0 \leq \alpha \leq 1$ the corresponding supporting hyper-plane is characterized by

$$\begin{aligned}\big(\alpha\bar{R}_1 + (1-\alpha)\bar{R}_2\big)H(W|Q,[W]_s) \\ - \Big(\alpha H(W_1|Q) + (1-\alpha)H(W_2|Q)\Big)\Big(\log_2 p^{(r-s)} - H(Z|[Z]_s)\Big) \leq 0,\end{aligned} \quad \text{(A.28)}$$

where $s \in [0, r-1]$. We use the support lemma for the above inequalities to bound $|\mathcal{Q}|$. To this end, we first show that the left-hand side of these inequalities are continuous functions of conditional PMF's of W_1 and W_2 given Q. Let $\mathscr{P}_r$ denote the set of all product PMF's on $\mathbb{Z}_{p^r} \times \mathbb{Z}_{p^r}$. Note $\mathscr{P}_r$ is a compact set. Fix $q \in \mathcal{Q}$. Denote $f(p(w_1|q)p(w_2|q)) = \alpha H(W_1|Q=q) + (1-\alpha)H(W_2|Q=q)$ and $g_s(p(w_1|q)p(w_2|q)) = H(W_1+W_2|Q=q,[W_1+W_2]_s)$, where $s \in [0:r-1]$. We show that $f(\cdot), g_s(\cdot)$ are real

valued continuous functions of $\mathscr{P}_r$. Since the entropy function is continuous then so is f. We can write $g_s(p(w_1|q)p(w_2|q)) = H(W_1+W_2|Q=q) - H([W_1+W_2]_s|Q=q)$. Note that $[\cdot]_s$ is a continuous function from $\mathscr{P}_r$ to $\mathscr{P}_r$. This implies that $H([\cdot]_s)$ is also continuous. So g_s is continuous. As a result, the left-hand side of the bounds in (A.28) are real valued continuous functions of $\mathscr{P}_r$. Therefore, we can apply the support lemma [5]. Since there are r bounds for different values of s, then $|\mathcal{Q}| \leq r$. $\square$

A.6 Proof of Theorem II.3

Fix positive integer n, and define $l \triangleq cn$, and $k \triangleq \tilde{c}n$, where $\tilde{c}$ and c are positive real numbers such that l and k are integers.

Codebook Generation We use two nested QGC's, one for each encoder. The codebook for Encoder 1 is an (n,k,l) nested QGC (as in Definition 5) with random variables (W_1, V_1, Q). Let $\mathcal{C}_{I,1}, \bar{\mathcal{C}}_1$, and $\mathcal{C}_{O,1}$ denote the corresponding inner code, shift code and the outer code (as in Definition 5), respectively. The codebook for Encoder 2 is an (n,k,l) nested QGC with random variables (W_2, V_2, Q), inner code $\mathcal{C}_{I,2}$, shift code $\bar{\mathcal{C}}_2$, and outer code $\mathcal{C}_{O,2}$. For the decoder, we use $\mathcal{C}_{O,1} + \mathcal{C}_{O,2}$ as a codebook. Conditioned on Q, the random variables (W_1, W_2, V_1, V_2) are mutually independent.

The nested QGCs and $\mathcal{C}_d$ have identical generator matrices but different translations and index random variables. Note that each nested QGC has two generator matrices/translations, one for the inner code and one for the shift code as in Definition 5. The generator matrix and the translation for the inner codes $\mathcal{C}_{I,i}, i = 1,2$, are denoted by $\mathbf{G}$ and $\mathbf{b}$, respectively. The generator matrix and the translation used for shift code $\mathcal{C}_{I,i}$, are denoted by $\bar{\mathbf{G}}$ and $\bar{\mathbf{b}}_i$, respectively, where $i = 1,2$. The elements of $\mathbf{G}, \bar{\mathbf{G}}, \mathbf{b}$, and $\bar{\mathbf{b}}_i, i = 1,2$ are generated randomly and independently from $\mathbb{Z}_{p^r}$. By R_i denote the rate of $\bar{\mathcal{C}}_i$, and let $R_{I,i}$ be the rate of $\mathcal{C}_{I,i}$, where $i = 1,2$.

Codebook Generation: We use two nested QGC's, one for each encoder. The

codebook used for the ith encoder is $(\mathcal{C}_{O,i}, \mathcal{C}_{I,i})$. With this notation, the random variables corresponding to $\mathcal{C}_{O,i}$ are $(W_i, V_i, Q), i = 1, 2$. For the decoder, we use $\mathcal{C}_{O,1} + \mathcal{C}_{O,2}$ as a codebook.

Encoding: Index the codewords of $\bar{\mathcal{C}}_i, i = 1, 2$. Upon receiving a message index θ_i, the ith encoder finds the codeword $\mathbf{c}_i \in \bar{\mathcal{C}}_i$ with that index. Then it finds $\mathbf{c}_{I,i} \in \mathcal{C}_{I,i}$ such that $\mathbf{c}_i + \mathbf{c}_{I,i}$ is ϵ-typical with respect to P_{X_i}. If such codeword was found, the encoder i sends $\mathbf{x}_i = \mathbf{c}_i + \mathbf{c}_{I,i}, i = 1, 2$. Otherwise, an error event $E_i, i = 1, 2$ is declared.

Decoding: The channel takes $\mathbf{x}_1$ and $\mathbf{x}_2$ and produces $\mathbf{y}$. Upon receiving $\mathbf{y}$ from the channel, the decoder wishes to decode $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$. It finds $\tilde{\mathbf{x}} \in \mathcal{C}_{O,1} + \mathcal{C}_{O,2}$ such that $\tilde{\mathbf{x}}$ and $\mathbf{y}$ are jointly $\tilde{\epsilon}$-typical with respect to the distribution $P_{X_1+X_2,Y}$. An error event E_d is declared, if no unique $\tilde{\mathbf{x}}$ was found.

Probability of Error: Let $(f_1(\cdot), f_2(\cdot))$ and $g(\cdot, \cdot)$ denote the encoding and decoding functions corresponding to the above coding scheme. The overall error event is defined as

$$E \triangleq \{g(Y^n) \neq f_1(M_1) + f_2(M_2)\}$$

For the achievability, we need to show that $P(E)$ can be made arbitrary small for sufficiently large n. If (X_1^n, X_2^n) denote the outputs of the encoders, define an error event E_c as the event in which $(X_1^n, X_2^n) \notin A_\epsilon^{(n)}(X_1, X_2)$. Next, using the aforementioned encoding and decoding error events we have

$$P(E) \leq P(E_1 \cup E_2 \cup E_d \cup E_c) + P(E|E_1^c \cap E_2^c \cap E_d^c \cap E_c^c)$$

Using standard arguments for typical sequences, we can show that when there is no encoding and decoding error (i.e., $E_1^c \cap E_2^c \cap E_d^c \cap E_c^c$) the error probability $P(E|E_1^c \cap E_2^c \cap E_d^c \cap E_c^c)$ approaches 0 as $n \to \infty$. As a result, the second term above is sufficiently small for large enough n. Therefore, for sufficiently large n and from the union bound

on the first term we obtain,

$$P(E) \leq P(E_1) + P(E_2) + P(E_d) + P(E_c) + \epsilon$$

We need to find conditions for which the probability of the error events E_1, E_2, E_d and E_c approach zero. For any $\mathbf{a} \in \mathbb{Z}_{p^r}^k$ and $\bar{\mathbf{a}} \in \mathbb{Z}_{p^r}^l$ define the map $\phi(\mathbf{a}, \bar{\mathbf{a}}) = \mathbf{aG} + \bar{\mathbf{a}}\bar{\mathbf{G}}$. By $\Phi(\cdot, \cdot)$ denote the map ϕ whose matrices are selected randomly and uniformly.

A.6.1 Analysis of E_1, E_2

For any sequence $\mathbf{v}_i \in \mathcal{V}_i$ define

$$\lambda_i(\mathbf{v}_i) = \sum_{\mathbf{w}_i \in \mathcal{W}_i} \sum_{\mathbf{x}_i \in A_\epsilon^{(n)}(X_i)} \mathbb{1}\{\mathbf{x}_i = \phi(\mathbf{w}_i, \mathbf{v}_i) + \mathbf{b} + \bar{\mathbf{b}}_i\},$$

where $i = 1, 2$. Therefore, E_i occurs if $\lambda_i(\mathbf{v}_i) = 0$. For more convenience, we weaken the definition of event E_i. We say E_i occurs, if $\lambda_i(\mathbf{v}_i) < \frac{1}{2}E(\lambda_i(v_i))$. Using Lemma 5 we can show that $P(E_i) \to 0$ as $n \to \infty$, if

$$\frac{k}{n} H([W_i]_t|Q) \geq \log_2 p^t - H([X_i]_t) + \gamma(\epsilon),\ i = 1, 2,\ 1 \leq t \leq r, \tag{A.29}$$

where $\lim_{\epsilon \to 0} \gamma(\epsilon) = 0$.

A.6.2 Analysis of E_c

Define the set

$$\mathcal{E} \triangleq \{(\mathbf{x}_1, \mathbf{x}_2) \in A_\epsilon^{(n)}(X_1) \times A_\epsilon^{(n)}(X_2) : (\mathbf{x}_1, \mathbf{x}_2) \notin A_\epsilon^{(n)}(X_1, X_2)\}.$$

Therefore, probability of E_c can be written as

$$P(E_c|E_1^c \cap E_2^c) = \sum_{(\mathbf{x}_1,\mathbf{x}_2)\in\mathcal{E}} P(e_1(\Theta_1) = \mathbf{x}_1, e_2(\Theta_2) = \mathbf{x}_2),$$

where e_i is the output of the ith encoder, and Θ_i is the random message to be transmitted by encoder i, where $i = 1, 2$. By $P(\mathbf{v}_i, \mathbf{w}_i, \mathbf{x}_i)$ denote the probability that $(\mathbf{v}_i, \mathbf{w}_i, \mathbf{x}_i)$ is selected at the ith encoder. Then, $P(\mathbf{v}_i, \mathbf{w}_i, \mathbf{x}_i) = \frac{1}{|\mathcal{V}_i|}\frac{1}{\lambda_i(\mathbf{v}_i)}\mathbb{1}\{\phi(\mathbf{w}_i, \mathbf{v}_i) + \mathbf{b} + \bar{\mathbf{b}}_i = \mathbf{x}_i\}$. By the definition of $\phi_1(\cdot)$ and $\phi_2(\cdot)$, we have

$$P(E_c|E_1^c \cap E_2^c) = \sum_{(\mathbf{x}_1,\mathbf{x}_2)\in\mathcal{E}} \prod_{i=1}^{2}\Bigg[\sum_{\mathbf{v}_i\in\mathcal{V}_i}\sum_{\mathbf{w}_i\in\mathcal{W}_i}\frac{1}{|\mathcal{V}_i|}\frac{1}{\lambda_i(\mathbf{v}_i)}\mathbb{1}\Big\{\mathbf{x}_i = \phi_i(\mathbf{w}_i, \mathbf{v}_i) + \mathbf{b} + \bar{\mathbf{b}}_i\Big\}\Bigg]$$

Since there is no encoding error (for the modified version), then $\lambda_i(\mathbf{v}_i) \geq \frac{1}{2}E[\lambda_i(\mathbf{v}_i)], i = 1, 2$. Therefore, replacing $\lambda_i(\mathbf{v}_i)$ in the above expression with $\frac{1}{2}E[\lambda_i(\mathbf{v}_i)]$ gives an upper bound on $P(E_c|E_1^c \cap E_2^c)$. Next, we take expectation over all ϕ_1 and ϕ_2. We have

$$\begin{aligned}
\mathbb{E}\{P(E_c|E_1^c \cap E_2^c)\} &\leq \sum_{(\mathbf{x}_1,\mathbf{x}_2)\in\mathcal{E}}\sum_{\mathbf{v}_i\in\mathcal{V}_i,i=1,2}\sum_{\mathbf{w}_i\in\mathcal{W}_i,i=1,2}\Bigg[\prod_{j=1}^{2}\frac{4}{|\mathcal{V}_j|E[\lambda_j(\mathbf{v}_j)]}\Bigg] \\
&\qquad P\{\mathbf{x}_i = \Phi_i(\mathbf{w}_i, \mathbf{v}_i) + \mathbf{B} + \bar{\mathbf{B}}_i, i = 1, 2\} \\
&\overset{(a)}{=} \sum_{(\mathbf{x}_1,\mathbf{x}_2)\in\mathcal{E}}\sum_{\mathbf{v}_i\in\mathcal{V}_i,i=1,2}\sum_{\mathbf{w}_i\in\mathcal{W}_i,i=1,2}\Bigg[\prod_{j=1}^{2}\frac{4}{|\mathcal{V}_j|E[\lambda_j(\mathbf{v}_j)]}\Bigg]p^{-2nr} \\
&= \sum_{(\mathbf{x}_1,\mathbf{x}_2)\in\mathcal{E}}|\mathcal{W}_1||\mathcal{W}_2|\frac{4}{E[\lambda_1(\mathbf{v}_1)]E[\lambda_2(\mathbf{v}_2)]}p^{-2nr}. \qquad \text{(A.30)}
\end{aligned}$$

Note that (a) is because $\mathbf{B}_1$ and $\mathbf{B}_2$ are independent random vectors with uniform distribution over $\mathbb{Z}_{p^r}^n$. From the definition of $\lambda_j(\mathbf{v}_j), j = 1, 2$, we have

$$E[\lambda_j(\mathbf{v}_j)] = |\mathcal{W}_j||A_\epsilon^{(n)}(X_i)|p^{-nr}$$

As a result of the above equation and (A.30),

$$\mathbb{E}\{P(E_c|E_1^c \cap E_2^c)\} \leq \sum_{(\mathbf{x}_1,\mathbf{x}_2)\in\mathcal{E}} 4|A_\epsilon^{(n)}(X_1)|^{-1}|A_\epsilon^{(n)}(X_2)|^{-1}$$

There exists a continuous function $\delta(\epsilon) > 0$ with $\delta(0) = 0$ such that for any $\mathbf{x}_i \in A_\epsilon^{(n)}(X_i)$, we have $P_{X_i}^n(\mathbf{x}_i) \geq |A_\epsilon^{(n)}(X_i)|^{-1}2^{-\delta(\epsilon)}$. Thus,

$$\mathbb{E}\{P(E_c \cap E_1^c \cap E_2^c)\} \leq \sum_{(\mathbf{x}_1,\mathbf{x}_2)\in\mathcal{E}} P_{X_1}^n(\mathbf{x}_1)P_{X_2}^n(\mathbf{x}_2)2^{n2\delta(\epsilon)} = 2^{n2\delta(\epsilon)}P_{X_1X_2}^n(\mathcal{E}).$$

Thus, $\mathbb{E}\{P(E_c|E_1^c \cap E_2^c)\} \to 0$ as $n \to \infty$.

A.6.3 Analysis of E_d

In what follows, to make the analysis tractable, we define an alternative decoding error. Upon receiving $\mathbf{y}$, the decoder finds $\tilde{\mathbf{w}} \in A_\epsilon^{(n)}(W_1+W_2)$ and $\tilde{\mathbf{v}} \in A_\epsilon^{(n)}(V_1+V_2)$ such that $\phi(\tilde{\mathbf{w}},\tilde{\mathbf{v}})+2\mathbf{b}+\bar{\mathbf{b}}_1+\bar{\mathbf{b}}_2$ is jointly typical with $\mathbf{y}$ with respect to $P_{X_1+X_2,Y}$. For the alternative decoder, we define a new decoding error. A decoding error E_d' occurs, if $(\tilde{\mathbf{w}},\tilde{\mathbf{v}})$ is not unique. With this definition $E_d \subseteq E_d'$. Because, the the mapping $\mathbf{x}_i = \phi(\mathbf{w}_i,\mathbf{v}_i) + \mathbf{b} + \bar{\mathbf{b}}_i$ is not necessarily injective. Note that the new decoder is required to decode $\mathbf{w}_1 + \mathbf{w}_2$ and $\mathbf{v}_1 + \mathbf{v}_2$. This is a more restrictive condition than decoding $\mathbf{x}_1 + \mathbf{x}_2$. Therefore, it is sufficient to show that $P(E_d') \to 0$ as $n \to \infty$. In what follows, we provide an upper bound on $P(E_d')$.

Since the the probability of the encoding errors E_1, E_2 and E_c are sufficiently small, then $P(E_d') \approx P(E_d' \cap E_1^c \cap E_2^c \cap E_c^c)$. We show that this probability approaches zero as $n \to \infty$. Fix $\phi, \mathbf{b}$ and $\bar{\mathbf{b}}_i, i = 1, 2$. Note that By $P(\mathbf{v}_i, \mathbf{w}_i, \mathbf{x}_i)$ denote the probability that $(\mathbf{v}_i, \mathbf{w}_i, \mathbf{x}_i)$ is selected at the ith encoder. Then, $P(\mathbf{v}_i, \mathbf{w}_i, \mathbf{x}_i) = \frac{1}{|\mathcal{V}_i|}\frac{1}{\lambda_i(\mathbf{v}_i)}\mathbb{1}\{\phi(\mathbf{w}_i,\mathbf{v}_i) + \mathbf{b} + \bar{\mathbf{b}}_i = \mathbf{x}_i\}$.

Then the probability of $E'_d \cap E_1^c \cap E_2^c \cap E_c^c$ equals

$$P(E'_d \cap E_1^c \cap E_2^c \cap E_c^c) = \left[\prod_{j=1}^{2} \sum_{\mathbf{v}_j \in \mathcal{V}_j} \sum_{\mathbf{w}_j \in \mathcal{W}_j} \mathbb{1}\Big\{\lambda_i(\mathbf{v}_i) \geq 1/2 \; E(\lambda_i(\mathbf{v}_i)), i = 1, 2\Big\}\right]$$
$$\sum_{(\mathbf{x}_1, \mathbf{x}_2) \in A_\epsilon^{(n)}(X_1, X_2)} \sum_{\mathbf{y} \in \mathcal{Y}^n} P(\mathbf{v}_i, \mathbf{w}_i, \mathbf{x}_i, i = 1, 2)$$
$$P_{Y|X_1X_2}^n(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) P(E_d \mid E_1^c \cap E_2^c \cap E_c^c, \mathbf{y}, \mathbf{x}_i, \mathbf{v}_i, \mathbf{w}_i, i = 1, 2)$$

Next, we bound $P(E'_d \mid E_1^c \cap E_2^c \cap E_c^c, \mathbf{y}, \mathbf{x}_i, \mathbf{v}_i, \mathbf{w}_i, i = 1, 2)$, and $P(\mathbf{v}_i \mathbf{w}_i, \mathbf{x}_i, i = 1, 2)$.

$$P(E'_d \mid E_1^c \cap E_2^c \cap E_c^c, \mathbf{y}, \mathbf{x}_i, \mathbf{v}_i, \mathbf{w}_i, i = 1, 2) =$$
$$\mathbb{1}\{\exists \; (\tilde{\mathbf{w}}, \tilde{\mathbf{v}}) \in \mathcal{W} \times \mathcal{V} : (\tilde{\mathbf{w}}, \tilde{\mathbf{v}}) \neq (\mathbf{w}_1 + \mathbf{w}_2, \mathbf{v}_1 + \mathbf{v}_2), \phi(\tilde{\mathbf{w}}, \tilde{\mathbf{v}}) + 2\mathbf{b} + \bar{\mathbf{b}}_1 + \bar{\mathbf{b}}_2 \in A_{\epsilon'}^n(Z|\mathbf{y})\},$$

where $\mathcal{W} \triangleq A_\epsilon^{(n)}(W_1 + W_2), \mathcal{V} \triangleq A_\epsilon^{(n)}(V_1 + V_2)$, and $Z \triangleq X_1 + X_2$. Using the union bound, we have

$$P(E'_d \mid E_1^c \cap E_2^c \cap E_c^c, \mathbf{y}, \mathbf{x}_i, \mathbf{v}_i, \mathbf{w}_i, i = 1, 2) \leq \tag{A.31}$$
$$\sum_{\substack{\tilde{\mathbf{w}} \in \mathcal{W} \\ \tilde{\mathbf{w}} \neq \mathbf{w}_1 + \mathbf{w}_2}} \sum_{\substack{\tilde{\mathbf{v}} \in \mathcal{V} \\ \tilde{\mathbf{v}} \neq \mathbf{v}_1 + \mathbf{v}_2}} \sum_{\tilde{\mathbf{z}} \in A_{\epsilon'}^{(n)}(Z|\mathbf{y})} \mathbb{1}\{\phi(\tilde{\mathbf{w}}, \tilde{\mathbf{v}}) + 2\mathbf{b} + \bar{\mathbf{b}}_1 + \bar{\mathbf{b}}_2 = \tilde{\mathbf{z}}\}$$

Note that $P(\mathbf{v}_i, \mathbf{w}_i, \mathbf{x}_i, i = 1, 2) = \prod_{i=1,2} P(\mathbf{v}_i, \mathbf{w}_i, \mathbf{x}_i)$. Since there is no encoding error, $\lambda_i(\mathbf{v}_i) \geq \frac{1}{2} E(\lambda_i(\mathbf{v}_i))$. As a result,

$$P(\mathbf{v}_i, \mathbf{w}_i, \mathbf{x}_i) \leq \frac{1}{|\mathcal{V}_i|} \frac{2}{E(\lambda_i(\mathbf{v}_i))} \mathbb{1}\{\phi(\mathbf{w}_i, \mathbf{v}_i) + \mathbf{b} + \bar{\mathbf{b}}_i = \mathbf{x}_i\} \tag{A.32}$$

