An algebraic framework for multi-terminal communication

by

Arun Raghuthama Padakandla

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical Engineering: Systems)
in The University of Michigan
2014

Dissertation Committee:

Associate Professor S. Sandeep Pradhan, Chair
Professor Venkat Anantharam, University of California at Berkeley
Associate Professor Achilleas Anastasopoulos
Professor David L. Neuhoff
Professor Roman Vershynin
Associate Professor Aaron B. Wagner, Cornell University

To Anna and Sindhu

Acknowledgements

I am deeply indebted to my advisor Prof. Sandeep Pradhan for making my PhD experience enlightening, enjoyable, and comfortable. I will be unable to express in words how much I owe to Sandeep. I will only describe a few among the many things I learned and experienced. Firstly, he led me into this beautiful area of multi-terminal information theory and taught me the key ideas and their interconnections. This viewpoint could not be obtained through a study of the literature, and proved to be useful. Recognizing these ideas and interconnections help to simplify the large volume of information, and I will do good to obtain such an understanding of any field I enter into. On a related note, he exposed me to the building blocks on which this thesis is built. He is the leading expert in using these building blocks and I am lucky to be tutored in this field by him. Secondly, he persisted on hard problems and never let me go significantly off track. I thank him for this persistence, without which I could not be feeling as proud of my thesis, as I am currently. More importantly, it gave me the opportunity to experience the journey involved in the conceptualization of a new theory. Thirdly, I thank him for his patience in dealing with my inefficiencies and procrastinations. There have been several occasions over the last 5 years when I should, and could have got tasks done within a shorter time frame. He never expressed his displeasure or lost his patience at my failings. Fourthly, he has spent a lot of his time over the last few years working with me on my thesis problems. He has actively thought on them and constantly communicated to me the picture evolving in his mind. This has enabled me make the right connections and come up with neat and elegant solutions. Fifthly, Sandeep has been very warm and friendly in all our interactions. This has played an extremely important role in keeping my emotion and confidence high. The health and well being of a PhD student depends heavily on his/her relationship with his/her advisor. I will be surprised to find anyone who has had a more comfortable, enjoyable, enriching PhD experience than I have had. I owe this to Sandeep.

I have had a great opportunity to interact and learn from my dissertation committee members. I thank all my dissertation committee members for their participation in my defense and proposal presentations. Prof. David Neuhoff has spent his time and efforts towards improving my thesis. I have benefited a lot from his suggestions and interactions. I have interacted with Prof. Achilleas Anastasopoulos throughout my PhD and it has been a wonderful learning experience. I benefited from his courses EECS 554, EECS 650. Prof. Achilleas Anastasopoulos also refined

my teaching skills when I was the GSI for the EECS 501 course he taught. He gave me considerable freedom in choosing the contents I presented and provided very useful mid-course corrections that has made me a better teacher. I also thank him for giving me the opportunity to present a few topics in his EECS 650 course. I thank Prof. Roman Vershynin for his participation in my dissertation committee. I attended Prof. Vershynin's course on functional analysis which enhanced my knowledge in the field of mathematical analysis.

In spite of being from a different university, and not having known me, Prof. Venkat Anantharam was very kind to my request, and participated in my dissertation committee. I thank him for patiently and promptly replying to my several emails. His inputs have enabled me better place my thesis contributions in context. I am privileged to have had Prof. Aaron Wagner on my dissertation committee. He raised interesting questions during these discussions that aided some aspects of my understanding. I thank Prof. Wagner for taking the time to familiarize himself about my findings and providing me useful feedback.

I have taken 11 out of my 12 courses in the mathematics department, where I also earned a masters. I would like to thank Prof. Robert Lazarsfeld, Prof. Sergey Fomin, Prof. Robert L. Griess, Dr. Pavlo Pylyavskyy, Prof. Berit Stensones for their insightful lectures. I would like to particularly thank Dr. Pavlo Pylyavskyy for his wonderful course on Algebraic Combinatorics and particularly his grading style that facilitated a very good understanding.

I have been very fortunate to have developed some amazing friends here at Ann Arbor. I have known Marimuthu Andiappan (Mari) from the very first week of my arrival here. Since then, Mari has always been there whenever I have felt low or wanted somebody to express my emotions. Mari has also helped me on numerous occasions and has been an amazing friend and confidant. I met Deepanshu in the Winter 2010 term and since then we have jelled very well. Deepanshu and I have interacted extensively on several topics including pure math, politics, information theory, biology, evolution, and the list goes on. I thank Deepanshu for the one-on-one sessions where we taught each other several topics including topology, analysis, measure theory, error exponents, dynamic programming etc. Curtis has always given a patient hearing to all my banter and and given me great support at times when I needed it most. I would like to thank him particularly for the following two conversations. Once in 2011, I felt very low when it seemed like I will not be able to go home for a vacation that summer. On an other occasion, in winter 2012, when I was finding it very hard to meet the unreasonable expectations I set for myself with regard to my teaching, I needed some encouragement. Curtis talked me out through these times and I felt a lot better. I will not forget the numerous dinners Curtis and I had together. Shang-Pin, Yi-Chin Wu, Qingsi Wang have kept my spirits high throughout.

My PhD experience would not have been so enjoyable had I not met Yun Zhang, Leeann and their two wonderful kids Ian and Zoe. Yun and Leeann have been extremely warm, affectionate and hospitable. Since I have known them, my Friday night dinners are invariably at their place after a fun filled round of UNO card game with Ian and Zoe. Yun has been very kind to help me out with many logistics here in Ann Arbor. Leeann has truly made me feel like I am their brother. I am unable to express how much Yun's family has meant to me over the last two years.

My uncle Dr. Bhaskar Padakandla has helped me in many ways to get myself setup and my feet moving here

in the US. Whenever I have had a logistic question with regard to any setup in the US, or required a suggestion to best get a task done, I have approached my uncle Bhaskar, and on every such occasion, he has provided me with the simplest and best solution I could have hoped for. I have called him up as soon as I suspect something wrong with my health and his advice has been invaluable and perfect. The cell phone I have carried for the last five years has also been part of his family plan that he extended to me. I thank him for his suggestions and the time he has spent counseling me.

EECS department at UMICH is composed of some of the nicest people around. Becky Turanski, my graduate coordinator, has answered all my inquiries about logistics, procedures, rules very promptly and more importantly directed me through the path of least difficulty. Among the numerous instances she