Therefore, using (A.32), we have

$$P(E'_d \cap E_1^c \cap E_2^c \cap E_c^c) \leq \sum_{(\mathbf{x}_1,\mathbf{x}_2)\in A_\epsilon^{(n)}(X_1,X_2)} \Big[\prod_{j=1}^{2} \sum_{\mathbf{v}_j\in\mathcal{V}_j} \sum_{\mathbf{w}_j\in\mathcal{W}_j} \mathbb{1}\Big\{\lambda_j(\mathbf{v}_j) \geq 1/2\ E(\lambda_j(\mathbf{v}_j))\Big\}$$
$$\frac{1}{|\mathcal{V}_j|}\frac{2}{E(\lambda_i(\mathbf{v}_j))}\mathbb{1}\{\phi(\mathbf{w}_j,\mathbf{v}_j)+\mathbf{b}+\bar{\mathbf{b}}_j=\mathbf{x}_j\}\Big]$$
$$\sum_{\mathbf{y}\in\mathcal{Y}^n} P^n_{Y|X_1X_2}(\mathbf{y}|\mathbf{x}_1,\mathbf{x}_2) P(E'_d \mid E_1^c\cap E_2^c\cap E_c^c, \mathbf{y}, \mathbf{x}_i, \mathbf{v}_i, \mathbf{w}_i, i=1,2)$$
$$\leq \sum_{(\mathbf{x}_1,\mathbf{x}_2)\in A_\epsilon^{(n)}(X_1,X_2)} \left[\prod_{j=1}^{2} \sum_{\mathbf{v}_j\in\mathcal{V}_j} \sum_{\mathbf{w}_j\in\mathcal{W}_j} \frac{1}{|\mathcal{V}_j|}\frac{2}{E(\lambda_i(\mathbf{v}_j))}\mathbb{1}\{\phi(\mathbf{w}_j,\mathbf{v}_j)+\mathbf{b}+\bar{\mathbf{b}}_j=\mathbf{x}_j\}\right]$$
$$\sum_{\mathbf{y}\in\mathcal{Y}^n} P^n_{Y|X_1X_2}(\mathbf{y}|\mathbf{x}_1,\mathbf{x}_2) P(E'_d \mid E_1^c\cap E_2^c\cap E_c^c, \mathbf{y}, \mathbf{x}_i, \mathbf{v}_i, \mathbf{w}_i, i=1,2) \tag{A.33}$$

The last inequality follows by eliminating the indicator function on $\{\lambda_i(\mathbf{v}_i) \geq 1/2\ E(\lambda_i(\mathbf{v}_i)), i=1,2\Big\}$. Note that for jointly ϵ-typical sequences $\mathbf{x}_1, \mathbf{x}_2$ and large enough n, we have $P(\mathbf{Y}^n \notin A_{\tilde{\epsilon}}^{(n)}(Y|\mathbf{x}_1,\mathbf{x}_2)) \leq \frac{c}{n\tilde{\epsilon}^2}$, where c is a constant. This follows from the standard arguments on typical sets. Thus, using (A.33) and (A.31) we get

$$P(E'_d \cap E_1^c \cap E_2^c \cap E_c^c) \leq \frac{c}{n\tilde{\epsilon}^2} +$$
$$\sum_{(\mathbf{x}_1,\mathbf{x}_2)\in A_\epsilon^{(n)}(X_1,X_2)} \left[\prod_{j=1}^{2} \sum_{\mathbf{v}_j\in\mathcal{V}_j} \sum_{\mathbf{w}_j\in\mathcal{W}_j} \frac{1}{|\mathcal{V}_j|}\frac{2}{E(\lambda_i(\mathbf{v}_j))}\mathbb{1}\{\phi(\mathbf{w}_j,\mathbf{v}_j)+\mathbf{b}+\bar{\mathbf{b}}_j=\mathbf{x}_j\}\right]$$
$$\sum_{\mathbf{y}\in A_{\tilde{\epsilon}}^n(Y|\mathbf{x}_1,\mathbf{x}_2)} P^n_{Y|X_1X_2}(\mathbf{y}|\mathbf{x}_1,\mathbf{x}_2) \sum_{\substack{\tilde{\mathbf{w}}\in\mathcal{W}\\ \tilde{\mathbf{w}}\neq\mathbf{w}_1+\mathbf{w}_2}} \sum_{\substack{\tilde{\mathbf{v}}\in\mathcal{V}\\ \tilde{\mathbf{v}}\neq\mathbf{v}_1+\mathbf{v}_2}} \sum_{\tilde{\mathbf{z}}\in A_{\epsilon'}^{(n)}(Z|\mathbf{y})} \mathbb{1}\{\phi(\tilde{\mathbf{w}},\tilde{\mathbf{v}})+2\mathbf{b}+\bar{\mathbf{b}}_1+\bar{\mathbf{b}}_2=\tilde{\mathbf{z}}\}$$

Next, we take the average of the above expression over all maps ϕ, and all vectors

$\mathbf{b}, \bar{\mathbf{b}}_i, i = 1, 2.$

$$\mathbb{E}\{P(E'_d \cap E_1^c \cap E_2^c \cap E_c^c)\} \leq \frac{c}{n\tilde{\epsilon}^2} + \left[\prod_{j=1}^{2} \sum_{\mathbf{v}_j \in \mathcal{V}_j} \sum_{\mathbf{w}_j \in \mathcal{W}_j} \frac{1}{|\mathcal{V}_j|} \frac{2}{E(\lambda_j(\mathbf{v}_j))}\right]$$
$$\sum_{(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}) \in A_\epsilon^{(n)}(X_1, X_2, Y)} P^n_{Y|X_1X_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) \sum_{\substack{\tilde{\mathbf{w}} \in \mathcal{W} \\ \tilde{\mathbf{w}} \neq \mathbf{w}_1 + \mathbf{w}_2}} \sum_{\substack{\tilde{\mathbf{v}} \in \mathcal{V} \\ \tilde{\mathbf{v}} \neq \mathbf{v}_1 + \mathbf{v}_2}} \sum_{\tilde{\mathbf{z}} \in A_{\epsilon'}^{(n)}(Z|\mathbf{y})}$$
$$P\{\tilde{z} = \Phi(\tilde{\mathbf{w}}, \tilde{\mathbf{v}}) + 2\mathbf{B} + \bar{\mathbf{B}}_1 + \bar{\mathbf{B}}_1, x_1 = \Phi(\mathbf{w}_1, \mathbf{v}_1) + \mathbf{B} + \bar{\mathbf{B}}_1, x_2 = \Phi(\mathbf{w}_2, \mathbf{v}_2) + \mathbf{B} + \bar{\mathbf{B}}_1\}$$

Notice that $\mathbf{B}, \bar{\mathbf{B}}_1$, and are $\bar{\mathbf{B}}_1$ are uniform over $\mathbb{Z}_{p^r}^n$ and independent of other random variables. Hence, the innermost term in the above summations is simplified to

$$p^{-2nr} P\{\tilde{\mathbf{z}} - \mathbf{x_1} - \mathbf{x_2} = \Phi(\tilde{\mathbf{w}} - (\mathbf{w_1} + \mathbf{w_2}), \tilde{\mathbf{v}} - (\mathbf{v_1} + \mathbf{v_2}))\} \tag{A.34}$$

Using Lemma 29, if $\tilde{\mathbf{w}} - (\mathbf{w_1} + \mathbf{w_2}), \tilde{\mathbf{v}} - (\mathbf{v_1} + \mathbf{v_2}) \in H_s^k \backslash H_{s+1}^k$ the expression in (A.34) equals

$$p^{-2nr} p^{-n(r-s)} \mathbb{1}\{\tilde{z} - \mathbf{x_1} - \mathbf{x_2} \in H_s^n\},$$

where $0 \leq s \leq r-1$. Therefore, $\mathbb{E}\{P(E'_d \cap E_1^c \cap E_2^c \cap E_c^c)\}$ is upper-bounded as

$$\mathbb{E}\{P(E'_d \cap E_1^c \cap E_2^c \cap E_c^c)\} \leq \frac{c}{n\tilde{\epsilon}^2} +$$
$$\left[\prod_{j=1}^{2} \sum_{\mathbf{v}_j \in \mathcal{V}_j} \sum_{\mathbf{w}_j \in \mathcal{W}_j} \frac{1}{|\mathcal{V}_j|} \frac{2}{E(\lambda_j(\mathbf{v}_j))}\right] \sum_{(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}) \in A_\epsilon^{(n)}(X_1, X_2, Y)} P^n_{Y|X_1X_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2)$$
$$\sum_{s=0}^{r-1} \sum_{\substack{\tilde{\mathbf{w}} \in \mathcal{W} \\ \tilde{\mathbf{w}} - (\mathbf{w_1} + \mathbf{w_2}) \in H_s^k}} \sum_{\substack{\tilde{\mathbf{v}} \in \mathcal{V} \\ \tilde{\mathbf{v}} - (\mathbf{v_1} + \mathbf{v_2}) \in H_s^k}} \sum_{\substack{\tilde{z} \in A_\epsilon^n(Z|y) \\ \tilde{\mathbf{z}} - \mathbf{x_1} - \mathbf{x_2} \in H_s^n}} p^{-2nr} p^{-n(r-s)} \tag{A.35}$$

Note the most inner term in the above summations does not depend on the value of $\tilde{\mathbf{z}}, \tilde{\mathbf{v}}$ and $\tilde{\mathbf{w}}$. Hence, we replace those summations by the size of the corresponding subsets. Using Lemma 30 we can bound the size of these subsets and get the following

bound on the probability of error

$$\mathbb{E}\{P(E'_d \cap E_1^c \cap E_2^c \cap E_c^c)\} \leq \frac{c}{n\tilde{\epsilon}^2} + \left[\prod_{j=1}^{2} \sum_{\mathbf{v}_j \in \mathcal{V}_j} \sum_{\mathbf{w}_j \in \mathcal{W}_j} \frac{1}{|\mathcal{V}_j|} \frac{2}{E(\lambda_j(\mathbf{v}_j))}\right] \sum_{(\mathbf{x}_1,\mathbf{x}_2,\mathbf{y}) \in A_\epsilon^{(n)}(X_1,X_2,Y)} P^n_{Y|X_1X_2}(\mathbf{y}|\mathbf{x}_1,\mathbf{x}_2)$$
$$\sum_{s=0}^{r-1} 2^{k(H(W|Q,[W]_s)+\eta_1(\epsilon))} 2^{l(H(V|Q,[V]_s)+\eta_2(\epsilon))}\ 2^{n(H(Z|Y[Z]_s)+\eta_3(\epsilon))} p^{-2nr} p^{-n(r-s)},$$

where $W = W_1 + W_2, V = V_1 + V_2$, and $\lim_{\epsilon\to 0} \eta_i(\epsilon) = 0, i = 1, 2, 3$. Note that $E(\lambda_i(\mathbf{v}_i)) = |\mathcal{W}_i||A_\epsilon^{(n)}(X_i)|p^{-nr}, i = 1, 2$. As the terms in the above expression do not depend on the values of $\mathbf{w}_i, \mathbf{v}_i, \mathbf{x}_i, i = 1, 2$ and $\mathbf{y}$, we can replace the summations over them with the corresponding sets. As a result, we have

$$\mathbb{E}\{P(E'_d \cap E_1^c \cap E_2^c \cap E_c^c)\} \leq \frac{c}{n\epsilon^2} + 4\sum_{s=0}^{r-1} p^{-n(r-s)} 2^{kH(W|Q,[W]_s)} 2^{lH(V|Q,[V]_s)}\ 2^{n(H(Z|Y,[Z]_s)+\delta'(\epsilon))},$$

where $\lim_{\epsilon\to 0} \delta'(\epsilon) = 0$. Therefore, the right-hand side of the above inequality approaches zero as $n \to \infty$, if the following bounds hold:

$$\frac{k}{n} H(W|Q,[W]_s) + \frac{l}{n} H(V|Q,[V]_s) \leq \log_2 p^{r-s} - H(Z|Y[Z]_s) - \delta(\epsilon), \quad \text{for } 0 \leq s \leq r-1. \tag{A.36}$$

Next, we apply the Fourier-Motzkin technique [5] to eliminate $\frac{k}{n}$ from (A.29) and (A.36). We get

$$\frac{l}{n} H(V|Q,[V]_s) \leq \log_2 p^{r-s} - H(Z|Y[Z]_s) - \frac{H(W|Q,[W]_s)}{H([W_i]_t|Q)}(\log_2 p^t - H([X_i]_t)) - o(\epsilon),$$

where $i = 1, 2,\ \ 0 \leq s \leq r-1$, and $1 \leq t \leq r$. Note by definition

$$R_i = \frac{1}{n}\log_2 |\bar{\mathcal{C}}_i| \leq \frac{1}{n}\log_2 |\mathcal{V}_i| \leq \frac{l}{n} H(V_i|Q).$$

Therefore, we obtain the bounds in the theorem. Using the same argument as in Lemma 26, we can bound the cardinality of Q by $|\mathcal{Q}| \leq r^2$. This completes the proof.

A.7 Proof of Lemma 7

Proof. Consider the bound on the sum-rate given in (2.16). The set of all (R_1, R_2) satisfying only this bound is an outer-bound for $\mathscr{R}_{GP}$. The time-sharing random variable Q is trivial for this outer-bound, because there is only one inequality on the rates, and because of the cost constraints $\mathbb{E}\{c_i(X_i)\} = 0, i = 1, 2$. For any distribution $P \in \mathscr{P}_{GP}$, we obtain

$$\begin{aligned} R_1 + R_2 &\leq I(U_1, U_2; Y) - I(U_1; S_1) - I(U_2; S_2) \\ &= H(Y) - H(Y|U_1, U_2) - H(S_1) + H(S_1|U_1) - H(S_2) + H(S_2|U_2) \\ &\leq H(S_1|U_1) + H(S_2|U_2) - H(Y|U_1, U_2) - 2 \\ &= \max_{P \in \mathscr{P}_{GP}} \sum_{u_1 \in \mathcal{U}_1} \sum_{u_2 \in \mathcal{U}_2} p(u_1, u_2) \Big(H(S_1|u_1) + H(S_2|u_2) - H(Y|u_1, u_2) - 2 \Big) \end{aligned} \tag{A.37}$$

where the second inequality holds, as $H(Y) \leq 2$, and $H(S_i) = 2$ for $i = 1, 2$. In the next step, we relax the conditions in $\mathscr{P}_{GP}$, and provide an upper-bound on (A.37). For $i = 1, 2$, and any $u_i \in \mathcal{U}_i$, define $\mathscr{P}_{u_i}$ as the collection of all conditional PMFs $p(s_i, x_i|u_i)$ on $\mathbb{Z}_4^2$ such that

1. $X_i = f_i(S_i, u_i)$ for some function f_i,
2. $E(c_i(X_i)|u_i) = 0$.

In the first condition, given u_i, $f_i(s_i, u_i)$ can be thought as a function g_{u_i} of s_i. For different u_i's we have different functions $g_{u_i}(s_i)$. The second condition is implied from the cost constraint $E(c_i(X_i)) = 0$, because without loss of generality we assume

$p(u_i) > 0$ for all $u_i \in \mathcal{U}_i$. Also, note that we removed the condition that S_i is uniform over $\mathbb{Z}_4$. Hence, $\mathscr{P}_{GP}$ is a subset of the set of all PMFs of the form $P = \prod_{i=1}^{2} p(u_i)p(s_i, x_i|u_i)$, where $p(s_i, x_i|u_i) \in \mathscr{P}_{u_i}, i = 1, 2$.

As a result, (A.37) is upper-bounded by

$$R_1 + R_2 \tag{A.38}$$

$$\leq \max_{p(u_1),p(u_2)} \max_{\substack{p(s_i,x_i|u_i)\in\mathscr{P}_{u_i} \\ i=1,2}} \sum_{u_1\in\mathcal{U}_1} \sum_{u_2\in\mathcal{U}_2} p(u_1, u_2)\Big(H(S_1|u_1) + H(S_2|u_2) - H(Y|u_1, u_2) - 2\Big) \tag{A.39}$$

$$\leq \max_{u_1\in\mathcal{U}_1,u_2\in\mathcal{U}_2} \max_{\substack{p(s_i,x_i|u_i)\in\mathscr{P}_{u_i} \\ i=1,2}} \Big(H(S_1|u_1) + H(S_2|u_2) - H(Y|u_1, u_2) - 2\Big) \tag{A.40}$$

Fix $u_2 \in \mathcal{U}_2$ and $p(s_2, x_2|u_2) \in \mathscr{P}_{u_2}$. We maximize over all $u_1 \in \mathcal{U}_1$ and $p(s_1, x_1|u_1) \in \mathscr{P}_{u_1}$. Let $N = X_2 + S_2$, where X_2 and S_2 are distributed according to $p(s_2, x_2|u_2)$. For fixed $u_2 \in \mathcal{U}_2$, by $Q_{u_2} \in \mathscr{P}_{u_2}$ denote the PMF $p(s_2, x_2|u_2)$. This maximization problem is equivalent to finding

$$R(u_2, Q_{u_2}) \triangleq H(S_2|u_2) + \max_{u_1\in\mathcal{U}_1} \max_{p(s_1,x_1|u_1)\in\mathscr{P}_{u_1}} H(S_1|u_1) - H(X_1 + S_1 + N|u_1) - 2. \tag{A.41}$$

Consider the problem of PtP channel with state, where the channel is $Y = X_1+S_1+N$. It can be shown that $R(u_2, Q_{u_2}) - H(S_2|u_2)$ is an upper-bound on the capacity of this problem. We proceed by the following lemma.

Lemma 27. *The following bound holds $R(u_2, Q_{u_2}) < 1$ for all $u_2 \in \mathcal{U}_2$ and $Q_{u_2} \in \mathscr{P}_{u_2}$*

.

Proof. The proof is given in Appendix A.8. □

Finally, as a result of the above lemma the proof is completed. □

A.8 Proof of Lemma 27

Proof. Note that for any fixed $u_2 \in \mathcal{U}_2$, the distribution of N depends on the conditional PMF $p(s_1|u_1)$, and the function $x_1 = f_1(s_1, u_1)$. For any $u \in \mathcal{U}_2$ define

$$\mathcal{L}_u := \{f_2(u, s) + s : s \in \mathbb{Z}_4\}.$$

For any given $i \in \{1, 2, 3, 4\}$, define

$$\mathcal{B}_i \triangleq \{u \in \mathcal{U}_2 : |\mathcal{L}_u| = i\}.$$

Note that $\mathcal{B}_i$'s are disjoint and $\mathcal{U}_2 = \bigcup_i \mathcal{B}_i$. Depending on u_2, we consider four cases. In what follows, for each case, we derive an upper bound on (A.41). Consider the PMF $p(\omega)$ on $\mathbb{Z}_4$. For brevity, we represent this PMF by the vector $\mathbf{p} := (p(0), p(1), p(2), p(3))$.

Case 1: $u_2 \in \mathcal{B}_1$

Since $|\mathcal{L}_{u_2}| = 1$, then for all $s_2 \in \mathbb{Z}_4$ the equality $s_2 + f_2(s_2, u_2) = a$ holds, where $a \in \mathbb{Z}_4$ is a constant that only depends on u_2. This implies that conditioned on u_2, $X_2 + S_2$ equals to a constant a, with probability one. Therefore,

$$H(X_1 + S_1 + X_2 + S_2|u_2u_1) = H(X_1 + S_1 + a|u_1, u_2) = H(X_1 + S_1|u_1)$$

Moreover,

$$H(S_2|u_2) = H(a \ominus X_2|u_2) = H(X_2|u_2).$$

By assumption $p(u_2) > 0$. Therefore, the cost constraint $\mathbb{E}(c_2(X_2)) = 0$ implies that $\mathbb{E}(c_2(X_2)|U_2 = u_2) = 0$. Hence, given $U_2 = u_2$, the random variable X_2 takes at most two values with positive probabilities. As a result, $H(X_2|u_2) \leq 1$. Given this

inequality, we obtain

$$R(u_2, Q_{u_2}) \leq H(S_1|u_1) - H(X_1 + S_1|u_1) - 1 \leq 0$$

where the last inequality follows by Lemma 32 in Appendix A.9.

Case 2: $u_2 \in \mathcal{B}_2$

For any fixed $u_2 \in \mathcal{B}_2$, $f_2(s_2, u_2) + s_2$ takes two values for all $s_2 \in \mathbb{Z}_4$. Assume these values are $a, b \in \mathbb{Z}_4$, where $a \neq b$. Given u_2 the random variable $X_2 + S_2$ is distributed over $\{a, b\}$. Therefore, $X_2 + S_2 \ominus a$ is distributed over $\{0, b \ominus a\}$, and

$$H(X_1 + S_1 + X_2 + S_2|u_2, u_1) = H(X_1 + S_1 + X_2 + S_2 \ominus a|u_2, u_1).$$

As a result, the case $\{a, b\}$ gives the same bound as $\{0, b \ominus a\}$, and we need to consider only the case in which $a = 0$. For the case in which $a = 0$, and $b = 3$, consider $X_2 + S_2 + 1$. Using a similar argument as above, we can show that when $b = 3$, we get the same bound when $b = 1$. Therefore, we only need to consider the cases in which $a = 0$, and $b \in \{1, 2\}$. We address these cases in the next Claim.

Claim 1. *Let* $P(X_2 + S_2 = 0|u_1) = p_0$. *The following holds:*

1) If $b = 2$, *then*

$$\begin{aligned} R(u_2, Q_{u_2}) \leq\ & \beta(H(S_1|u_1) - H(X_1 + S_1 + N_{(2/3,0,1/3,0)}|u_1)) \\ & + (1 - \beta)(H(S_1|u_1) - H(X_1 + S_1 + N_{(1/3,0,2/3,0)}|u_1)) + H(S_2|u_2) - 2 \end{aligned}$$

2) If $b = 1$, *then*

$$\begin{aligned} R(u_2, Q_{u_2}) \leq\ & \beta(H(S_1|u_1) - H(X_1 + S_1 + N_{(2/3,1/3,0,0)}|u_1)) \\ & + (1 - \beta)(H(S_1|u_1) - H(X_1 + S_1 + N_{(1/3,2/3,0,0)}|u_1)) + H(S_2|u_2) - 2 \end{aligned}$$

Proof. The proof is given in Appendix A.10. □

Using the claim and applying Lemma 32, we have

$$R(u_2, Q_{u_2}) < 1 + H(S_2|u_2) - 2 \leq 1.$$

Case 3: $u_2 \in \mathcal{B}_3$

We need only to consider the case when $\mathbf{p} = (p_0, p_1, p_2, 0)$. We proceed by the following claim.

Claim 2. *If $u_2 \in \mathcal{B}_3$, the following bound holds*

$$\begin{aligned} R(u_2, Q_{u_2}) \leq\ & \beta_0(H(S_1|u_1) - H(X_1 + S_1 + N_{(2/4,1/4,1/4,0)}|u_1)) \\ & + \beta_1(H(S_1|u_1) - H(X_1 + S_1 + N_{(1/4,2/4,1/4,0)}|u_1)) \\ & + \beta_2(H(S_1|u_1) - H(X_1 + S_1 + N_{(1/4,1/4,2/4,0)}|u_1)) + H(S_2|u_2) - 2, \end{aligned}$$

where $\beta_i = 4p_i - 1,\ i = 0, 1, 2.$

Proof. Similar to Claim 1, we can write $\mathbf{p}$ as a linear combination of three distributions of the form

$$\mathbf{p} = \beta_0(2/4, 1/4, 1/4, 0) + \beta_1(1/4, 2/4, 1/4, 0) + \beta_2(1/4, 1/4, 2/4, 0),$$

where $\beta_i = 4p_i - 1,\ i = 0, 1, 2.$ The proof then follows from the concavity of the entropy. □

Therefore, by Lemma 32, we obtain

$$R(u_2, Q_{u_2}) < 1 + H(S_2|u_2) - 2 \leq 1.$$

Case 4: $u_2 \in \mathcal{B}_4$

In this case, there is a 1-1 correspondence between $x_2(s_2, u_2)+s_2$ and s_2. Therefore $H(S_2|u_1, u_2) = H(S_2 + X_2|u_1, u_2)$, and we obtain

$$H(S_2|u_1, u_2) - H(X_1 + S_1 + X_2 + S_2|u_1, u_2) = H(S_2 + X_2|u_1, u_2) - H(X_1 + S_1 + X_2 + S_2|u_1, u_2)$$
$$\leq 0$$

Therefore $H(S_1|u_1) + H(S_2|u_2) - H(Y|u_1u_2) - 2 \leq H(S_1|u_1) - 2 \leq 0$.

Finally, considering all four cases $R(u_2, Q_{u_2}) < 1$ for all $u_2 \in \mathcal{U}_2$. This completes the proof. □

A.9 Useful Lemmas

Lemma 28. *Let X and Y be independent random variables with marginal distributions P_X and P_Y, respectively. Suppose X and Y take values from a group $\mathbb{Z}_m$. Then*

1. $A_{\epsilon/2}^{(n)}(X + Y) \subseteq A_\epsilon^{(n)}(X) + A_\epsilon^{(n)}(Y)$,

2. *there exists a function $\delta(\cdot)$ with $\lim_{\epsilon\to 0} \delta(\epsilon) = 0$ such that*

$$\frac{\left|A_{\delta(\epsilon)}^{(n)}(X, Y)\right|}{\left|A_\epsilon^{(n)}(X)\right|\left|A_\epsilon^{(n)}(Y)\right|} \geq 1 - 2^{-n\frac{\epsilon}{m}}.$$

Proof. For the first statement take an arbitrary element $\mathbf{z} \in A_{\epsilon/2}^{(n)}(X + Y)$. We show that such an element can be written as $\mathbf{z} = \mathbf{x} + \mathbf{y}$ for some element $\mathbf{x} \in A_\epsilon^{(n)}(X)$ and $\mathbf{y} \in A_\epsilon^{(n)}(Y)$. For that, select an arbitrary $\mathbf{y} \in A_{\epsilon/2}^{(n)}(Y|\mathbf{z})$. From standard arguments on typical sequences, $\mathbf{y}$ is $\epsilon/2$- typical with respect to P_Y. In addition, $(\mathbf{z}, \mathbf{y}) \in A_\epsilon^{(n)}(X + Y, Y)$. As a result, $(\mathbf{z} - \mathbf{y}, \mathbf{y}) \in A_\epsilon^{(n)}(X, Y)$. Set $\mathbf{x} = \mathbf{z} - \mathbf{y}$. We

showed that, $(\mathbf{x},\mathbf{y}) \in A_\epsilon^{(n)}(X,Y)$, and $\mathbf{x}+\mathbf{y}=\mathbf{z}$. Since $\mathbf{x}$ and $\mathbf{y}$ are jointly ϵ-typical, then $\mathbf{x} \in A_\epsilon^{(n)}(X)$ and $\mathbf{y} \in A_\epsilon^{(n)}(Y)$. This completes the proof for the first statement.

For the second statement, given $\tilde{\epsilon} > 0$ we have

$$1 - \frac{\left|A_{\tilde{\epsilon}}^{(n)}(X,Y)\right|}{\left|A_\epsilon^{(n)}(X)\right|\left|A_\epsilon^{(n)}(Y)\right|} \leq \frac{\left|A_{\tilde{\epsilon}}^{(n)}(X,Y)^c\right|}{\left|A_\epsilon^{(n)}(X)\right|\left|A_\epsilon^{(n)}(Y)\right|} = \sum_{(\mathbf{x},\mathbf{y})\notin A_{\tilde{\epsilon}}^{(n)}(X,Y)} \frac{1}{\left|A_\epsilon^{(n)}(X)\right|\left|A_\epsilon^{(n)}(Y)\right|}$$

Let $P_{X,Y}^n = \prod_{i=1}^n P_X P_Y$. From standard arguments for ϵ-typical sequences the above expression does not exceed

$$\sum_{(\mathbf{x},\mathbf{y})\notin A_{\tilde{\epsilon}}^{(n)}(X,Y)} 2^{n\epsilon\frac{\alpha}{m}} P_{X,Y}^n(\mathbf{x},\mathbf{y}) = P_{X,Y}^n\{A_{\tilde{\epsilon}}^{(n)}(X,Y)^c\}2^{n\epsilon\frac{\alpha}{m}} \leq 2^{n\epsilon\frac{\alpha}{m}} 2^{-\frac{\tilde{\epsilon}^2 n}{m^2 \ln 4}}$$

where

$$\alpha = -\frac{3}{m} \sum_{\substack{a,b\in\mathbb{Z}_m \\ P_{X,Y}(a,b)>0}} \log P_{X,Y}(a,b).$$

The last inequality holds as (X,Y) are independent. Define the function $\delta(\epsilon) =\triangleq [m\epsilon(1+\alpha)\ln 4]^{1/2}$ and set $\tilde{\epsilon} = \delta(\epsilon)$. As a result, the right-hand side of the above inequality is simplified to $2^{-n\frac{\epsilon}{m}}$. Thus, the second statement of the lemma is established. □

Lemma 29 ([67]). *Suppose that* $\mathbf{G}$ *is a* $k\times n$ *matrix with elements generated randomly and uniformly from* $\mathbb{Z}_{p^r}$. *If* $\mathbf{u} \in H_s^k \backslash H_{s+1}^k$, *then*

$$P\{\mathbf{u}\mathbf{G}_i = \mathbf{x}\} = p^{-n(r-s)}\mathbb{1}\{x \in H_s^n\}.$$

Lemma 30. *Given* $(X,Y) \sim P_{XY}$, *and sequences* $\mathbf{x},\mathbf{y}$ *such that* $([\mathbf{x}]_s,\mathbf{y}) \in A_\epsilon^{(n)}([X]_s,Y)$, *let* $\mathcal{A} = \{\mathbf{x}' \mid (\mathbf{x}',\mathbf{y}) \in A_\epsilon^n(XY), \mathbf{x}'-\mathbf{x} \in H_s^n\}$. *Then*

$$A_{c_1\epsilon}^{(n)}(X|[\mathbf{x}]_s,\mathbf{y}) \subseteq \mathcal{A} \subseteq A_{c_2\epsilon}^{(n)}(X|[\mathbf{x}]_s,\mathbf{y}),$$

and we have,

$$(1-c_1\epsilon)2^{n(H(X|Y[X]_s)-c_1\delta(\epsilon))} \leq |\mathcal{A}| \leq 2^{n(H(X|Y[X]_s)+c_2\delta(\epsilon))},$$

where $\delta(\epsilon) = \frac{\epsilon}{|\mathcal{Y}|}\sum_{a\in\mathcal{X}}\sum_{b\in\mathcal{Y}:p(b|a)>0}\log_2 p(b|a)$, $c_1 = \frac{1}{|\mathcal{X}|+|\mathcal{Y}|}$, *and* $c_2 = p^{r-s}\frac{|\mathcal{X}|+1}{|\mathcal{Y}|}$.

Proof. Suppose $\mathbf{x}' \in \mathcal{A}$. Then $\mathbf{x}' - \mathbf{x} \in H_s^n$, which implies $[\mathbf{x}']_s = [\mathbf{x}]_s$. In addition, $(\mathbf{x}', \mathbf{y}) \in A_\epsilon^{(n)}(X,Y)$. Therefore, $(\mathbf{x}', [\mathbf{x}]_s, \mathbf{y}) \in A_{\epsilon'}^{(n)}(X,[X],Y)$, where $\epsilon' = \epsilon p^{r-s}$. Thus, $\mathbf{x}' \in A_{\epsilon''}^{(n)}(X|[\mathbf{x}]_s, \mathbf{y})$, where $\epsilon'' = \frac{|\mathcal{X}|+1}{|\mathcal{Y}|}\epsilon'$. On the other hand, if $\mathbf{x}' \in A_{\tilde{\epsilon}}^{(n)}(X|[\mathbf{x}]_s\mathbf{y})$, then $[\mathbf{x}']_s = [\mathbf{x}]_s$, and $\mathbf{x}' \in A_\epsilon^{(n)}(X|\mathbf{y})$, where $\epsilon = \tilde{\epsilon}(|\mathcal{X}|+|\mathcal{Y}|)$. □

Lemma 31. *Let* X *and* Y *be two independent random variables over* $\mathbb{Z}_m$ *with distributions* $\mathbf{p} = (p_0, p_1, ..., p_{m-1})$ *and* $\mathbf{q} = (q_0, q_1, ..., q_{m-1})$, *respectively. Then* $H(X \oplus_m Y) = H(Y)$ *if and only if there exists* $i \in [1:m]$ *such that* $\mathbf{p} \circledast_m \mathbf{q} = \pi^i(\mathbf{q})$, *where* $\circledast_m$ *is the circular convolution and is defined as*

$$(\mathbf{p} \circledast_m \mathbf{q})(a) \triangleq \sum_{b\in\mathbb{Z}_m} p_b q_{a\ominus b}, \quad \forall a \in \mathbb{Z}_m,$$

$\pi((q_0, q_1, ..., q_{m-1})) = (q_{m-1}, q_0, q_1, ..., q_{m-2})$, *and* π^i *is the composition of the function* π *with itself for* i *times.*

Proof. First note that as X is independent of Y, we have $H(X \oplus_m Y) - H(Y) = I(X; X \oplus_m Y) \geq 0$. We want to find all distributions $\mathbf{p}$ and $\mathbf{q}$ for which the right-hand side equals zero. We first fix a distribution $\mathbf{q}$ and find all $\mathbf{p}$ such that the equality holds. This is equivalent to the solution of the following minimization problem:

$$\min_{\mathbf{p}\in\Delta_m} H(\mathbf{p} \circledast_m \mathbf{q}) - H(\mathbf{q}), \tag{A.42}$$

where $\Delta_m \triangleq \{(q_0, q_1, ..., q_{m-1}) \in \mathbb{R}^m : \sum_{i=0}^{m-1} q_i = 1,\ q_i \geq 0,\ i \in [0:m-1]\}$. Note that Δ_m is a $m-1$-dimensional simplex in $\mathbb{R}^m$. Define the map $\varphi_{\mathbf{q}} : \Delta_m \mapsto \Delta_m,\ \varphi_{\mathbf{q}}(\mathbf{p}) =$

$\mathbf{p} \circledast_m \mathbf{q}$ for all $\mathbf{p}, \mathbf{q} \in \Delta_m$. Note that $\varphi_{\mathbf{q}}$ is a linear map. Let $\varphi_{\mathbf{q}}(\Delta_m)$ denote the image of Δ_m under $\varphi_{\mathbf{q}}$. Since $\varphi_{\mathbf{q}}$ is a linear map, $\varphi_{\mathbf{q}}(\Delta_m)$ is a simplex. Therefore, (A.42) is equivalent to $\min_{\mathbf{p}' \in \varphi_{\mathbf{q}}(\Delta_m)} H(\mathbf{p}') - H(\mathbf{q})$. It is well-known that the entropy function is strictly concave. Hence, the minimum points are the extreme points of the simplex $\varphi_{\mathbf{q}}(\Delta_m)$. Extreme points of $\varphi_{\mathbf{q}}(\Delta_m)$ are the image of the extreme points of Δ_m. Define the map $\pi : \Delta_m \mapsto \Delta_m$ as in the statement of the lemma. Extreme points of $\varphi_{\mathbf{q}}(\Delta_m)$ are characterized by $\pi^i(\mathbf{q}), i \in [1 : m]$, where π^i is the composition of π with itself for i times. Therefore, the minimum points of (A.42) are described as $\bigcup_{i=1}^m \varphi_{\mathbf{q}}^{-1}(\pi^i(\mathbf{q}))$, where $\varphi^{-1}(\mathbf{a})$ is the pre-image of $\mathbf{a}, \forall \mathbf{a} \in \Delta_m$.

Next, we range over all $\mathbf{q} \in \Delta_m$. Define the set

$$\mathcal{A}_i \triangleq \{(\mathbf{p}, \mathbf{q}) \in \Delta_m \times \Delta_m : \mathbf{p} \circledast_m \mathbf{q} = \pi^i(\mathbf{q})\}.$$

Then, the set of all $(\mathbf{p}, \mathbf{q})$ such that $H(\mathbf{p} \circledast_m \mathbf{q}) = H(\mathbf{q})$ is characterized by the set $\bigcup_{i=1}^m \mathcal{A}_i$. This is equivalent to the statement of the lemma. □

Lemma 32. *Suppose S and $N_{\mathbf{p}}$ are independent random variables over $\mathbb{Z}_4$, where $\mathbf{p}$ is the distribution of $N_{\mathbf{p}}$. Let $f : \mathbb{Z}_4 \mapsto \mathbb{Z}_4$ be a function of S, and denote $X \triangleq f(S)$. Suppose for the cost functions (c_1, c_2) given in Example 4, the equality $\mathbb{E}\{c_1(X)\} = 0$ holds. Then the following bounds hold:*

$$H(S) - H(X + S) \leq 1$$
$$H(S) - H(X + S + N_{\mathbf{p}}) < 1,$$

where $\mathbf{p} \in \{(1/3, 0, 2/3, 0), (1/3, 2/3, 0, 0), (1/4, 1/4, 1/2, 0)\}$.

Proof. For the first equality, we start with the following relations

$$\begin{aligned} H(X+S) &= H(X,S) - H(X|X+S) \\ &= H(S) - H(X|X+S). \end{aligned}$$

Therefore, we obtain

$$H(S) - H(X+S) = H(X|X+S) \leq H(X) \overset{(a)}{\leq} 1.$$

Note (a) is true, because X takes at most two values with positive probabilities.

For the second inequality we have

$$\begin{aligned} H(S) - H(X+S+N_{\mathbf{p}}) &= H(S) - H(X+S) + H(X+S) - H(X+S+N_{\mathbf{p}}) \\ &\leq 1 - (H(X+S+N_{\mathbf{p}}) - H(X+S)) \leq 1. \end{aligned} \tag{A.43}$$

Let $\mathbf{q}$ be the distribution of $X+S$. We find the conditions on $\mathbf{p}$ and $\mathbf{q}$ for which $H(X+S+N_{\mathbf{p}}) - H(X+S) = 0$. Since $N_{\mathbf{p}}$ is independent of $X+S$, we can use Lemma 31 in which $Y = N_{\mathbf{p}}$ and $X = X+S$. Therefore, $H(X+S+N_{\mathbf{p}}) = H(X+S)$, if and only if $\mathbf{p} \circledast_4 \mathbf{q} = \pi^i(\mathbf{q})$ for some $i \in [1:4]$. For fixed i and $\mathbf{p}$, the map defined by $\mathbf{q} \mapsto \mathbf{p} \circledast_4 \mathbf{q} - \pi^i(\mathbf{q})$ is a linear map. In addition, the null space of this map characterizes the set of all $\mathbf{q}$ that satisfies the equality in Lemma 31. For $\mathbf{p} = (1/3, 0, 2/3, 0)$ this map can be represented by the matrix

$$A_{i,(1/3,0,2/3,0)} = \begin{bmatrix} -\frac{2}{3} & 0 & \frac{2}{3} & 0 \\ 0 & -\frac{2}{3} & 0 & \frac{2}{3} \\ \frac{2}{3} & 0 & -\frac{2}{3} & 0 \\ 0 & \frac{2}{3} & 0 & -\frac{2}{3} \end{bmatrix}$$

The null space of $\mathbf{A}_{i,(1/3,0,2/3,0)}$ is the subspace spanned by $(1/2, 0, 1/2, 0)$ and $(1/4, 1/4, 1/4, 1/4)$. Using the same approach, we can show that for any $i \in [1:4]$ and

$$\mathbf{p} \in \{(1/3, 0, 2/3, 0), (1/3, 2/3, 0, 0), (1/4, 1/4, 1/2, 0)\},$$

the null space of $\mathbf{A}_{i,\mathbf{p}}$ is contained in the subspace spanned by $(1/2, 0, 1/2, 0)$ and $(1/4, 1/4, 1/4, 1/4)$. This implies that $q_0 = q_2$ and $q_1 = q_3$.

Table A.1: The conditions on $x(\cdot)$ and S.

$X+S$	0	1	2	3
$(s, x(s))$	$(0,0), (2,2)$	$(1,0), (3,2)$	$(0,2), (2,0)$	$(1,2), (3,0)$

Note $\mathbf{q}$ is the distribution of $x(S)+S$. Next, we find all functions $x(\cdot)$ and random variables S such that $q_0 = q_2$ and $q_1 = q_3$. For each $a \in \mathbb{Z}_4$, we characterize $(s, x(s))$ such that $x(s) + s = a$, where $x(s) \in \{0, 2\}$. We present such a characterization in Table A.1. Using Table A.1, if $q_0 > 0$, then $p(S = 0) = p(S = 2) = q_0$ and $x(0) = x(2)$. Similarly, if $q_1 > 0$, then $p(S = 1) = p(S = 3) = q_1$ and $x(1) = x(3)$. Therefore, if $q_0, q_1 > 0$, the distribution of S equals to $\mathbf{q} = (q_0, q_1, q_0, q_1)$. If $q_0 = 0$, then $q_1 = 1/2$. This implies $p(S = 1) = p(S = 3) = 1/2$. Similarly, If $q_1 = 0$, then $p(S = 0) = p(S = 2) = q_1 = 1/2$. As a result of this argument, $H(S) = H(X + S)$. Also by Lemma 31, the equality $H(X + S) = H(X + S + N_{\mathbf{p}})$ holds. Therefore, in this case, $H(S) - H(X + S + N_{\mathbf{p}}) = 0$. To sum-up, we proved that if $\mathbf{p} \in \{(1/3, 0, 2/3, 0), (1/3, 2/3, 0, 0), (1/4, 1/4, 1/2, 0)\}$ and $H(X + S) = H(X + S + N_{\mathbf{p}})$, then $H(S) - H(X + S + N_{\mathbf{p}}) = 0$. Therefore, using this argument and (A.43), we proved that if $\mathbf{p} \in \{(1/3, 0, 2/3, 0), (1/3, 2/3, 0, 0), (1/4, 1/4, 1/2, 0)\}$, then $H(X + S) - H(X + S + N_{\mathbf{p}}) < 1$. □

A.10 Proof of Claim 1

1) Let $a = 0, b = 2$, and $P(X_2+S_2 = 0|u_1) = p_0$, and $P(X_2+S_2 = 2|u_1) = 1-p_0$. We represent this PMF by the vector $\mathbf{p} = (p_0, 0, 1-p_0, 0)$. This probability distribution is a linear combination of the form

$$\mathbf{p} = \beta(2/3, 0, 1/3, 0) + (1-\beta)(1/3, 0, 2/3, 0), \tag{A.44}$$

where $\beta = 3p_0 - 1$.

Remark 21. Let $Z = X + Y$, where the PMF of X is $\mathbf{p} = (p_0, p_1, p_2, p_3)$, and the PMF of Y is $\mathbf{q} = (q_0, q_1, q_2, q_3)$. If $\mathbf{t}$ is the PMF of Z, then $\mathbf{t} = \mathbf{p} \circledast_4 \mathbf{q}$, where $\circledast_4$ is the circular convolution in $\mathbb{Z}_4$. In addition, the map $(\mathbf{p}, \mathbf{q}) \longmapsto \mathbf{p} \circledast_4 \mathbf{q}$ is a bi-linear map.

Let $t_i = p(X_1 + S_1 + X_2 + S_2 = i|u_1u_2)$ and $q_i = p(X_1 + S_1 = i|u_1)$ for all $i \in \mathbb{Z}_4$. Also denote $\mathbf{q} = (q_0, q_1, q_2, q_3)$, and $\mathbf{t} = (t_0, t_1, t_2, t_3)$. Using Remark 21 and equation (A.44) we obtain

$$\mathbf{t} = \beta\big((2/3, 0, 1/3, 0) \circledast_4 \mathbf{q}\big) + (1-\beta)\big((1/3, 0, 2/3, 0) \circledast_4 \mathbf{q}\big).$$

This implies that, $\mathbf{t}$ is also a linear combination of two PMFs. From the concavity of entropy, we get the following lower-bound:

$$\begin{aligned}
H(X_1 + S_1 + X_2 + S_2|u_1u_2) &= H(\mathbf{t}) \\
&= H(\beta\big((2/3, 0, 1/3, 0) \circledast_4 \mathbf{q}\big) + (1-\beta)\big((1/3, 0, 2/3, 0) \circledast_4 \mathbf{q}\big)) \\
&\geq \beta H((2/3, 0, 1/3, 0) \circledast_4 \mathbf{q}) + (1-\beta)H((1/3, 0, 2/3, 0) \circledast_4 \mathbf{q}) \\
&= \beta H(X_1 + S_1 + N_{(2/3,0,1/3,0)}|u_1) + (1-\beta)H(X_1 + S_1 + N_{(1/3,0,2/3,0)}|u_1),
\end{aligned}$$

where in the last equality, $N_{(\lambda_0,\lambda_1,\lambda_2,\lambda_3)}$ denotes a random variable with PMF $(\lambda_0, \lambda_1, \lambda_2, \lambda_3)$ that is also independent of u_1 and X_1+S_1. As a result of the above argument, equation (A.37) is bounded by

$$\begin{aligned}
&H(S_1|u_1) + H(S_2|u_2) - H(Y|u_1u_2) - 2 \\
&\qquad \leq H(S_1|u_1) + H(S_2|u_2) - \beta H(X_1 + S_1 + N_{(2/3,0,1/3,0)}|u_1) \\
&\qquad - (1-\beta)H(X_1 + S_1 + N_{(1/3,0,2/3,0)}|u_1) - 2 \\
&\qquad = \beta(H(S_1|u_1) - H(X_1 + S_1 + N_{(2/3,0,1/3,0)}|u_1)) \\
&\qquad + (1-\beta)(H(S_1|u_1) - H(X_1 + S_1 + N_{(1/3,0,2/3,0)}|u_1)) + H(S_2|u_2) - 2
\end{aligned}$$

2) Let $a = 0, b = 2$, and $P(X_2 + S_2 = 0|u_1) = p_0$, and $P(X_2 + S_2 = 1|u_1) = 1 - p_0$. In this case $\mathbf{p} = (p_0, 1-p_0, 0, 0)$. Also,

$$\mathbf{p} = \beta(2/3, 1/3, 0, 0) + (1-\beta)(1/3, 2/3, 0, 0),$$

where $\beta = 3p_0 - 1$. Similar to case 1), we use Remark 21 and the concavity of the entropy to get,

$$\begin{aligned}
&H(S_1|u_1) + H(S_2|u_2) - H(Y|u_1u_2) - 2 \\
&\qquad \leq \beta(H(S_1|u_1) - H(X_1 + S_1 + N_{(2/3,1/3,0,0)}|u_1)) \\
&\qquad + (1-\beta)(H(S_1|u_1) - H(X_1 + S_1 + N_{(1/3,2/3,0,0)}|u_1)) + H(S_2|u_2) - 2
\end{aligned}$$

APPENDIX B

Proofs for Chapter III

B.1 Proof of Theorem III.1

Proof. There are two error events, E_0 and E_1. E_0 occurs if no $\underline{\tilde{\mathbf{s}}}$ was found. E_1 is declared if $\underline{\tilde{\mathbf{s}}} \neq \underline{\mathbf{s}}$. To show that E_0 is small, we need the next lemma. Suppose $v_i()$ and $v_i()$ are a realization of random functions generated as in the outline of the proof of Theorem III.1.

Lemma 33. *Suppose* $\mathbf{s}_i, i = 1, 2, 3$ *are jointly typical with respect to* $P_{\underline{\mathbf{S}}}$. *Then*

$$\big(v_1(\mathbf{s}_1), v_2(\mathbf{s}_2), v_3(\mathbf{s}_3), x_1(\mathbf{s}_1, v_1(\mathbf{s}_1)), x_2(\mathbf{s}_2, v_2(\mathbf{s}_2)), x_3(\mathbf{s}_3, v_3(\mathbf{s}_3))\big) \in A_\epsilon^{(n)}(V_1V_2V_3X_1X_2X_3|\mathbf{s}_1\mathbf{s}_2\mathbf{s}_3).$$

Proof. The proof is straightforward. □

As a result, the sequences $\mathbf{s}_i, \mathbf{v}_i, \mathbf{x}_i, i = 1, 2, 3$ are jointly typical with $\mathbf{y}$ with respect to $P_{\underline{\mathbf{S}},\underline{\mathbf{V}},\underline{\mathbf{X}},Y}$. This implies that $P(E_0)$ approaches 0 as $n \to \infty$. Next, we calculate $P(E_1 \cap E_0^c)$. For a given $\underline{\mathbf{s}} \in A_\epsilon(\underline{S})$, using the definition of E_1 and the union bound

we obtain,

$$P(E_1 \cap E_0^c|\underline{\mathbf{s}}) \leq \sum_{(\underline{\mathbf{v}},\underline{\mathbf{x}})\in A_\epsilon(\underline{\mathbf{V}},\underline{\mathbf{X}}|\underline{\mathbf{s}})} \mathbb{1}\{\mathbf{v}_i = v_i(\mathbf{s}_i), \mathbf{x}_i = x_i(\mathbf{s}_i,\mathbf{v}_i),\ i=1,2,3\} \sum_{\mathbf{y}\in A_\epsilon(Y|\underline{\mathbf{x}})} p(\mathbf{y}|\underline{\mathbf{x}})$$
$$\sum_{\substack{(\underline{\tilde{\mathbf{s}}},\underline{\tilde{\mathbf{v}}},\underline{\tilde{\mathbf{x}}})\in A_\epsilon(\underline{S},\underline{V},\underline{X}|\mathbf{y}) \\ \underline{\tilde{\mathbf{s}}}\neq\underline{\mathbf{s}}}} \mathbb{1}\{\tilde{\mathbf{v}}_j = v_j(\tilde{\mathbf{s}}_j), \tilde{\mathbf{x}}_j = x_j(\tilde{\mathbf{s}}_j,\tilde{\mathbf{v}}_j), j=1,2,3\}$$

Taking expectation over random functions $X_i(,)$ and $V_i()$ gives,

$$p_e(\underline{\mathbf{s}}) = \mathbb{E}\{P(E_1|\underline{\mathbf{s}})\} \leq \sum_{(\underline{\mathbf{v}},\underline{\mathbf{x}},\mathbf{y})\in A_\epsilon(\underline{V},\underline{X},Y|\underline{\mathbf{s}})} p(\mathbf{y}|\underline{\mathbf{x}}) \sum_{\substack{(\underline{\tilde{\mathbf{s}}},\underline{\tilde{\mathbf{v}}},\underline{\tilde{\mathbf{x}}})\in A_\epsilon(\underline{S},\underline{V},\underline{X}|\mathbf{y}) \\ \underline{\tilde{\mathbf{s}}}\neq\underline{\mathbf{s}}}} \tag{B.1}$$
$$P\{\mathbf{v}_l = V_l(\mathbf{s}_l), \mathbf{x}_l = X_l(\mathbf{s}_l,\mathbf{v}_l), \tilde{\mathbf{v}}_l = V_l(\tilde{\mathbf{s}}_l), \tilde{\mathbf{x}}_l = X_l(\tilde{\mathbf{s}}_l,\tilde{\mathbf{v}}_l) \text{ for } l=1,2,3\}$$

Note that $V_i()$ and $\mathbf{X}_i(\ ,\ )$ are generated independently. So the most inner term in (B.1) is simplified to

$$P\{\mathbf{v}_j = V_j(\mathbf{s}_j), \tilde{\mathbf{v}}_j = V_j(\tilde{\mathbf{s}}_j)\ j=1,2\} P\{\mathbf{x}_l = X_l(\mathbf{s}_l,\mathbf{v}_l), \tilde{\mathbf{x}}_l = X_l(\tilde{\mathbf{s}}_l,\tilde{\mathbf{v}}_l)\ l=1,2,3\}. \tag{B.2}$$

Note $j=3$ is redundant because, $\mathbf{v}_3 = \mathbf{v}_1 \oplus_q \mathbf{v}_2$ and $\tilde{\mathbf{v}}_3 = \tilde{\mathbf{v}}_1 \oplus_q \tilde{\mathbf{v}}_2$. By definition, $V_j(\mathbf{s}_j) = \mathbf{s}_j\mathbf{G} + \mathbf{B}_j, j=1,2$, where B_1, B_2 are uniform and independent of $\mathbf{G}$. Then

$$P\{\mathbf{v}_j = \Phi_j(\mathbf{s}_j), \tilde{\mathbf{v}}_j = \Phi_j(\tilde{\mathbf{s}}_j)\ j=1,2\} = \frac{1}{q^{2n}} P\{(\tilde{\mathbf{s}}_j - \mathbf{s}_j)\mathbf{G} = \tilde{\mathbf{v}}_j - \mathbf{v}_j,\ j=1,2\} \tag{B.3}$$

The following lemma determines the above term.

Lemma 34. *Suppose elements of* $\mathbf{G}$ *are generated randomly and uniformly from* $\mathbb{F}_q$.

If $\mathbf{s}_1$ *or* $\mathbf{s}_2$ *is nonzero, the following holds:*

$$P\{\mathbf{s}_j\mathbf{G}=\mathbf{v}_j,\ j=1,2\}=\begin{cases} q^{-n}\mathbb{1}\{\mathbf{v}_j=\mathbf{0}\}, & \text{if } \mathbf{s}_j=\mathbf{0} \\ q^{-n}\mathbb{1}\{\mathbf{v}_1=\mathbf{v}_2\}, & \text{if } \mathbf{s}_1\neq\mathbf{0},\mathbf{s}_2\neq\mathbf{0},\mathbf{s}_1=\mathbf{s}_2. \\ q^{-2n}, & \text{if otherwise.} \end{cases}$$

Outline of the proof. We can write $\mathbf{s}_j\mathbf{G}=\sum_{i=1}^{n}\mathbf{s}_{ji}\mathbf{G}_i$, where $\mathbf{s}_{ji}$ is the ith component of $\mathbf{s}_j$ and $\mathbf{G}_i$ is the ith row of $\mathbf{G}$. Not that $\mathbf{G}_i$ are independent random variables with uniform distribution over $\mathbb{F}_q^n$. Hence, if $\mathbf{s}_j\neq\mathbf{0}$, then $\mathbf{s}_j\mathbf{G}$ is uniform over $\mathbb{F}_q^n$. If $\mathbf{s}_1\neq\mathbf{s}_2$, one can show that $\mathbf{s}_1\mathbf{G}$ is independent of $\mathbf{s}_2\mathbf{G}$. The proof follows by arguing that if a random variables X is independent of Y and is uniform over $\mathbb{F}_q$, then $X\oplus_q Y$ is also uniform over $\mathbb{F}_q$ and is independent of Y. □

Finally, we are ready to characterize the conditions in which $p_e\to 0$. We divide the last summation in (B.1) into the following cases:

Case 1, $\tilde{\mathbf{s}}_1\neq\mathbf{s}_1,\tilde{\mathbf{s}}_2=\mathbf{s}_2$ In this case, using Lemma 34, (B.3) equals to $q^{-3n}\mathbb{1}\{\tilde{\mathbf{v}}_2=\mathbf{v}_2\}$. Therefore, (B.1) is simplified to

$$p_{e_1}(\underline{\mathbf{s}}):=\sum_{(\underline{\mathbf{v}},\underline{\mathbf{x}},\underline{\mathbf{y}})\in A_\epsilon(\underline{V},\underline{X},Y|\underline{\mathbf{s}})} p(\mathbf{y}|\underline{\mathbf{x}}) \sum_{\substack{(\underline{\tilde{\mathbf{s}}},\underline{\tilde{\mathbf{v}}},\underline{\tilde{\mathbf{x}}})\in A_\epsilon(\underline{S},\underline{V},\underline{X}|\mathbf{y}) \\ \underline{\tilde{\mathbf{s}}}\neq\underline{\mathbf{s}},\tilde{\mathbf{s}}_2=\mathbf{s}_2,\tilde{\mathbf{v}}_2=\mathbf{v}_2}} q^{-3n}P\{\mathbf{x}_l=\mathbf{X}_l(\mathbf{s}_l,\mathbf{v}_l),\tilde{\mathbf{x}}_l=\mathbf{X}_l(\tilde{\mathbf{s}}_l,\tilde{\mathbf{v}}_l)\ l=1,2,3\}.$$

Note that $\mathbf{X}_l(\mathbf{s}_l,\mathbf{v}_l)$ is independent of $\mathbf{X}_k(\tilde{\mathbf{s}}_k,\tilde{\mathbf{v}}_k)$, if $l\neq k$ or $\mathbf{s}_l\neq\tilde{\mathbf{s}}_l$ or $\mathbf{v}_l\neq\tilde{\mathbf{v}}_l$. Moreover, $P\{\mathbf{x}_l=\mathbf{X}_l(\mathbf{s}_l,\mathbf{v}_l)\}\approx 2^{nH(X_l|S_lV_l))}$. As $\mathbf{s}_2=\tilde{\mathbf{s}}_2$ and $\mathbf{v}_2=\tilde{\mathbf{v}}_2$, then $X_2(\tilde{\mathbf{s}}_2,\tilde{\mathbf{v}}_2)=X_2(\mathbf{s}_2,\mathbf{v}_2)$. Therefore,

$$P\{\mathbf{x}_l=X_l(\mathbf{s}_l,\mathbf{v}_l),\tilde{\mathbf{x}}_l=X_l(\tilde{\mathbf{s}}_l,\tilde{\mathbf{v}}_l)\ l=1,2,3\}=2^{-n[2H(X_1|S_1V_1)+H(X_2|S_2V_2)+2H(X_3|S_3V_3)]}\mathbb{1}\{\tilde{\mathbf{x}}_2=\mathbf{x}_2\}.$$

Hence, we have:

$$p_{e_1}(\underline{\mathbf{s}}) \approx 2^{nH(\underline{V},\underline{X}|\underline{S})} 2^{nH(S_1,V_1,X_1,S_3,V_3,X_3|YS_2V_2X_2)} \frac{1}{q^{3n}} 2^{-n[2H(X_1|S_1V_1)+H(X_2|S_2V_2)+2H(X_3|S_3V_3)]}.$$

Note that $H(\underline{V},\underline{X}|\underline{S}) = 2\log_2 q + \sum_{i=1}^{3} H(X_i|S_i,V_i)$. Therefore, $p_{e_1} \to 0$, if

$$H(S_1,V_1,X_1,S_3,V_3,X_3|YS_2V_2X_2) \leq \log_2 q + H(X_1|S_1V_1) + H(X_3|S_3V_3) \qquad \text{(B.4)}$$

The right-hand side in the above inequality equals to $H(X_1X_3V_1V_3|S_1S_2S_3X_2V_2)$. We simplify the left-hand side. Observe that

$$H(S_1,V_1,X_1,S_3,V_3,X_3|YS_2V_2X_2) = H(V_1,X_1,V_3,X_3|YS_2V_2X_2) + H(S_1|S_2\underline{V}\ \underline{X}),$$

where Y is removed from the second term, because conditioned on $\underline{X}$, Y is independent of S_1. Note that

$$\begin{aligned} H(S_1|S_2\underline{V}\ \underline{X}) &= H(S_1|S_2X_2V_2) - I(S_1;X_1V_1X_3V_3|S_2V_2X_2) \\ &= H(S_1|S_2) - I(S_1;X_1V_1X_3V_3|S_2V_2X_2). \end{aligned}$$

Therefore, using the above argument the inequality in (B.4) is simplified to

$$\begin{aligned} H(S_1|S_2) \leq\ & I(S_1;X_1V_1X_3V_3|S_2V_2X_2) - H(V_1,X_1,V_3,X_3|YS_2V_2X_2) \\ & + H(X_1X_3V_1V_3|S_1S_2S_3X_2V_2) \\ &= I(X_1V_1X_3V_3;Y|S_2V_2X_2) \\ &= I(X_1X_3;Y|S_2V_2X_2.) \end{aligned}$$

Case 2, $\tilde{\mathbf{s}}_1 = \mathbf{s}_1, \tilde{\mathbf{s}}_2 \neq \mathbf{s}_2$ A similar argument as in the first case gives $H(S_2|S_1) \leq I(X_2X_3;Y|S_1V_1X_1)$.

Case 3, $\tilde{\mathbf{s}}_1 \neq \mathbf{s}_1, \tilde{\mathbf{s}}_2 \neq \mathbf{s}_2, \tilde{\mathbf{s}}_1 \oplus_q \tilde{\mathbf{s}}_2 = \mathbf{s}_1 \oplus_q \mathbf{s}_2$ Using Lemma 34,

$$P\{\mathbf{v}_j = \Phi_j(\mathbf{s}_j), \tilde{\mathbf{v}}_j = \Phi_j(\tilde{\mathbf{s}}_j)\ j = 1, 2\} = q^{-3n} \mathbb{1}\{\tilde{\mathbf{v}}_1 \oplus_q \tilde{\mathbf{v}}_2 = \mathbf{v}_1 \oplus_q \mathbf{v}_2\}$$

Therefore, the above probability is nonzero only when $\tilde{\mathbf{v}}_3 = \mathbf{v}_3$. Hence, as $\mathbf{s}_3 = \tilde{\mathbf{s}}_3$, we get $X_3(\tilde{\mathbf{s}}_3, \tilde{\mathbf{v}}_3) = X_3(\mathbf{s}_3, \mathbf{v}_3)$. This implies that,

$$P\{\mathbf{x}_l = \mathbf{X}_l(\mathbf{s}_l, \mathbf{v}_l), \tilde{\mathbf{x}}_l = \mathbf{X}_l(\tilde{\mathbf{s}}_l, \tilde{\mathbf{v}}_l)\ l = 1, 2, 3\} = 2^{-n[2H(X_1|S_1V_1)+2H(X_2|S_2V_2)+H(X_3|S_3V_3)]} \mathbb{1}\{\tilde{\mathbf{x}}_3 = \mathbf{x}_3\}.$$

As a result, (B.1), in this case, is simplified to :

$$p_{e_3}(\underline{\mathbf{s}}) \approx 2^{nH(S_1,V_1,X_1,S_2,V_2,X_2|YS_3V_3X_3)} q^{-n} 2^{-n[H(X_1|S_1V_1)+H(X_2|S_2V_2)]}.$$

Therefore, $p_{e3} \to 0$, if $H(S_1, V_1, X_1, S_2, V_2, X_2|YS_3V_3X_3) \leq H(X_1, X_2, V_1, V_2|S_1S_2S_3V_3X_3)$. Using a similar argument as in the first case, this inequality is equivalent to $H(S_1S_2|S_3) \leq I(X_1, X_2; Y|S_3V_3X_3)$.

Case 4, $\tilde{\mathbf{s}}_i \neq \mathbf{s}_i, i = 1, 2, 3$ Observe that,

$$P\{\mathbf{v}_j = \Phi_j(\mathbf{s}_j), \tilde{\mathbf{v}}_j = \Phi_j(\tilde{\mathbf{s}}_j)\ j = 1, 2\} = q^{-4n}$$
$$P\{\mathbf{x}_l = \mathbf{X}_l(\mathbf{s}_l, \mathbf{v}_l), \tilde{\mathbf{x}}_l = \mathbf{X}_l(\tilde{\mathbf{s}}_l, \tilde{\mathbf{v}}_l)\ l = 1, 2, 3\} = 2^{-2n \sum_{l=1}^{3} H(X_l|S_lV_l)}.$$

Therefore, (B.1), in this case, is simplified to $p_{e_4}(\underline{\mathbf{s}}) \approx q^{-2n} 2^{nH(\underline{S},\underline{V},\underline{X}|Y)} 2^{-n \sum_{l=1}^{3} H(X_l|S_lV_l)}$. As a result, one can show that $P_{e4} \to 0$, if $H(S_1S_2S_3) \leq I(X_1X_2X_3; Y)$.

Finally, note that $P_e(\underline{\mathbf{s}}) \leq \sum_{i=1}^{4} P_{ei}(\underline{\mathbf{s}})$. Moreover, $P_{ei}(\underline{\mathbf{s}})$ depends on $\underline{\mathbf{s}}$ only through its PMF. Therefore, for any typical $\underline{\mathbf{s}}$, P_e approaches zero as $n \to \infty$, if the following

bounds are satisfied:

$$H(S_1|S_2) \leq I(X_1X_3;Y|S_2V_2X_2)$$
$$H(S_2|S_1) \leq I(X_2X_3;Y|S_1V_1X_1)$$
$$H(S_1S_2|S_1 \oplus_q S_2) \leq I(X_1X_2;Y|S_1 \oplus_q S_2, V_3X_3)$$
$$H(S_1, S_2) \leq I(X_1X_2X_3;Y).$$

□

B.2 Proof of Lemma 12

Proof. For the setup in Example 9, the bounds given in Theorem III.1 are simplified to

$$h(\gamma) \leq I(X_2X_3;Y|X_1S_1V_1) \tag{B.5}$$
$$h(\sigma) \leq I(X_1X_2;Y|X_3S_3V_3) \tag{B.6}$$
$$h(\gamma) + h(\sigma) - h(\sigma * \gamma) \leq I(X_1X_3;Y|X_2S_2V_2) \tag{B.7}$$
$$h(\gamma) + h(\sigma) \leq I(X_1X_2X_3;Y). \tag{B.8}$$

Set $X_i = V_i, i = 1, 2, 3$, where the distribution of these random variables are given in Theorem III.1. One can verify that the source corresponding to $\sigma = 0$ and $\gamma = \gamma^*$ satisfies the above inequalities and therefore can be transmitted.

No that all the terms in (B.5)-(B.8) are entropy functions and mutual information. Therefore, they are continuous with respect to conditional density $p(\underline{x}|\underline{s}, \underline{v})$. Hence,

one can show that $\forall \epsilon_0 > 0$, there exist a conditional density $p(\underline{x}|\underline{s}, \underline{v})$ such that

$$\begin{aligned} I(X_2X_3; Y|X_1S_1V_1) &\geq 2 - H(N) - \eta(\epsilon_0) \\ I(X_1X_2; Y|X_3S_3V_3) &\geq \epsilon_0 \\ I(X_1X_3; Y|X_2S_2V_2) &\geq 2 - H(N) - \eta(\epsilon_0) \\ I(X_1X_2X_3; Y) &\geq 2 - H(N) - \eta(\epsilon_0), \end{aligned}$$

where $\eta()$ is function of ϵ such that $\eta(\epsilon) \to 0$ as $\epsilon \to 0$.

Note also that the left-hand sides in (B.5)-(B.8) are continuous in σ and γ. Hence $\exists \epsilon' > 0$ such that when $\sigma \leq \epsilon', |\gamma - \gamma^*| \leq \epsilon'$, we have

$$\begin{aligned} h(\gamma) &\leq 2 - H(N) - \eta(\epsilon_0) \\ h(\sigma) &\leq \epsilon_0 \\ h(\gamma) + h(\sigma) - h(\sigma * \gamma) &\leq 2 - H(N) - \eta(\epsilon_0) \\ h(\gamma) + h(\sigma) &\leq 2 - H(N) - \eta(\epsilon_0) \end{aligned}$$

This implies that the source corresponding to $\sigma \leq \epsilon', \gamma \leq \gamma^* - \epsilon'$ can be transmitted reliably and the proof is complete. □

APPENDIX C

Proofs for Chapter IV

C.1 Proof of Theorem IV.1

Proof. We build upon QLCs and propose a new coding scheme. Let W_i be a random variable with distribution P_{W_i}. Fix integer k and n. Consider the set of all ϵ-typical sequences W_i^k. Without loss of generality assume that the new message at the ith encoder is a sequence w_i^k which is selected randomly and uniformly from $A_\epsilon^{(k)}(W_i)$. In this case $M_i = |A_\epsilon^{(k)}(W_i)|$.

Define $\mathcal{L}[l-2]$ as the list of highly likely messages corresponding to the block $l-2$ at the decoder. This list is defined as

$$\mathcal{L}[l-2] \triangleq \{(\hat{w}_1, \hat{w}_2, \hat{w}_3) \in A_\epsilon^{(n)}(W_1, W_2, W_3) : (Y_{[l-2]}, U_{[l-2]}, S_{1,[l-2]}, S_{2,[l-2]}, S_{3,[l-2]}) \in A_\epsilon^{(n)}(\tilde{Y}, \tilde{U}, \tilde{S}_1, \tilde{S}$$

Codebook Construction: For each $1 \leq l \leq L$ generate $M_{0,[l]}$ sequences $U_{[l,m]}$, each according to P_U^n, where $1 \leq m \leq M_{0,[l]}$. For any vector $w_i^k \in \mathbb{F}_2^k$, denote

$$t_i(w_i^k) \triangleq w_i^k \mathbf{G} + b_i^n, \quad i = 1, 2, 3,$$

where $\mathbf{G}$ is a $k \times n$ matrix with elements chosen randomly and uniformly from $\mathbb{F}_2$, and b_i^n is a vector selected randomly and uniformly from $\mathbb{F}_2^n$.

For each $u^n \in \mathcal{U}^n$ and $t^n, v^n \in \mathbb{F}_2^n$ generate M_i sequences $X^n_{i,[l,m]}$ randomly with conditional distribution $\prod_{j=1}^n P(\cdot|u_j, t_j, v_j)$, where $m \in [1 : M_i]$. Denote such sequences by $x_i(u^n, t^n, v^n, m_i)$.

Initialization: For block $l = 0$, set $M_{0,[0]} = 1, U_{[0,1]} = 0$ and . For block $l = 1$, set $M_{0,[1]} = 1, U_{[1,1]} = 0, \mathbf{v}_{i,[1]} = 0$.

Encoding

Block $l = 1$ At block $l = 1$, given a message $\mathbf{w}_{i,[1]} \in A_\epsilon^{(k)}(W_i)$, the ith encoder calculates $t_i(\mathbf{w}_{i,[1]})$. This sequence is denoted by $t_{i,[1]}$. Next the encoder i calculates $x_i(u_{[0,1]}, t_{i,[1]}, v_{i,[1]}, \mathbf{w}_{i,[1]})$. Denote such sequence by $\mathbf{x}_{i,[1]}$. Finally, the $i's$ encoder sends $\mathbf{x}_{i,[1]}$.

Block $l = 2$ At the beginning of the block $l = 2$, each encoder i receives $Y_{[1]}$ as a feedback from the channel. The encoder i wishes to decode sum of the messages of the other two encoders. The first encoder finds unique $\hat{w}_{23} \in A_\epsilon^{(k)}(W_2 + W_3)$ such that

$$(\hat{w}_{23}\mathbf{G} + b_2 + b_3, Y_{[0]}) \in A_\epsilon^{(n)}(T_2 + T_3, Y|u_{[0]}t_{1,[0]}, x_{1,[l]}).$$

Otherwise an encoding error will be declared. If $\hat{w}_{23}$ was unique, the encoder sets $v_{1,[2]} = \hat{w}_{23}\mathbf{G} + b_2 + b_3$. Similarly encoder 2 finds unique $\hat{w}_{13}$ and determines $v_{2,[2]}$. Also encoder 3 finds unique $\hat{w}_{12}$, and determines $v_{3,[2]}$.

Block $l > 2$ At the beginning of the block $l > 2$, each encoder i receives $Y_{[l-1]}$ as a feedback from the channel. The encoder i wishes to decode sum of the messages of the other two encoders from block $l-1$. Next, given $Y_{[l-2]}$, the encoder i decodes the messages of the other two encoders from block $l-2$.

The first decoding process is the same as the decoding process in block $l = 2$.

Suppose $\hat{w}_{jk}$ and $v_{i,[l]}$ are the outputs of this decoding process at the encoder i. The next stage of the decoding process is as follows. The first encoder finds unique $\hat{w}_{2,[l-2]} \in A_\epsilon^{(k)}(W_2)$ and $\hat{w}_{3,[l-2]} \in A_\epsilon^{(k)}(W_3)$ such that

1) $\hat{w}_{2,[l-2]} + \hat{w}_{3,[l-2]} = \hat{w}_{23}$.

2)

$$\Big(t_2(\hat{w}_{2,[l-2]}), x_2\big(u^n, t_2(\hat{w}_{2,[l-2]}), v_{2,[l-2]}, \hat{w}_{2,[l-2]}\big),$$
$$t_3(\hat{w}_{3,[l-2]}), x_3\big(u^n, t_3(\hat{w}_{3,[l-2]}), v_{3,[l-2]}, \hat{w}_{3,[l-2]}\big), Y_{[l-2]}\Big) \in A_\epsilon^{(n)}(\tilde{T}_2\tilde{X}_2\tilde{T}_3\tilde{X}_3\tilde{Y}|s_{1,[l-2]}v_{2,[l-2]}, v_{3,[l-2]})$$

3) $(\hat{v}_{2,[l-1]}, \hat{v}_{2,[l-1]}, Y_{[l-1]}) \in A_\epsilon^{(n)}(V_2V_3Y|u_{[l-1]}s_{1[l-1])})$,

where $v_{i,[l-2]}$ is known at the encoder from the previous blocks, and $\hat{v}_{2,[l-1]}, \hat{v}_{3,[l-1]}$ are defined as

$$\hat{v}_{2,[l-1]} = (w_{1,[l-2]} + \hat{w}_{3,[l-2]})\mathbf{G} + b_1 + b_3$$
$$\hat{v}_{3,[l-1]} = (w_{1,[l-2]} + \hat{w}_{2,[l-2]})\mathbf{G} + b_1 + b_2.$$

If the messages are not unique, an error will be declared.

The next step, the encoder creates the list $\mathcal{L}[l-2]$ as defined in the above. If $(w_{1,[l-2]}, \hat{w}_{2,[l-2]}, \hat{w}_{3,[l-2]}) \in \mathcal{L}[l-2]$, then the first encoder finds the index m corresponding to $(w_{1,[l-2]}, \hat{w}_{2,[l-2]}, \hat{w}_{3,[l-2]})$. Then the encoder calculates the corresponding $u_{[l-2,m}$. Denote such sequence by $u_{[l]}$. This sequence is used for transmission of new messages at block l. If the decoding processes are successful, then the sequences $v_{1,[l]}$ and $u_{[l]}$ are determined. The next step is the encoding process, which is the same as in the block $l = 1$.

Decoding at block l The decoder knows the list of highly likely messages . This list is $\mathcal{L}[l-2]$ as defined in the above. Given $Y_{[l]}$ the decoder wishes to decode $U_{[l]}$. Note that $U_{[l]}$ determines the index of the messages in $\mathcal{L}[l-2]$ which were transmitted at

block $l-2$. This decoding process is performed by finding unique index $m \in [1 : M_{0,[l]}]$ such that

$$(U_{[l,m]}, Y_{[l]}) \in A_\epsilon^{(n)}(U, Y | u_{[l-1]}, y_{[l-1]})$$

Error Analysis There are three types of decoding errors: 1) error in decoding sum of the messages of the other two encoders, i.e., $\hat{w}_{jk}$ is not unique at the encoder i. 2) error in the decoding of the individual messages of the other encoders, i.e., $\hat{w}_{j,[l]}, \hat{w}_{k,[l]}$ are not unique at the encoder i. 3) error at the decoder, i.e. the index m is not unique. Using standard arguments for each type of the errors we get the following bounds:

The probability of the first type of the errors approaches zero, if fro any distinct $i, j, k \in \{1, 2, 3\}$ the following bound holds:

$$\frac{k}{n} H(W_j + W_k) \leq I(T_j + T_k; Y | U T_k V_k X_k). \tag{C.1}$$

The probability of the second type of the errors approaches zero, if

$$\frac{k}{n} H(W_i | W_j + W_k) \leq I(\tilde{X}_i \tilde{X}_j; \tilde{Y} | \tilde{U} \tilde{S}_k \tilde{V}_1 \tilde{V}_2 \tilde{V}_3) \tag{C.2}$$

Note that the third type of error occurs with high probability, if $|\mathcal{L}[l]| > 2^{nI(U;Y|\tilde{U},\tilde{Y})}$. It can be shown that for sufficiently large n,

$$P\{|\mathcal{L}[l]| < 2^{n \max_{\mathcal{A} \subseteq \{1,2,3\}} F_{\mathcal{A}} + o(\epsilon)}\} > 1 - \epsilon,$$

where

$$F_{\mathcal{A}} \triangleq \frac{k}{n} H(W_{\mathcal{A}}) - I(X_{\mathcal{A}}; Y | U S_{\mathcal{A}^c} \tilde{V}_1, \tilde{V}_2, \tilde{V}_3)$$

Therefore, the probability of third type of the errors approaches zero, if the following

bounds hold:

$$F_{\mathcal{A}} \leq I(U;Y|\tilde{U},\tilde{Y}),$$

Using the definition of $F_{\mathcal{A}}$ and the above bound, we can get the following bound:

$$\frac{k}{n}H(W_{\mathcal{A}}) \leq I(X_{\mathcal{A}};Y|US_{\mathcal{A}^c}\tilde{V}_1,\tilde{V}_2,\tilde{V}_3) + I(U;Y|\tilde{U},\tilde{Y}) \tag{C.3}$$

Note that the effective rate of our coding scheme is $R_i \triangleq \frac{1}{n}\log_2 M_i = \frac{k}{n}H(W_i)$ for $i = 1,2,3$. Finally, it can be shown that using this equation and the bounds in (C.1), (C.2), and (C.3), the following bounds are achievable

$$\begin{aligned}
R_{\mathcal{A}} &\leq I(X_{\mathcal{A}};Y|US_{\mathcal{A}^c}\tilde{V}_1\tilde{V}_2\tilde{V}_3) + I(U;Y|\tilde{U}\tilde{Y}) \\
R_i + R_j &\leq I(T_i \oplus T_j;Y|UT_kX_k\tilde{V}_1\tilde{V}_2\tilde{V}_3) \\
&\quad + I(\tilde{X}_i\tilde{X}_j;\tilde{Y}|\tilde{U}\tilde{S}_k\tilde{V}_1\tilde{V}_2\tilde{V}_3V_k) \\
&\quad + I(\tilde{X}_i\tilde{X}_j;Y|\tilde{U}\tilde{S}_k\tilde{V}_1\tilde{V}_2\tilde{V}_3US_k\tilde{Y}) \\
R_i + R_j &\leq \frac{H(W_i)+H(W_j)}{H(W_i \oplus W_j)} I(T_i \oplus T_j;Y|UT_kX_k).
\end{aligned}$$

□

C.2 Proof of Lemma 13

Outline of the proof. We start by proposing a coding scheme. There are L blocks of transmissions in this scheme, with new messages available at each user at the beginning of each block. The scheme sends the messages with n uses of the channel. Let $\mathbf{W}_{i,[l]}^k$ denotes the message of the ith transmitter at the lth block, where $i = 1,2,3$, and $1 \leq l \leq L$. Let $\mathbf{W}_{i,[l]}^k$ take values randomly and uniformly from $\mathbb{F}_2^k$. In this case, the transmission rate of each user is $R_i = \frac{k}{n}, i = 1,2,3$. The first and the second

outputs of the ith encoder in block l is denoted by $\mathbf{X}^n_{i1,[l]}$ and $\mathbf{X}^n_{i2,[l]}$, respectively.

Codebook Construction: Select a $k \times n$ matrix $\mathbf{G}$ randomly and uniformly from $\mathbb{F}_2^{k\times n}$. This matrix is used as the generator matrix of a linear code. Each encoder is given the matrix $\mathbf{G}$. Therefore, the encoders use an identical linear code generated by $\mathbf{G}$.

Encoder 1 and 2: For the first block set $\mathbf{X}^n_{i2,[1]} = 0$, for $i = 1, 2, 3$. For the block l, encoder 1 sends $\mathbf{X}^n_{11,[l]} = \mathbf{W}^k_{1,[l]}\mathbf{G}$ through its first output. For the second output, encoder 1 sends $\mathbf{X}^n_{11,[l-1]}$ from block $l-1$, that is $\mathbf{X}^n_{12,[l]} = \mathbf{X}^n_{11,[l-1]}$. Similarly, the outputs of the second encoder are $\mathbf{X}^n_{21,[l]} = \mathbf{W}^k_{2,[l]}\mathbf{G}$, and $\mathbf{X}^n_{22,[l]} = \mathbf{X}^n_{21,[l-1]}$.

Encoder 3: The third encoder sends $\mathbf{X}^n_{31,[l]} = \mathbf{W}^k_{3,[l]}\mathbf{G}$ though its first output. This encoder receives the feedback from the block $l-1$ of the channel. This encoder wishes to decode $\mathbf{W}^k_{1,[l-1]} \oplus \mathbf{W}^k_{2,[l-1]}$ using $\mathbf{Y}^n_{1,[l-1]}$. For this purpose, it subtracts $\mathbf{X}^n_{31,[l-1]}$ from $\mathbf{Y}^n_{1,[l-1]}$. Denote the resulting vector by $\mathbf{Z}^n$. Then, it finds a unique vector $\tilde{\mathbf{w}}^k \in \mathbb{F}_2^k$ such that $(\tilde{\mathbf{w}}^k\mathbf{G}, \mathbf{Z}^n)$ is ϵ-typical with respect to P_{XZ}, where X is uniform over $\mathbb{F}_2$, and $Z = X \oplus \tilde{N}_\delta$. If the decoding process is successful, the third encoder sends $\mathbf{X}^n_{32,[l]} = \tilde{\mathbf{w}}^k_{[l-1]}\mathbf{G}$. Otherwise, an event $E_{1,[l]}$ is declared.

Decoder: The decoder receives the outputs of the channel from the lth block, that is $\mathbf{Y}^n_{1,[l]}$ and $\mathbf{Y}^n_{2,[l]}$. The decoding is performed in three steps. First, the decoder uses $\mathbf{Y}^n_{2,[l]}$ to decode $\mathbf{W}^k_{1,[l-1]}$, and $\mathbf{W}^k_{2,[l-1]}$. In particular, it finds unique $\tilde{\mathbf{w}}^k_1, \tilde{\mathbf{w}}^k_2 \in \mathbb{F}_2^k$ such that $(\tilde{\mathbf{w}}^k_1\mathbf{G}, \tilde{\mathbf{w}}^k_2\mathbf{G}, \mathbf{Y}^n_{2,[l]})$ are jointly ϵ-typical with respect to $P_{X_{12}X_{22}Y_2}$. Otherwise, an error event $E_{2,[l]}$ will be declared.

Suppose the first part of the decoding process is successful. At the second step, the decoder calculates $\mathbf{X}^n_{11,[l-1]}$, and $\mathbf{X}^n_{21,[l-1]}$. This is possible, because $\mathbf{X}^n_{11,[l-1]}$, and $\mathbf{X}^n_{21,[l-1]}$ are functions of the messages. The decoder, then, subtracts $\mathbf{X}^n_{11,[l-1]} \oplus \mathbf{X}^n_{21,[l-1]}$ from $Y_{1,[l-1]}$. The resulting vector is

$$\tilde{\mathbf{Y}}^n = \mathbf{X}^n_{31,[l-1]} \oplus \tilde{N}^n_\delta.$$

In this situation, the channel from X_{31} to $\tilde{Y}$ is a binary additive channel with δ as the bias of the noise. At the third step, the decoder uses $\tilde{\mathbf{Y}}^n$ to decode the message of the third user, i.e., $\mathbf{W}_{3,[l-1]}^k$. In particular, the decoder finds unique $\tilde{\mathbf{w}}_3^k \in \mathbb{F}_2^k$ such that $(\tilde{\mathbf{w}}_3^k \mathbf{G}, \tilde{\mathbf{Y}}^n)$ are jointly ϵ-typical with respect to $P_{X_{31}\tilde{Y}}$. Otherwise, an error event $E_{3,[l]}$ is declared.

Error Analysis: We can show that this problem is equivalent to a point-to-point channel coding problem, where the channel is described by $Z = X \oplus \tilde{N}_\delta$. The average probability of error approaches zero, if $\frac{k}{n} \leq 1 - h(\delta)$.

Suppose there is no error in the decoding process of the third user. That is $E_{1,[l]}^c$ occurs. Therefore, $\mathbf{X}_{32,[l]}^n = \mathbf{X}_{22,[l]}^n \oplus \mathbf{X}_{12,[l]}^n$ with probability one. As a result, the channel in Fig. 4.4 is in the first state. This implies that the corresponding channel consists of two parallel binary additive channel with independent noises and bias δ. Similar to the argument for E_1, it can be shown that $P(E_{2,[l]}|E_{1,[l]}) \to 0$, if $\frac{k}{n} \leq 1 - h(\delta)$. Lastly, we can show that conditioned on $E_{1,[l]}^c$ and $E_{2,[l]}^c$, the probability of $E_{3,[l]}$ approaches zero, if $\frac{k}{n} \leq 1 - h(\delta)$.

As a result of the above argument, the average probability of error approaches 0, if $\frac{k}{n} \leq 1 - h(\delta)$. This implies that the rates $R_i = 1 - h(\delta), i = 1, 2, 3$ are achievable, and the proof is completed. □

C.3 Proof of Lemma 14

Proof. Let R_i be the rate of the ith encoder. We have $R_i \geq 1 - h(\delta) - \epsilon$. We apply the generalized Fano's inequality (Lemma 4.3 in [47]) for decoding of the messages. More precisely, as $\bar{P} \leq \epsilon$, we have

$$\frac{1}{M_1 M_2 M_3} H(\Theta_1, \Theta_2, \Theta_3 | \mathbf{Y}^N) \leq h(\bar{P}) \leq h(\epsilon)$$

By the definition of the rate we have

$$\begin{aligned} R_1 + R_2 + R_3 &= \frac{1}{N} H(\Theta_1, \Theta_2, \Theta_3) \\ &\leq \frac{1}{N} I(\Theta_1, \Theta_2, \Theta_3; \mathbf{Y}^n) + o(\epsilon) \\ &\overset{(a)}{\leq} \frac{1}{N} I(\mathbf{X}_1^n, \mathbf{X}_2^n, \mathbf{X}_3^n; \mathbf{Y}^N) + o(\epsilon) \\ &\overset{(b)}{\leq} 3 - \frac{1}{N} H(\mathbf{Y}^n | \mathbf{X}^n) + o(\epsilon), \end{aligned} \tag{C.4}$$

where (a) is because of (6.2), and for (b) we use the fact that Y is a vector of three binary random variables, which implies $\frac{1}{N} H(Y^N) \leq 3$. As the channel is memoryless, and since (6.2) holds, we have

$$\frac{1}{N} H(\mathbf{Y}^n | \mathbf{X}^n) = \frac{1}{N} \sum_{l=1}^{N} H(Y_l | X_{1,l} X_{2,l} X_{3,l}).$$

Let $P(X_{32,l} \neq X_{12,l} \oplus X_{12,l}) = q_l$, for $l \in [1 : N]$. Denote $\bar{q}_l = 1 - q_l$. We can show that,

$$H(Y_l | X_{1,l} X_{2,l} X_{3,l}) = (1 + 2\bar{q}_l) h(\delta) + 2q_l.$$

We use the above argument, and the last inequality in (C.4) to give the following bound

$$\begin{aligned} R_1 + R_2 + R_3 &\leq 3 - \frac{1}{N} \sum_{l=1}^{N} [(1 + 2\bar{q}_l) h(\delta) + 2q_l] + o(\epsilon) \\ &= 3 - 3h(\delta) + \frac{1}{N} 2(1 - h(\delta)) \sum_{l=1}^{N} q_l + o(\epsilon) \end{aligned}$$

By assumption $R_1 + R_2 + R_3 \geq 3(1 - h(\delta) - \epsilon)$. Therefore, using the above bound we

obtain,

$$\frac{3\epsilon + o(\epsilon)}{2(1-h(\delta))} \geq \frac{1}{N}\sum_{l=1}^{N} q_l \overset{(a)}{\geq} \frac{1}{N}\sum_{l\in\mathcal{I}_c^N} q_l,$$

where (a) holds, because we remove the summation over all $l \notin \mathcal{I}_c^N$. We defined $\mathcal{I}_c^N$ as in the statement of this Lemma. Note that if $l \in \mathcal{I}_c^N$, then $q_l \geq c$. Finally, we obtain

$$\frac{|\mathcal{I}_c^N|}{N} \leq \frac{3\epsilon + o(\epsilon)}{2c(1-h(\delta))}$$

□

C.4 Proof of Lemma 15

Proof. Let $\mathcal{I}_c^N$ be as in Lemma 14. The average probability of error for decoding $X_{12}^N \oplus X_{22}^N$ is bounded as

$$\begin{aligned}
\bar{P}_e &= \frac{1}{N}\sum_{l=1}^{N} P(X_{32,l} \neq X_{12,l} \oplus X_{22,l}) \\
&= \frac{1}{N}\sum_{l\in\mathcal{I}_c^N} P(X_{32,l} \neq X_{12,l} \oplus X_{22,l}) + \frac{1}{N}\sum_{l\notin\mathcal{L}_c^N} P(X_{32,l} \neq X_{12,l} \oplus X_{22,l}) \\
&\leq \frac{|\mathcal{I}_c^N|}{N} + c(1 - \frac{|\mathcal{I}_c^N|}{N}) \\
&= (1-c)\frac{|\mathcal{I}_c^N|}{N} + c \\
&\leq (1-c)\frac{\eta(\epsilon)}{2c(1-h(\delta))} + c
\end{aligned}$$

As a result as $\epsilon \to 0$, then $\bar{P}_e \to c$. Since $c > 0$ is arbitrary, $\bar{P}_e$ can be made arbitrary small. Hence, for any $\epsilon' > 0$, and there exist $\epsilon > 0$ and large enough N such that $\bar{P}_e < \epsilon'$. Note that X_{32}^N is a function of M_3, Y_1^N, Y_{12}^N and Y_{22}^N. Next we argue that to get $\bar{P}_e < \epsilon'$, it is enough for X_{32}^N to be a function of M_3, Y_1^N. More precisely,

given $X_{32,l}$, the random variables $Y_{12,l}$ and $Y_{22,l}$ are independent of $X_{12,l} \oplus X_{22,l}$. To see this, we need to consider two cases. If $X_{32,l} = X_{12,l} \oplus X_{22,l}$ then the argument follows trivially. Otherwise, $Y_{12,l} = X_{12,l} \oplus N_{1/2}$, where $N_{1/2} \sim Ber(1/2)$, and it is independent of $X_{12,l}$. Hence in this case, $Y_{12,l}$ is independent of $X_{12,l}$. Similarly, $Y_{22,l}$ is independent of $X_{22,l}$.

By subtracting X_{31}^N from Y_1^N, we get $Z^N := X_{11}^N \oplus X_{21}^N \oplus N_\delta^N$. Next, we argue that the third encoder uses Z^N to decode $X_{12}^N \oplus X_{22}^N$. Since M_3 is independent of M_1 and M_2, it is independent of X_{1j}^N, X_{j2}^N for $j = 1, 2$. Therefore Z^N is independent of M_3. Hence, X_{32}^N is function of Z^N. Intuitively, we convert the problem of decoding $X_{11}^N \oplus X_{21}^N$ to a point to point channel coding problem. The channel in this case is a binary additive channel with noise $N_\delta \sim Ber(\delta)$. In this channel coding problem the codebook at the encoder is $\mathcal{C}_{12} \oplus \mathcal{C}_{22}$. The capacity of this channel equals $1 - h_b(\delta)$. Since the average probability of error is small, we can use the generalized Fano's inequality to bound the rate of the encoder. As a result, it can be shown that

$$\frac{1}{N} \log_2 ||\mathcal{C}_{12} \oplus \mathcal{C}_{22}|| \leq 1 - h_b(\delta) + \eta(\epsilon), \tag{C.5}$$

where $\eta(\epsilon) \to 0$ as $\epsilon \to 0$.

Claim 3. *The following bound holds*

$$\frac{1}{N} \log_2 ||\mathcal{C}_{j2}|| \geq 1 - h_b(\delta) - \gamma_j(\epsilon), \tag{C.6}$$

where $j = 1, 2$ and $\gamma_j(\epsilon) \to 0$ as $\epsilon \to 0$.

Outline of the proof. First, we show that the decoder must decode M_3 from Y_1^N. We argued in the above that X_{32}^N is independent of M_3. Hence, the message M_3 is encoded only to X_{31}^N. Since X_{31}^N is sent though the first channel in Example 1, the decoder must decode M_3 from Y_1^N. Next, we argue that the receiver must decode M_1 and M_2 from

Y_{21}^N and Y_{22}^N, respectively. Note that the rate of the third encoder is $1 - h_b(\delta)$, which equals to the capacity of the first channel given $X_{11}^N \oplus X_{21}^N$. Therefore, the decoder can decode M_3, if it has $X_{11}^N \oplus X_{21}^N$. Hence, the decoder must reconstruct $X_{11}^N \oplus X_{21}^N$ from the second channel. It can be shown that this is possible, if the decoder can decode M_1 and M_2 from the second channel. As a result, from Fano's inequality, the bounds in the Claim hold. □

Finally, using (7) and (C.6) we get

$$0 \leq \frac{1}{N} \log_2 ||\mathcal{C}_{12} \oplus \mathcal{C}_{22}|| - \frac{1}{N} \log_2 ||\mathcal{C}_{j2}|| \leq \eta(\epsilon) + \gamma_j(\epsilon), \quad j = 1, 2.$$

This completes the proof. □

APPENDIX D

Proofs for Chapter V

D.1 Proof of Theorem V.2

Proof. The pair of the nested linear codes are denoted by $(\mathcal{C}_i, \mathcal{C}_o)$ and $(\mathcal{C}_i', \mathcal{C}_o')$. For any ϵ-typical sequence $\mathbf{x} \in \mathcal{X}^n$ with respect to P_X, define

$$\theta(x) \triangleq \sum_{\mathbf{u}\in\mathcal{C}_o, \mathbf{v}\in\mathcal{C}_o'} \mathbb{1}\{(\mathbf{u}, \mathbf{v}) \in A_\epsilon^n(U, V|x)\}$$

Note that $\theta(\mathbf{x})$ equals the number of codewords $(\mathbf{u}, \mathbf{v})$ selected from the two nested linear codes such that $(\mathbf{u}, \mathbf{v}, \mathbf{x})$ are jointly ϵ-typical with respect to $P_{U,V,X}$. Assume the generator matrices and the dither for $(\mathcal{C}_i, \mathcal{C}_o)$ are denoted by $(\mathbf{G}, \mathbf{\Delta G})$ and $\mathbf{b}$, respectively. Also, let $(\mathbf{G}, \mathbf{\Delta G}')$ and $\mathbf{b}'$ denote the generator matrices and the dither of $(\mathcal{C}_i', \mathcal{C}_o')$, respectively. Note that the inner codes share the same generator matrix. With this notation $\delta(\mathbf{x})$ can be written as

$$\theta(\mathbf{x}) = \sum_{\mathbf{m}\in\mathbb{F}_q^l, \mathbf{m}'\in\mathbb{F}_q^{l'}} \sum_{\mathbf{a},\mathbf{a}'\in\mathbb{F}_q^k} \sum_{(\mathbf{u},\mathbf{v})\in A_\epsilon^n(U,V|x)} \mathbb{1}\{\mathbf{aG} + \mathbf{m\Delta G} + \mathbf{b} = \mathbf{u}, \mathbf{aG} + \mathbf{m}'\mathbf{\Delta G} + \mathbf{b}' = \mathbf{v}\}$$

Suppose the elements of the generator matrices and the dithers are selected randomly independently and uniformly from $\mathbb{F}_q$. Then the following lemma holds.

Lemma 35. *For any* $\mathbf{a} \in \mathbb{F}_q^k$ *and* $\mathbf{m} \in \mathbb{F}_q^l$, *define* $g(\mathbf{a}, \mathbf{m}) \triangleq \mathbf{aG} + \mathbf{m\Delta G} + \mathbf{b}$. *Similarly define* $g'(\mathbf{a}', \mathbf{m}') \triangleq \mathbf{a}'\mathbf{G} + \mathbf{m}'\mathbf{\Delta G}' + \mathbf{b}'$ *where* $\mathbf{a}' \in \mathbb{F}_q^k$ *and* $\mathbf{m}' \in \mathbb{F}_q^l$. *Suppose the elements of the matrices and the dithers are selected randomly independently and uniformly from* $\mathbb{F}_q$. *Then the followings hold:*

1. $g(a, m)$ *and* $g'(b, m')$ *are uniform over* $\mathbb{F}_q^n$.
2. $g(a, m)$ *is independent of* $g(\tilde{a}, m)$ *when* $a \neq \tilde{a}$.
3. $g'(b, m')$ *is independent of* $g'(\tilde{b}, m')$ *when* $b \neq \tilde{b}$.
4. $g(a, m)$ *and* $g'(a, m)$ *are independent.*

Proof. Follows from [67] Lemma III.2 and III.3 and the fact that $\mathbf{b}, \mathbf{b}'$ are independent and uniform. □

The theorem follows by showing that $\theta(\mathbf{x}) = 0$ with probability sufficiently close to one. In what follows we prove a stronger statement. We show that $\theta(\mathbf{x}) \geq \frac{1}{2}\mathbb{E}[\theta(\mathbf{x})]$ with probability approaching one. For that applying Chebyshev's inequality gives

$$\begin{aligned}\mathbb{P}\Big\{\theta(x) \leq \frac{1}{2}\mathbb{E}[\theta(x)]\Big\} &\leq \mathbb{P}\Big\{\big|\theta(x) - \mathbb{E}[\theta(x)]\big| \geq \frac{1}{2}\mathbb{E}[\theta(x)]\Big\} \\ &\leq \frac{4\mathrm{var}[\theta(x)]}{\mathbb{E}[\theta(x)]^2}\end{aligned}$$

The expectation of $\theta(x)$ equals

$$\begin{aligned}\mathbb{E}\{\theta(x)\} &= \sum_{\mathbf{m}\in\mathbb{F}_q^l, \mathbf{m}'\in\mathbb{F}_q^{l'}} \sum_{\mathbf{a},\mathbf{a}'\in\mathbb{F}_q^k} \sum_{(\mathbf{u},\mathbf{v})\in A_\epsilon^n(U,V|x)} \mathbb{P}\{g(\mathbf{a}, \mathbf{m}) = \mathbf{u}, g'(\mathbf{a}', \mathbf{m}') = \mathbf{v}\} \\ &= \sum_{\mathbf{m}\in\mathbb{F}_q^l, \mathbf{m}'\in\mathbb{F}_q^{l'}} \sum_{\mathbf{a},\mathbf{a}'\in\mathbb{F}_q^k} \sum_{(\mathbf{u},\mathbf{v})\in A_\epsilon^n(U,V|x)} \frac{1}{q^{2n}} \\ &\leq \frac{q^{l+l'}q^{2k}}{q^{2n}} 2^{n(H(U,V|X)+O(\epsilon))}\end{aligned}$$

where the first equality follows by Lemma 35. Next, we calculate $\mathbb{E}[\theta(x)^2]$. For that we have

$$\mathbb{E}[\theta(x)^2] = \sum_{\substack{\mathbf{m},\tilde{\mathbf{m}}\in\mathbb{F}_q^l,\\ \mathbf{m}',\tilde{\mathbf{m}}'\in\mathbb{F}_q^{l'}}} \sum_{\substack{\mathbf{a},\tilde{\mathbf{a}}\in\mathbb{F}_q^k,\\ \mathbf{a}',\tilde{\mathbf{a}}'\in\mathbb{F}_q^k}} \sum_{\substack{(\mathbf{u},\mathbf{v})\in A_\epsilon^n(U,V|x)\\ (\tilde{\mathbf{u}},\tilde{\mathbf{v}})\in A_\epsilon^n(U,V|x)}} \mathbb{P}\{g(\mathbf{a},\mathbf{m})=\mathbf{u}, g'(\mathbf{a}',\mathbf{m}')=\mathbf{v}, g(\tilde{\mathbf{a}},\tilde{\mathbf{m}})=\tilde{\mathbf{u}}, g'(\tilde{\mathbf{a}}',\tilde{\mathbf{m}}')=\tilde{\mathbf{v}}\}$$

Note that from Lemma 35, the probability above equals

$$\frac{1}{q^{2n}}\mathbb{P}\{g_0(\mathbf{a}-\tilde{\mathbf{a}},\mathbf{m}-\tilde{\mathbf{m}})=\mathbf{u}-\tilde{\mathbf{u}}, g_0'(\mathbf{a}'-\tilde{\mathbf{a}}',\mathbf{m}'-\tilde{\mathbf{m}}')=\mathbf{v}-\tilde{\mathbf{v}}\}$$

where $g_0(a,m)=a\mathbf{G}+m\mathbf{\Delta G}, g_0'(a',m')=a'\mathbf{G}+m'\mathbf{\Delta G}'$. The following cases can be considered regarding the value of the above probability

Cases 1: $\mathbf{m}=\tilde{\mathbf{m}}, \mathbf{m}'=\tilde{\mathbf{m}}'$:

Case 1.1 $a=\tilde{a}, a'=\tilde{a}' \Rightarrow \mathbb{P}=\frac{1}{q^{2n}}\mathbb{1}\{u=\tilde{u}, v=\tilde{v}\}$

Case 1.2 $a=\tilde{a}, a'\neq\tilde{a}' \Rightarrow \mathbb{P}=\frac{1}{q^{3n}}\mathbb{1}\{u=\tilde{u}\}$

Case 1.3 $a\neq\tilde{a}, a'=\tilde{a}' \Rightarrow \mathbb{P}=\frac{1}{q^{3n}}\mathbb{1}\{v=\tilde{v}\}$

Case 1.4 $a\neq\tilde{a}, a'\neq\tilde{a}', (a-\tilde{a})=i(a'-\tilde{a}'), i\in\mathbb{F}_q-\{0\} \Rightarrow \mathbb{P}=\frac{1}{q^{3n}}\mathbb{1}\{(u-\tilde{u})=(iv-\tilde{v})\}$

Case 1.5 $a\neq\tilde{a}, a'\neq\tilde{a}', (a-\tilde{a})\neq i(a'-\tilde{a}')\forall i\in\mathbb{F}_q \Rightarrow \mathbb{P}=\frac{1}{q^{4n}}$

Cases 2: $\mathbf{m}\neq\tilde{\mathbf{m}}, \mathbf{m}'=\tilde{\mathbf{m}}'$:

Case 2.1 $a=\tilde{a}, a'=\tilde{a}' \Rightarrow \mathbb{P}=\frac{1}{q^{3n}}\mathbb{1}\{v=\tilde{v}\}$

Case 2.2 $a=\tilde{a}, a'\neq\tilde{a}' \Rightarrow \mathbb{P}=\frac{1}{q^{4n}}$

Case 2.3 $a\neq\tilde{a}, a'=\tilde{a}' \Rightarrow \mathbb{P}=\frac{1}{q^{3n}}\mathbb{1}\{v=\tilde{v}\}$

Case 2.4 $a\neq\tilde{a}, a'\neq\tilde{a}' \Rightarrow \mathbb{P}=\frac{1}{q^{4n}}$

Cases when $m = \tilde{m}, m' \neq \tilde{m}'$ and $m \neq \tilde{m}, m' \neq \tilde{m}'$ are very similar and will be discussed shortly. Considering the above cases when $m = \tilde{m}, m' = \tilde{m}'$ it gives:

$$\mathbb{E}\{\theta(X)^2 | m = \tilde{m}, m' = \tilde{m}'\} =$$
$$\sum_{m,m'} \Bigg[\sum_{a=\tilde{a}} \sum_{b=\tilde{b}} \sum_{(u,v)\in A_\epsilon^n(U,V|x)} \frac{1}{q^{2n}} + \sum_{a=\tilde{a}} \sum_{b\neq\tilde{b}} \sum_{(u,v),(u,\tilde{v})\in A_\epsilon^n(U,V|x)} \frac{1}{q^{3n}}$$
$$+ \sum_{a\neq\tilde{a}} \sum_{b=\tilde{b}} \sum_{(u,v),(\tilde{u},v)\in A_\epsilon^n(U,V|x)} \frac{1}{q^{3n}} + \sum_{i\in\mathbb{F}_q-\{0\}} \sum_{a\neq\tilde{a}} \sum_{\substack{b\neq\tilde{b} \\ b-\tilde{b}=i(a-\tilde{a})}} \sum_{\substack{(u,v),(\tilde{u},\tilde{v})\in A_\epsilon^n(V|x) \\ v-\tilde{v}=i(u-\tilde{u})}} \frac{1}{q^{3n}}$$
$$+ \sum_{a\neq\tilde{a}} \sum_{\substack{b\neq\tilde{b} \\ b-\tilde{b}\neq i(a-\tilde{a}) \\ \forall i\in\mathbb{F}_q-0}} \sum_{\substack{(u,v),(\tilde{u},\tilde{v})\in A_\epsilon^n(V|x) \\ v-\tilde{v}\neq i(u-\tilde{u})}} \frac{1}{q^{4n}} \Bigg]$$

Consequently

$$\begin{aligned}\mathbb{E}\{\theta(X)^2 | m = \tilde{m}, m' = \tilde{m}'\} \leq{}& \frac{q^{l+l'}q^{2k}}{q^{2n}} 2^{n(H(U,V|X)+O(\epsilon))} + \frac{q^{l+l'}q^{3k}}{q^{3n}} 2^{n(H(U,V|X)+H(V|X,U)+O(\epsilon))} \\ &+ \frac{q^{l+l'}q^{3k}}{q^{3n}} 2^{n(H(U,V|X)+H(U|X,V)+O(\epsilon))} \\ &+ \frac{q^{l+l'}q^{3k}}{q^{3n}} 2^{n(H(U,V|X)+\max_{i\neq 0} H(U,V|X,V+iU)+O(\epsilon))} \\ &+ \frac{q^{l+l'}q^{4k}}{q^{4n}} 2^{2n(H(U,V|X)+O(\epsilon))}\end{aligned}$$

For the case where $(m \neq \tilde{m}, m' = \tilde{m}')$ using a similar argument we obtain

$$\begin{aligned}\mathbb{E}\{\theta(X)^2 | m \neq \tilde{m}, m' = \tilde{m}'\} \leq{}& \frac{q^{2l+l'}q^{2k}}{q^{3n}} 2^{n(H(U,V|X)+H(U|X,V)+O(\epsilon))} \\ &+ \frac{q^{2l+l'}q^{3k}}{q^{4n}} 2^{n(2H(U,V|X)+O(\epsilon))} \\ &+ \frac{q^{2l+l'}q^{4k}}{q^{4n}} 2^{n(2H(U,V|X)+O(\epsilon))}\end{aligned}$$

Similarly for the case $(m = \tilde{m}, m' \neq \tilde{m}')$ we have

$$\begin{aligned}\mathbb{E}\{\theta(X)^2 | m = \tilde{m}, m' \neq \tilde{m}'\} \leq & \frac{q^{l+2l'}q^{2k}}{q^{3n}} 2^{n(H(U,V|X)+H(V|X,U)+O(\epsilon))} \\ & + \frac{q^{l+2l'}q^{3k}}{q^{4n}} 2^{n(2H(U,V|X)+O(\epsilon))} \\ & + \frac{q^{l+2l'}q^{4k}}{q^{4n}} 2^{n(2H(U,V|X)+O(\epsilon))}\end{aligned}$$

When $(m \neq \tilde{m}, m' \neq \tilde{m}')$, for any value of $a, \tilde{a}, b, \tilde{b}$ we have $\mathbb{P} = \frac{1}{q^{4n}}$. The reason is that $(m - \tilde{m})\mathbf{\Delta G}$ and $(m' - \tilde{m}')\mathbf{\Delta G}'$ are independent and uniform over $\mathbb{F}_q^n$. Therefore, for this case we have:

$$\mathbb{E}\{\theta(X)^2 | m = \tilde{m}, m' \neq \tilde{m}'\} \leq \frac{q^{2l+2l'}q^{4k}}{q^{4n}} 2^{n(2H(U,V|X)+O(\epsilon))}$$

Finally we have:

$$\begin{aligned}\frac{\text{var}\{\theta(x)\}}{\mathbb{E}\{\theta(x)\}^2} \leq & \frac{q^{2n}}{q^{l+l'}q^{2k}} 2^{-n(H(U,V|X)+O(\epsilon))} + \frac{q^n}{q^{l+l'}q^k} 2^{-n(H(U|X)+O(\epsilon))} \\ & + \frac{q^n}{q^{l+l'}q^k} 2^{-n(H(V|X)+O(\epsilon))} \\ & + \frac{q^n}{q^{l+l'}q^k} 2^{-n(H(U,V|X)-\max_{i\neq 0} H(U,V|X,V+iU)+O(\epsilon))} \\ & + \frac{q^n}{q^l q^k} 2^{-n(H(U|X)+O(\epsilon))} + \frac{q^n}{q^{l'}q^k} 2^{-n(H(V|X)+O(\epsilon))} \\ & + \frac{1}{q^l} + \frac{1}{q^{l'}} + \frac{1}{q^{l+l'}} + \frac{1}{q^{l+k}} + \frac{1}{q^{l'+k}}\end{aligned}$$

Let r_o, r_o' denote the rate of $\mathcal{C}_o$ and $\mathcal{C}_o'$, respectively. Also let r_i, r_i' denote the rate of the inner-codes $\mathcal{C}_i$ and $\mathcal{C}_i'$, respectively. Then $\mathbb{P}\{\theta(x) = 0\} \to 0$ as $n \to \infty$ if the

following bounds are satisfied

$$
\begin{aligned}
r_o + r'_o &\geq 2\log p - H(U,V|X), \\
r_o + r'_o - r_i &\geq \log p - H(U|X) \\
r_o + r'_o - r_i &\geq \log p - H(V|X) \\
r_o + r'_o - r_i &\geq \log p - H(U,V|X) + \max_{i\neq 0} H(U,V|X,V+iU) \\
r_o &\geq \log p - H(U|X), \qquad r_o \geq r_i, \qquad r_o \geq 0 \\
r'_o &\geq \log p - H(V|X), \qquad r'_o \geq r_i, \qquad r'_o \geq 0
\end{aligned}
$$

Observe that

$$
\begin{aligned}
H(U,V|X,V+iU) &= H(U,V,V+iU|X) - H(V+iU|X) \\
&= H(U,V|X) - H(V+iU|X)
\end{aligned}
$$

Note that the second and third inequalities are redundant. This is because $r_o \geq r_i$ and $r'_o \geq r_i$. As a result, the above bounds are simplified to the set of bounds given in the statement of the theorem.

□

APPENDIX E

Proofs for Chapter VI

E.1 Proof of Theorem VI.1

Proof. At each block a re-transmission occurs with probability q, an error occurs with probability P_{eb} and a correct decoding process happens with probability $1 - q - P_{eb}$. The probability of a re-transmission at each block is

$$q = P(\hat{\Theta} = \Theta_1).$$

The probability of error at each block is

$$P_{eb} = P(\Theta_1)P(\hat{\Theta} = \Theta_0|\Theta_1).$$

Therefore, with this setting the total probability of error for the transmission of a message is

$$P_e = \sum_{k=0}^{\infty} q^k P_{eb} = \frac{P_{eb}}{1-q}. \tag{E.1}$$

The number of blocks required to complete the transmission of one message is a geometric random variable with probability of success $1-q$. Thus, the expected number of blocks for transmission of a message is $\frac{1}{1-q}$.

Next, we derive an upper-bound for q and P_{eb}. For shorthand, denote $H_{12} = (H_1, H_2), \hat{H}_{12} = (\hat{H}_1, \hat{H}_2)$. Then

$$\begin{aligned} P_{eb} &= P\left(\hat{H}_{12} = 00, H_{12} \neq 00\right) \\ &= \sum_{a \in \{01,10,11\}} P(H_{12} = a) P(\hat{H}_{12} = 00 | H_{12} = a). \end{aligned}$$

Note that the effective rates of this transmission scheme are $(\frac{R_1}{1-\gamma}, \frac{R_2}{1-\gamma})$. Suppose $(\frac{R_1}{1-\gamma}, \frac{R_2}{1-\gamma})$ is inside the feedback-capacity region of the channel. Then, from the definition of the capacity region, there exist a sequence $\zeta_n, n \geq 1$ with $\zeta_n \to 0$ such that after the first stage

$$P((\hat{W}_1, \hat{W}_2) \neq (W_1, W_2)) \leq \zeta_n.$$

Equivalently, the effective rates are inside the capacity region, if the following inequality holds for any $\lambda_i \geq 0, i = 1, 2, 3$:

$$\frac{1}{1-\gamma} \left(\lambda_1 R_1 + \lambda_2 R_2 + \lambda_3 (R_1 + R_2)\right) < C_{\underline{\lambda}}, \tag{E.2}$$

where $C_{\underline{\lambda}}$ is given in Definition 42. Denote $R_3 = R_1 + R_2$ and define

$$\gamma^* = \min_{\substack{\lambda_1, \lambda_2, \lambda_3 \geq 0 \\ \lambda_1 + \lambda_2 + \lambda_3 = 1}} \left(1 - \frac{\sum_i \lambda_i R_i}{C_{\underline{\lambda}}}\right). \tag{E.3}$$

Then, (E.2) implies that $\gamma < \gamma^*$. The probability of error is therefore bounded by

$$P_{eb} \leq \sum_{a \in \{01,10,11\}} P(\hat{H}_{12} = 00 | H_{12} = a) \tag{E.4}$$

Suppose $(X_1(0), X_1(1), X_2(0), X_2(1))$ are random variables with joint distribution P_n. Then for $i, j \in \{0, 1\}$ define

$$\bar{D}_{\mathsf{P}_n}(00||ij) = \mathbb{E}_{\mathsf{P}_n}\Big[D_Q\big(X_1(0), X_2(0)||X_1(i), X_2(j)\big)\Big].$$

From the description of the transmission scheme, the codewords for the confirmation stage are selected with joint-type P_n. In addition, the decoding process is performed using ML decoding. Therefore, the following bounds hold for $a \in \{01, 10, 11\}$:

$$P(\hat{H}_{12} = 00 | H_{12} = a) \leq 2^{-n\gamma \bar{D}_{\mathsf{P}_n}(00||a)}.$$

Thus, from (E.4), the probability of error is upper bounded by

$$P_{eb} \leq 3 \times 2^{-n\gamma D_{l,n}} \tag{E.5}$$

Where $D_{l,n} = \max_{\mathsf{P}_n} \min_{a \in \{01,10,11\}} \bar{D}_{\mathsf{P}_n}(00||a)$.

Next we derive an upper bound for q. We have

$$\begin{aligned} q &= P(\hat{\Theta} = \Theta_1) \\ &= P(\Theta_0)P(\hat{\Theta} = \Theta_1|\Theta_0) + P(\Theta_1)P(\hat{\Theta} = \Theta_1|\Theta_1) \\ &\leq P(\hat{\Theta} = \Theta_1|\Theta_0) + \zeta_n, \end{aligned}$$

where the last inequality holds because of the following inequalities 1) $P(\Theta_1) \leq \zeta_n$,

and 2) $P(\Theta_0), P(\hat{\Theta} = \Theta_1|\Theta_1) \leq 1$. Note that

$$\begin{aligned} P(\hat{\Theta} = \Theta_1|\Theta_0) &= \sum_{a\in\{01,10,11\}} P(\hat{H}_{12} = a|H_{12} = 00) \\ &\leq \sum_{a\in\{01,10,11\}} 2^{-n\gamma \bar{D}_{\mathsf{P}_n}(a||00)} \\ &\leq 3 \times 2^{-n\gamma \tilde{D}_{l,n}}, \end{aligned}$$

where $\tilde{D}_{l,n} = \min_{a\in\{01,10,11\}} \bar{D}_{\mathsf{P}_n}(a||00)$. Therefore, there exists a sequence $\{q_n\}_{n\geq 1}$ with $q_n \to 0$ such that $q < q_n + \zeta_n$. Using this inequality and the inequality at (E.5), we derive the following upper-bound for the total probability of error given in (E.1)

$$P_e \leq \frac{3}{1 - q_n - \zeta_n} 2^{-n\gamma D_{l,n}}.$$

Therefore, the error exponent is bounded from below as

$$\frac{-\log_2 P_e}{\mathbb{E}[T]} \geq \sup \frac{\gamma D_{l,n}}{(1 - q_n - \zeta_n)} + \xi_n$$

where $\xi_n = \frac{1}{n} \frac{\log_2(\frac{1-q_n-\zeta_n}{3})}{1-q_n-\zeta_n}$. Note that for any $\epsilon > 0$ there exists large enough n such that $q_n + \zeta_n < \epsilon, D_{l,n} > D_l - \epsilon, \xi_n < \epsilon$. Set $\gamma = \gamma^* - \epsilon$. Then

$$\frac{-\log_2 P_e}{\mathbb{E}[T]} \geq \gamma^* D_l - \sigma(\epsilon)$$

where σ is a function of ϵ such that $\lim_{\epsilon\to 0} \sigma(\epsilon) = 0$. Finally, the proof is complete by replacing γ^* from (E.3).

□

E.2 Proof of Lemma 20

Proof. Given $Y^t = y^t, W_1 = m_1, W_2 = m_2$, we obtain

$$\begin{aligned}
&\mathbb{E}[H^1_{t+1} - H^1_t | m_2, y^t] \\
&\quad = -I(W_1; Y_{t+1} | m_2, y^t) \\
&\quad = -I(W_1; Y_{t+1} | m_2, x_2^{t+1}, y^t) \\
&\quad = -H(Y_{t+1} | m_2, x_2^{t+1}, y^t) + H(Y_{t+1} | m_2, x_2^{t+1}, W_1, y^t) \\
&\quad = -H(Y_{t+1} | m_2, x_2^{t+1}, y^t) \\
&\qquad + H(Y_{t+1} | m_2, x_2^{t+1}, W_1, X_1^{t+1}, y^t) \\
&\quad \overset{(a)}{=} -H(Y_{t+1} | m_2, x_2^{t+1}, y^t) + H(Y_{t+1} | x_2^{t+1}, X_1^{t+1}, y^t) \\
&\quad \triangleq -J^1_{t+1}(m_2, x_2^{t+1}, y^t)
\end{aligned} \tag{E.6}$$

where (a) follows because condition on the channel inputs $X_{1,t+1}, X_{2,t+1}$, the output Y_{t+1} is independent of W_1, W_2. We denote the right-hand side of (a) by $J^1_{t+1}(.)$ as in (E.6). Similarly for the case when $i = 2$ the following lower-bound holds

$$\begin{aligned}
&\mathbb{E}[H^2_{t+1} - H^2_t | m_1, y^t] \\
&\quad = -H(Y_{t+1} | m_1, x_1^{t+1}, y^t) + H(Y_{t+1} | X_2^{t+1}, x_1^{t+1}, y^t) \\
&\quad \triangleq -J^2_{t+1}(m_1, x_1^{t+1}, y^t).
\end{aligned} \tag{E.7}$$

Using a similar argument for the case when $i = 3$, we can show that the following inequality holds

$$\begin{aligned}
\mathbb{E}[H^3_{t+1} - H^3_t | y^t] &\geq -I(X_1^{t+1}, X_2^{t+1}; Y_{t+1} | y^t) \\
&\triangleq -J^3_{t+1}(y^t).
\end{aligned} \tag{E.8}$$

Consider the quantities at the right-hand side of (E.6), (E.7) and (E.8), i.e., the functions $J^1_{t+1}, J^2_{t+1}, J^3_{t+1}$. We proceed by the following lemma.

Lemma 36. *The vector* $(J^1_{t+1}, J^2_{t+1}, J^3_{t+1})$ *is inside the feedback-capacity region* $\mathcal{C}$ *almost surely.*

Proof. We use the alternative representation for $\mathcal{C}$ which is given in Fact 1. For any non-negative numbers $\lambda_1, \lambda_2, \lambda_3$, let

$$J_\lambda(m_1, m_2, x_1^{t+1}, x_2^{t+1}, y^t) = \lambda_1 J^1_{t+1}(m_2, x_2^{t+1}, y^t) + \lambda_2 J^2_{t+1}(m_1, x_1^{t+1}, y^t) + \lambda_3 J^3_{t+1}(y^t)$$

Note that

$$J_\lambda(m_1,m_2, x_1^{t+1}, x_2^{t+1}, y^t) \leq \sup_{P_{W_1 W_2 X_1^{t+1} X_2^{t+1}|Y^{t+1}}} \mathbb{E}\{J_\lambda(W_1, W_2, X_1^{t+1}, X_2^{t+1}, y^t)\}, \quad \text{(E.9)}$$

where the supremum is taken over all $P_{X_1^{t+1} X_2^{t+1}|Y^{t+1}}$ that factors as in (6.4). The right-hand side of the above inequality equals $\sum_i \mathbb{E}[\lambda_i J^i_{t+1}]$. Each expectation inside the summation can be bounded as follows

$$\begin{aligned}
\mathbb{E}\{J^1_{t+1}(W_2, X_2^{t+1}, y^t)\} &= H(Y_{t+1}|W_2, X_2^{t+1}, y^t) - H(Y_{t+1}|X_2^{t+1}, X_1^{t+1}, y^t) \\
&\leq H(Y_{t+1}|X_2^{t+1}, y^t) - H(Y_{t+1}|X_2^{t+1}, X_1^{t+1}, y^t) \\
&= I(X_{1,t+1}, Y_{t+1}|X_2^{t+1}, y^t)
\end{aligned}$$

Similarly,

$$\begin{aligned}
\mathbb{E}\{J^2_{t+1}(W_1, X_1^{t+1}, y^t)\} &\leq I(X_{2,t+1}; Y_{t+1}|X_1^{t+1} y^t) \\
\mathbb{E}\{J^3_{t+1}(y^t)\} &\leq I(X_{1,t+1}, X_{2,t+1}; Y_{t+1}|y^t)
\end{aligned}$$

Therefore, since the channel is memoryless using the above bounds we have

$$\begin{aligned}
&\mathbb{E}\{J_\lambda(W_1, W_2, X_1^{t+1}, X_2^{t+1}, y^t)\} \\
&\leq \lambda_1 I(X_{1,t+1}, Y_{t+1}|X_2^{t+1}, y^t) + \lambda_2 I(X_{2,t+1}; Y_{t+1}|X_1^{t+1} y^t) + \lambda_3 I(X_{1,t+1}, X_{2,t+1}; Y_{t+1}|y^t) \\
&\leq C_{\underline{\lambda}}
\end{aligned}$$

□

Since the vector $(J_{t+1}^1, J_{t+1}^2, J_{t+1}^3)$ is inside the capacity for all $1 \leq t \leq N$, then, by definition, $\forall \epsilon > 0$ there exist L and $P_{X_1^L X_2^L Y^L}$ factoring as in (6.4) such that

$$J_{t+1}^i \leq I_L^i + \epsilon, \quad i = 1, 2, 3$$

holds for all $1 \leq t \leq N$. This implies the statement of the lemma.

□

E.3 Proof of Lemma 24

Proof. We prove the first statement of the lemma. The second and the third statements follow by a similar argument. Given $Y^t = y^t, W_2 = m$, define the following quantities

$$\begin{aligned}
f_{i|m} &= P(W_1 = i|Y^t = y^t, W_2 = m) \\
f_{i|m}(y_{t+1}) &= P(W_1 = i|Y^t = y^t, W_2 = m, Y_{t+1} = y_{t+1}) \\
Q_{i,m}(y_{t+1}) &= P(Y_{t+1} = y_{t+1}|W_1 = i, W_2 = m, Y^t = y^t),
\end{aligned}$$

where $i \in [1 : M_1], y_{t+1} \in \mathcal{Y}$. Since $H_t^1 < \epsilon$, then there exist ϵ' (as a function of ϵ) and an index $l \in [1 : M_1]$ such that $f_{l|m} \geq 1 - \epsilon'$ and $f_{i|m} \leq \frac{\epsilon'}{M_1 - 1}$ for all $i \in [1 : M_1], i \neq l$.

Denote

$$\hat{f}_{i|m} = \frac{f_{i|m}}{1 - f_{l|m}}, \quad i \neq l.$$

Using the grouping axiom we have

$$H_t^1 = H(W_1|W_2 = m, y^t) = h_b(f_{l|m}) + (1 - f_{l|m})H(\hat{X})$$

where $\hat{X}$ is a random variable with probability distribution $P(\hat{X} = i) = \hat{f}_{i|m},\ i \in [1 : M_1], i \neq l$. Note that

$$h_b(f_{l|m}) \approx -(1 - f_{l|m}) \log(1 - f_{l|m}).$$

Therefore,

$$\begin{aligned} H_t^1 &\approx -(1 - f_{l|m})(\log(1 - f_{l|m}) - H(\hat{X})) \\ &\approx (1 - f_{l|m}) \log(1 - f_{l|m}) \end{aligned} \tag{E.10}$$

where the last approximation is due to the fact that $-\log(1 - f_{l|m}) \gg H(\hat{X})$. Next, we derive an approximation for H_{t+1}^1. Note that

$$f_{l|m}(y_{t+1}) = \frac{f_{l|m} Q_{l,m}(y_{t+1})}{\sum_j f_{j|m} Q_{j,m}(y_{t+1})}$$

The denominator can be written as

$$f_{l|m} Q_{l,m}(y_{t+1}) + (1 - f_{l|m}) \sum_{j \neq l} \hat{f}_{j|m} Q_{j,m}(y_{t+1}).$$

The above quantity is approximately equals to $Q_{l,m}(y)$. Therefore,

$$\begin{aligned}(1 - f_{l|m}(y_{t+1})) &= (1 - f_{l|m})\frac{\sum_{j\neq l}\hat{f}_{j|m}Q_{j,m}(y_{t+1})}{\sum_j f_{j|m}Q_{j,m}(y_{t+1})} \\ &\approx (1 - f_{l|m})\frac{\sum_{j\neq l}\hat{f}_{j|m}Q_{j,m}(y_{t+1})}{Q_{l,m}(y_{t+1})}\end{aligned}$$

This implies that $f_{l|m}(y_{t+1}) \approx 1$. Therefore, using the same argument for H_t^1 we have

$$\begin{aligned}H_{t+1}^1 &\approx -(1 - f_{l|m}(y_{t+1}))(\log(1 - f_{l|m}(y_{t+1})) \\ &= -(1 - f_{l|m}(y_{t+1}))\Big[\log(1 - f_{l|m}) + \log(\frac{\sum_{j\neq l}\hat{f}_{j|m}Q_{j,m}(y_{t+1})}{Q_{l,m}(y_{t+1})})\Big] \\ &\approx -(1 - f_{l|m}(y_{t+1}))\log(1 - f_{l|m}). \qquad \text{(E.11)}\end{aligned}$$

As a result of the approximations in (E.10) and (E.11), we obtain

$$\begin{aligned}\frac{H_{t+1}^1}{H_t^1} &\approx \frac{(1 - f_{l|m}(y))\log(1 - f_{l|m})}{(1 - f_{l|m})\log(1 - f_{l|m})} \\ &= \frac{\sum_{j\neq l}\hat{f}_{j|m}Q_{j,m}(y)}{Q_{l,m}(y)}\end{aligned}$$

Note that

$$P(Y_{t+1} = y|W_2 = m, y^t) \approx Q_{l,m}(y)$$

Therefore,

$$\begin{aligned}
\mathbb{E}\{\log \frac{H_{t+1}^1}{H_t^1}|y^t\} &\approx \mathbb{E}\{\log \frac{\sum_{j\neq l}\hat{f}_{j|m}Q_{j,m}(Y_{t+1})}{Q_{l,m}(Y_{t+1})}\} \\
&= \sum_y Q_{l,m}(y)\log\frac{\sum_{j\neq l}\hat{f}_{j|m}Q_{j,m}(y)}{Q_{l,m}(y)} \\
&\overset{(a)}{=} -D(Q_{l,m}||\sum_{j\neq l}\hat{f}_{j|m}Q_{j,m}) \\
&\overset{(b)}{\geq} -\sum_{j\neq l}\hat{f}_{j|m}D(Q_{l,m}||Q_{j,m}) \\
&\geq -\max_{j\neq l} D(Q_{l,m}||Q_{j,m}) \\
&\overset{(c)}{\geq} -(D_1+\epsilon)
\end{aligned}$$

where (a) is due to the definition of Kullback–Leibler divergence, (b) is due to the convexity of Kullback–Leibler divergence, and (c) is due to the definition of D_1. □

E.4 Proof of Theorem VI.2

Proof. Since $\{Z_t\}$ is a submartingale, then $Z_0 \leq \mathbb{E}[Z_T]$. By the definition of $\{Z_t\}$ we have $\mathbb{E}[Z_T] = \sum_{i=1}^3 \alpha_i \mathbb{E}[Z_T^i]$. For any of processes $\{Z_t^i\}$, the following hold:

$$\begin{aligned}
\mathbb{E}[Z_T^i] &= \mathbb{E}\left[\frac{H_T^i-\epsilon}{I_L^i+\epsilon}1_{\{H_T^i\geq\epsilon\}}\right] + \mathbb{E}\left[\left(\frac{\log H_T^i-\log\epsilon}{D_i+\epsilon}+f_i(\log\frac{H_T^i}{\epsilon})\right)1_{\{H_T^i\leq\epsilon\}}\right] + \mathbb{E}[T] \\
&\leq \mathbb{E}\left[\frac{H_T^i-\epsilon}{I_L^i+\epsilon}\right] + \mathbb{E}\left[\frac{\log H_T^i-\log\epsilon}{D_i+\epsilon}+f_i(\log\frac{H_T^i}{\epsilon})\right] + \mathbb{E}[T] \\
&\overset{(a)}{\leq} \mathbb{E}\left[\frac{H_T^i-\epsilon}{I_L^i+\epsilon}\right] + \mathbb{E}\left[\frac{\log H_T^i-\log\epsilon}{D_i+\epsilon}\right] + \frac{1}{\mu_i D_i} + \mathbb{E}[T] \\
&= \frac{\mathbb{E}[H_T^i]-\epsilon}{I_L^i+\epsilon} + \frac{\mathbb{E}[\log H_T^i]-\log\epsilon}{D_i+\epsilon} + \frac{1}{\mu_i D_i} + \mathbb{E}[T] \\
&\overset{(b)}{\leq} \frac{\mathbb{E}[H_T^i]-\epsilon}{I_L^i+\epsilon} + \frac{\log\mathbb{E}[H_T^i]-\log\epsilon}{D_i+\epsilon} + \frac{1}{\mu_i D_i} + \mathbb{E}[T]
\end{aligned} \tag{E.12}$$

where (a) follows from the inequality $f_i(y) \leq \frac{1}{\mu_i D_i}$, and (b) follows by applying Jensen's inequality for the function $\log(x)$.

Define $\eta(P_e) = h_b(P_e) + P_e \log(M_1 M_2)$. Using Lemma 23, the right-hand side of (E.12) is upper bounded as

$$\begin{aligned}
&\leq \frac{\eta(P_e) - \epsilon}{(I_L^i + \epsilon)} + \frac{\log(\eta(P_e)) - \log \epsilon}{D_i + \epsilon} + \frac{1}{\mu_i D_i} + \mathbb{E}[T] \\
&= \frac{\eta(P_e) - \epsilon}{(I_L^i + \epsilon)} + \frac{\log P_e + \log \frac{\eta(P_e)}{P_e} - \log \epsilon}{D_i + \epsilon} + \frac{1}{\mu_i D_i} + \mathbb{E}[T] \\
&\leq \frac{\log P_e}{D_i + \epsilon} + \mathbb{E}[T](1 + \delta_i(P_e, M_1 M_2, \epsilon)), \qquad \text{(E.13)}
\end{aligned}$$

where the function δ_i is defined as

$$\delta_i(P_e, M_1 M_2, \epsilon) = \| \frac{\eta(P_e) - \epsilon}{(I_L^i + \epsilon)\frac{\log M_1 M_2}{R_N^{(3)}}} + \frac{\log \frac{\eta(P_e)}{P_e} - \log \epsilon}{(D_i + \epsilon)\frac{\log M_1 M_2}{R_N^{(3)}}} + \frac{1}{\mu_i D_i \frac{\log M_1 M_2}{R_N^{(3)}}} \|$$

Note that we use the equation $\mathbb{E}[T] = \frac{\log M_1 M_2}{R_N^{(3)}}$ in the definition of δ_i. Observe that

$$\lim_{P_e \to 0} \lim_{M_1 M_2 \to \infty} \delta_i(P_e, M_1 M_2, \epsilon) = 0.$$

Note that $Z_0^i \leq \mathbb{E}[Z_T^i], i = 1, 2, 3$, where $Z_0^i = \frac{\log M_i - \epsilon}{I_L^i + \epsilon}$. Therefore,

$$\frac{\log M_i - \epsilon}{I_L^i + \epsilon} \leq \frac{\log P_e}{D_i + \epsilon} + \mathbb{E}[T](1 + \delta_i(P_e, M_1 M_2, \epsilon))$$

Multiplying both sides by $\frac{D_i + \epsilon}{\mathbb{E}[T]}$ and rearranging the terms give

$$\begin{aligned}
-\frac{\log P_e}{\mathbb{E}[T]} \leq (D_i + \epsilon) &\left(1 - \frac{R_N^{(i)}}{I_L^i + \epsilon}\right) \\
&+ \frac{\epsilon(D_i + \epsilon)}{(I_L^i + \epsilon)\mathbb{E}[T]} + (D_i + \epsilon)\delta_i(P_e, M_1 M_2, \epsilon),
\end{aligned}$$

Define

$$\tilde{\delta}(P_e, M_1M_2, \epsilon) = \max_i \frac{(D_i+\epsilon)\ \epsilon\ R_N^{(3)}}{(I_L^i+\epsilon')\log M_1M_2} + (D_i+\epsilon)\delta_i(P_e, M_1M_2, \epsilon).$$

For any non-negative numbers $\lambda_i, i = 1, 2, 3$ the following inequality holds:

$$\begin{aligned}
-\frac{\log P_e}{\mathbb{E}[T]} &\leq (D_i+\epsilon)\left(1 - \frac{R_N^{(i)}}{I_L^i+\epsilon}\right) + \tilde{\delta}, \qquad \text{(E.14)}\\
&\leq (D_i+\epsilon)\left(1 - \frac{\lambda_i R_N^{(i)}}{\lambda_i I_L^i+\epsilon}\right) + \tilde{\delta},\\
&\leq (D_i+\epsilon)\left(1 - \frac{\lambda_i R_N^{(i)}}{\sum_j \lambda_j I_L^j+\epsilon'}\right) + \tilde{\delta},\\
&\leq (D_i+\epsilon)\left(1 - \frac{\lambda_i R_N^{(i)}}{\sup \sum_j \lambda_j I_L^j+\epsilon}\right) + \tilde{\delta},\\
&= (D_i+\epsilon)\left(1 - \frac{\lambda_i R_N^{(i)}}{C_\lambda+\epsilon}\right) + \tilde{\delta},
\end{aligned}$$

Since the transmission rates are inside the capacity region, $\lambda_i R_N^{(i)} \leq C_\lambda$ and we obtain

$$\begin{aligned}
\frac{\log P_e}{\mathbb{E}[T]} &\leq D_i\left(1 - \frac{\lambda_i R_N^{(i)}}{C_\lambda+\epsilon}\right) + \epsilon + \tilde{\delta}(P_e, M_1M_2, \epsilon),\\
&\overset{(a)}{=} D_i\left(1 - \frac{\lambda_i R_N^{(i)}}{C_\lambda}\right) + D_i\frac{\lambda_i R_N^{(i)}\epsilon}{C_\lambda(C_\lambda+\epsilon)} + \epsilon + \tilde{\delta}(P_e, M_1M_2, \epsilon),\\
&\overset{(b)}{\leq} D_i\left(1 - \frac{\lambda_i R_N^{(i)}}{C_\lambda}\right) + D_{\max}\frac{\epsilon}{C_\lambda} + \epsilon + \tilde{\delta}(P_e, M_1M_2, \epsilon),
\end{aligned}$$

where $D_{\max} = \max\{D_1, D_2, D_3\}$, (a) follows by adding and subtracting the term $D_i(\frac{\lambda_i R_N^{(i)}}{C_\lambda})$, and (b) follows as $\frac{\lambda_i R_N^{(i)}}{C_\lambda+\epsilon} \leq 1$. Define $\delta(P_e, M_1M_2, \epsilon) = \epsilon(1 + \frac{D_{\max}}{C_\lambda}) + \tilde{\delta}(P_e, M_1M_2, \epsilon)$. The theorem follows by taking the minimum over $\lambda_i, i = 1, 2, 3$ and

the fact that the following condition is satisfied:

$$\lim_{\epsilon\to 0}\lim_{P_e\to 0}\lim_{M_1 M_2\to\infty}\delta(P_e, M_1M_2, \epsilon) = 0.$$

Note that in the above proof it is assumed that the capacity region is nonempty. This assumption implies that $C_\lambda > 0$ for all $\underline{\lambda} \neq \underline{0}$ with non-negative components. □

E.5 Proof of Corollary 3

From (E.14) in the proof of Theorem VI.2, we obtain:

$$\begin{aligned}
-\frac{\log P_e}{\mathbb{E}[T]} &\leq \min_{i\in\{1,2,3\}}(D_i+\epsilon)\left(1-\frac{R_N^{(i)}}{I_L^i+\epsilon}\right)+\tilde{\delta}\\
&\leq D_{\max}\min_{i\in\{1,2,3\}}\left(1-\frac{R_N^{(i)}}{I_L^i}\right)+\delta\\
&= D_{\max}\min_{\substack{\alpha_1,\alpha_2,\alpha_3\geq 0\\ \alpha_1+\alpha_2+\alpha_3=1}}\left(1-\sum_{i=1}^{3}\alpha_i\frac{R_N^{(i)}}{I_L^i}\right)+\delta, \qquad \text{(E.15)}
\end{aligned}$$

where $D_{\max} = \max\{D_1, D_2, D_3\}$, and $\delta = \tilde{\delta} + \epsilon\sup(1+\frac{D_{\max}}{I_L^i})$. For non-negative $\lambda_i, i = 1,2,3$, set $\alpha_i = \frac{\lambda_i I_L^i}{\sum_j \lambda_j I_L^j}$. Next, replace $\alpha_i, i = 1,2,3$ in (E.15) with the above term. Therefore, (E.15) does not exceed the following

$$D_{\max}\min_{\substack{\lambda_1,\lambda_2,\lambda_3\geq 0\\ \lambda_1+\lambda_2+\lambda_3=1}}\left(1-\frac{\sum_i \lambda_i R_N^{(i)}}{\sum_j \lambda_j I_L^j}\right)+\delta.$$

The proof is completed by noting that $\sum_j \lambda_j I_L^j \leq C_\lambda$.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 4, pp. 623–656, Oct 1948.

[2] J. A. T. Thomas M. Cover, *Elements of Information Theory.* Wiley John + Sons, 2006.

[3] I. Csiszar and J. Korner, *Information Theory.* Cambridge University Press, 2011.

[4] R. G. Gallager, *Information Theory and Reliable Communication.* John Wiley & Sons, Inc., 1968.

[5] A. E. Gamal and Y.-H. Kim, *Network Information Theory.* Cambridge University Press, 2011.

[6] R. Ahlswede, "Multi-way communication channels," in *Second International Symposium on Information Theory: Tsahkadsor, Armenia, USSR, Sept. 2-8, 1971*, 1973.

[7] J. Korner and K. Marton, "How to encode the modulo-two sum of binary sources (corresp.)," *IEEE Transactions on Information Theory*, vol. 25, no. 2, pp. 219–221, 1979.

[8] C. Nair, L. Xia, and M. Yazdanpanah, "Sub-optimality of han-kobayashi achievable region for interference channels," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Jun. 2015, pp. 2416–2420.

[9] T. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE transactions on information theory*, vol. 27, no. 1, pp. 49–60, 1981.

[10] D. Krithivasan and S. S. Pradhan, "Distributed source coding using abelian group codes: A new achievable rate-distortion region," *IEEE Transactions on Information Theory*, vol. 57, no. 3, pp. 1495–1519, March 2011.

[11] R. Ahlswede and T. Han, "On source coding with side information via a multiple-access channel and related problems in multi-user information theory," *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 396–412, May 1983.

[12] T. Han and K. Kobayashi, "A unified achievable rate region for a general class of multiterminal source coding systems," *IEEE Transactions on Information Theory*, vol. 26, no. 3, pp. 277–288, May 1980.

[13] ——, "A dichotomy of functions $F(X, Y)$ of correlated sources (X, Y)," *IEEE Transactions on Information Theory*, vol. 33, no. 1, pp. 69–76, January 1987.

[14] B. Nazer and M. Gastpar, "Computation over multiple-access channels," *IEEE Transactions on Information Theory*, vol. 53, no. 10, pp. 3498–3516, Oct. 2007.

[15] S. H. Lim, C. Feng, B. Nazer, and M. Gastpar, "A joint typicality approach to compute-forward," in *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, sep 2015.

[16] M. Heidari, F. Shirani, and S. S. Pradhan, "Beyond group capacity in multiterminal communications," in *2015 IEEE International Symposium on Information Theory (ISIT)*, June 2015, pp. 2081–2085.

[17] M. Heidari and S. S. Pradhan, "How to compute modulo prime-power sums," in *2016 IEEE International Symposium on Information Theory (ISIT)*, July 2016, pp. 1824–1828.

[18] A. Padakandla, "Computing sum of sources over an arbitrary multiple access channel," in *Proc. IEEE Int. Symp. Information Theory*, Jul. 2013, pp. 2144–2148.

[19] J. Zhan, S. Y. Park, M. Gastpar, and A. Sahai, "Linear function computation in networks: Duality and constant gap results," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 4, pp. 620–638, April 2013.

[20] R. Appuswamy, M. Franceschetti, N. Karamchandani, and K. Zeger, "Linear codes, target function classes, and network computing capacity," *IEEE Transactions on Information Theory*, vol. 59, no. 9, pp. 5741–5753, Sept 2013.

[21] T. Philosof, R. Zamir, U. Erez, and A. J. Khisti, "Lattice strategies for the dirty multiple access channel," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 5006–5035, Aug 2011.

[22] T. Philosof and R. Zamir, "On the loss of single-letter characterization: The dirty multiple access channel," *IEEE Transactions on Information Theory*, vol. 55, no. 6, pp. 2442–2454, Jun. 2009.

[23] A. Padakandla and S. S. Pradhan, "Achievable rate region based on coset codes for multiple access channel with states," in *2013 IEEE International Symposium on Information Theory*, July 2013, pp. 2641–2645.

[24] M. Heidari, F. Shirani, and S. S. Pradhan, "A new achievable rate region for multiple-access channel with states," in *IEEE International Symposium on Information Theory (ISIT)*, 2017.

[25] ——, "New sufficient conditions for multiple-access channel with correlated sources," in *2016 IEEE International Symposium on Information Theory (ISIT)*, July 2016, pp. 2019–2023.

[26] F. Shirani, M. Heidari, and S. S. Pradhan, "New lattice codes for multiple-descriptions," in *2015 IEEE International Symposium on Information Theory (ISIT)*, June 2015, pp. 1580–1584.

[27] S. Sridharan, A. Jafarian, S. Vishwanath, S. A. Jafar, and S. Shamai, "A layered lattice coding scheme for a class of three user gaussian interference channels," in *2008 46th Annual Allerton Conference on Communication, Control, and Computing*, Sept 2008, pp. 531–538.

[28] S. N. Hong and G. Caire, "On interference networks over finite fields," *IEEE Transactions on Information Theory*, vol. 60, no. 8, pp. 4902–4921, Aug 2014.

[29] G. Bresler, A. Parekh, and D. N. C. Tse, "The approximate capacity of the many-to-one and one-to-many gaussian interference channels," *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4566–4592, Sept 2010.

[30] U. Niesen and M. A. Maddah-Ali, "Interference alignment: From degrees of freedom to constant-gap capacity approximations," *IEEE Transactions on Information Theory*, vol. 59, no. 8, pp. 4855–4888, Aug 2013.

[31] A. Jafarian and S. Vishwanath, "Achievable rates for k-user gaussian interference channels," *IEEE Transactions on Information Theory*, vol. 58, no. 7, pp. 4367–4380, 2012.

[32] O. Ordentlich, U. Erez, and B. Nazer, "The approximate sum capacity of the symmetric gaussian k-user interference channel," in *2012 IEEE International Symposium on Information Theory Proceedings*, July 2012, pp. 2072–2076.

[33] F. Shirani and S. S. Pradhan, "Trade-off between communication and cooperation in the interference channel," in *2016 IEEE International Symposium on Information Theory (ISIT)*, July 2016, pp. 2214–2218.

[34] A. Padakandla and S. S. Pradhan, "Achievable rate region for three user discrete broadcast channel based on coset codes," in *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*. IEEE, 2013, pp. 1277–1281.

[35] M. Heidari, F. Shirani, and S. S. Pradhan, "On the necessity of structured codes for communications over mac with feedback," in *IEEE International Symposium on Information Theory (ISIT)*, 2017.

[36] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. 19, no. 4, pp. 471–480, Jul. 1973.

[37] A. G. Sahebi and S. S. Pradhan, "Multilevel channel polarization for arbitrary discrete memoryless channels," *IEEE Transactions on Information Theory*, vol. 59, no. 12, pp. 7839–7857, Dec 2013.

[38] C. Shannon, "The zero error capacity of a noisy channel," *IRE Transactions on Information Theory*, vol. 2, no. 3, pp. 8–19, Sep. 1956.

[39] R. L. Dobrushin, "An asymptotic bound for the probability error of information transmission through a channel without memory using the feedback," *Problemy Peredachi Informatsii*, vol. 8, p. 160–161, 1962.

[40] E. A. Haroutunian, "Lower bound for error probability in channels with feedback," *Problemy Peredachi Informatsii*, vol. 13, pp. 36–44, 1977.

[41] M. Horstein, "Sequential transmission using noiseless feedback," *IEEE Transactions on Information Theory*, vol. 9, no. 3, pp. 136–143, Jul. 1963.

[42] M. V. Burnashev, "Data transmission over a discrete channel with feedback. random transmission time," *Problemy peredachi informatsii*, vol. 12, no. 4, pp. 10–30, 1976.

[43] N. Gaarder and J. Wolf, "The capacity region of a multiple-access discrete memoryless channel can increase with feedback (corresp.)," *IEEE Transactions on Information Theory*, vol. 21, no. 1, pp. 100–102, Jan. 1975.

[44] E. van der Meulen, "A survey of multi-way channels in information theory: 1961-1976," *IEEE Transactions on Information Theory*, vol. 23, no. 1, pp. 1–37, Jan. 1977.

[45] J. Schalkwijk, "The binary multiplying channel–a coding scheme that operates beyond Shannon's inner bound region (corresp.)," *IEEE Transactions on Information Theory*, vol. 28, no. 1, pp. 107–110, Jan. 1982.

[46] T. Han, "A general coding scheme for the two-way channel," *IEEE Transactions on Information Theory*, vol. 30, no. 1, pp. 35–44, Jan. 1984.

[47] G. Kramer, "Directed information for channels with feedback," Ph.D. dissertation, Swiss Federal Institute of Technology, Zurich, 1998.

[48] T. Cover, A. E. Gamal, and M. Salehi, "Multiple access channels with arbitrarily correlated sources," *IEEE Transactions on Information Theory*, vol. 26, no. 6, pp. 648–657, Nov. 1980.

[49] D. Slepian and J. K. Wolf, "A coding theorem for multiple access channels with correlated sources," *The Bell System Technical Journal*, vol. 52, no. 7, pp. 1037–1076, Sep. 1973.

[50] P. Gacs and J. Körner, "Common information is far less than mutual information," *Problems of Control and Information Theory*, vol. 2, Jan. 1973.

[51] H. S. Witsenhausen, "On sequences of pairs of dependent random variables," *SIAM Journal on Applied Mathematics*, vol. 28, no. 1, pp. 100–113, jan 1975.

[52] T. Cover and C. Leung, "An achievable rate region for the multiple-access channel with feedback," *IEEE Transactions on Information Theory*, vol. 27, no. 3, pp. 292–298, May 1981.

[53] M. Naghshvar and T. Javidi, "Active sequential hypothesis testing," *The Annals of Statistics*, vol. 41, no. 6, pp. 2703–2738, 2013. [Online]. Available: http://www.jstor.org/stable/23566746

[54] U. Erez and R. Zamir, "Achieving $\frac{1}{2}\log(1+SNR)$ on the AWGN channel with lattice encoding and decoding," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2293–2314, Oct 2004.

[55] U. Erez, S. Litsyn, and R. Zamir, "Lattices which are good for (almost) everything," *IEEE Transactions on Information Theory*, vol. 51, no. 10, pp. 3401–3416, Oct. 2005.

[56] A. R. Calderbank, "The art of signaling: fifty years of coding theory," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2561–2595, Oct. 1998.

[57] A. G. Sahebi and S. S. Pradhan, "Nested lattice codes for arbitrary continuous sources and channels," in *Proc. IEEE Int Symp. Information Theory*, Jul. 2012, pp. 626–630.

[58] D. Slepian, "Group codes for the gaussian channel," *Bell Labs Technical Journal*, vol. 47, no. 4, pp. 575–602, 1968.

[59] H.-A. Loeliger, "Signal sets matched to groups," *IEEE Transactions on Information Theory*, vol. 37, no. 6, pp. 1675–1682, 1991.

[60] E. Şaşoğlu, E. Telatar, and E. Arikan, "Polarization for arbitrary discrete memoryless channels," in *Information Theory Workshop, 2009. ITW 2009. IEEE*. IEEE, 2009, pp. 144–148.

[61] E. Abbe and E. Telatar, "Polar codes for the m-user multiple access channel," *IEEE Transactions on Information Theory*, vol. 58, no. 8, pp. 5437–5448, Aug 2012.

[62] W. Park and A. Barg, "Polar codes for q-ary channels, $q = 2^r$," *IEEE Transactions on Information Theory*, vol. 59, no. 2, pp. 955–969, Feb 2013.

[63] R. Ahlswede, "Group codes do not achieve shannon's channel capacity for general discrete channels," *The Annals of Mathematical Statistics*, pp. 224–240, 1971.

[64] R. Ahlswede and J. Gemma, "Bounds on algebraic code capacities for noisy channels. i," *Information and Control*, vol. 19, no. 2, pp. 124–145, 1971.

[65] ——, "Bounds on algebraic code capacities for noisy channels. ii," *Information and Control*, vol. 19, no. 2, pp. 146–158, 1971.

[66] G. Como and F. Fagnani, "The capacity of finite abelian group codes over symmetric memoryless channels," *IEEE Transactions on Information Theory*, vol. 55, no. 5, pp. 2037–2054, 2009.

[67] A. G. Sahebi and S. S. Pradhan, "Abelian group codes for channel coding and source coding," *IEEE Transactions on Information Theory*, vol. 61, no. 5, pp. 2399–2414, May 2015.

[68] A. Padakandla, A. G. Sahebi, and S. S. Pradhan, "A new achievable rate region for the 3-user discrete memoryless interference channel," in *2012 IEEE International Symposium on Information Theory Proceedings*, July 2012, pp. 2256–2260.

[69] S. Jafar, "Capacity with causal and noncausal side information: A unified view," *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5468–5474, Dec 2006.

[70] H. Liao, "A coding theorem for multiple access communications," *in Proc. Int. Symp. Information Theory, Asilomar, CA*, 1972.

[71] G. Dueck, "A note on the multiple access channel with correlated sources (corresp.)," *IEEE Transactions on Information Theory*, vol. 27, no. 2, pp. 232–235, Mar. 1981.

[72] T. Han and M. Costa, "Broadcast channels with arbitrarily correlated sources," *IEEE Transactions on Information Theory*, vol. 33, no. 5, pp. 641–650, Sep. 1987.

[73] M. Salehi and E. Kurtas, "Interference channels with correlated sources," in *Proc. IEEE Int. Symp. Information Theory*, Jan. 1993, p. 208.

[74] S. Choi and S. S. Pradhan, "A graph-based framework for transmission of correlated sources over broadcast channels," *IEEE Transactions on Information Theory*, vol. 54, no. 7, pp. 2841–2856, Jul. 2008.

[75] F. Willems, "The feedback capacity region of a class of discrete memoryless multiple access channels (corresp.)," *IEEE Transactions on Information Theory*, vol. 28, no. 1, pp. 93–95, Jan. 1982.

[76] L. Ozarow, "The capacity of the white gaussian multiple access channel with feedback," *IEEE Transactions on Information Theory*, vol. 30, no. 4, pp. 623–629, Jul. 1984.

[77] S. I. Bross and A. Lapidoth, "An improved achievable region for the discrete memoryless two-user multiple-access channel with noiseless feedback," *IEEE Transactions on Information Theory*, vol. 51, no. 3, pp. 811–833, Mar. 2005.

[78] R. Venkataramanan and S. S. Pradhan, "A new achievable rate region for the multiple-access channel with noiseless feedback," *IEEE Transactions on Information Theory*, vol. 57, no. 12, pp. 8038–8054, Dec. 2011.

[79] F. Shirani, M. Heidari, and S. S. Pradhan, "Quasi linear codes: Application to point-to-point and multi-terminal source coding," in *2016 IEEE International Symposium on Information Theory (ISIT)*, July 2016, pp. 730–734.

[80] S. Servetto, V. Vaishampayan, and N. Sloane, "Multiple description lattice vector quantization," in *Data Compression Conference*, Mar. 1999, pp. 13–22.

[81] J. Chen, C. Tian, T. Berger, and S. S. Hemami, "Multiple description quantization via gram schmidt orthogonalization," *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5197–5217, Dec. 2006.

[82] Z. Zhang and T. Berger, "New results in binary multiple descriptions," *IEEE Transactions on Information Theory*, vol. 33, no. 4, pp. 502–521, Jul. 1987.

[83] S. S. Pradhan, R. Puri, and K. Ramchandran, "n-channel symmetric multiple descriptions - part i: (n, k) source-channel erasure codes," *IEEE Transactions on Information Theory*, vol. 50, no. 1, pp. 47–61, Jan. 2004.

[84] R. Venkataramani, G. Kramer, and V. K. Goyal, "Multiple description coding with many channels," *IEEE Transactions on Information Theory*, vol. 49, no. 9, pp. 2106–2114, Sep. 2003.

[85] C. Tian and J. Chen, "New coding schemes for the symmetric k-description problem," *IEEE Transactions on Information Theory*, vol. 56, no. 10, pp. 5344–5365, Oct. 2010.

[86] E. Akyol, K. Viswanatha, and K. Rose, "Combinatorial message sharing and random binning for multiple description coding," in *Proc. IEEE Int Symp. Information Theory*, Jul. 2012, pp. 1371–1375.

[87] F. Shirani and S. S. Pradhan, "An achievable rate-distortion region for the multiple descriptions problem," in *Proc. IEEE Int. Symp. Information Theory*, Jun. 2014, pp. 576–580.

[88] H. Yamamoto and K. Itoh, "Asymptotic performance of a modified schalkwijk-barron scheme for channels with noiseless feedback (corresp.)," *IEEE Transactions on Information Theory*, vol. 25, no. 6, pp. 729–733, Nov. 1979.

[89] P. Berlin, B. Nakiboglu, B. Rimoldi, and E. Telatar, "A simple converse of burnashev's reliability function," *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 3074–3080, Jul. 2009.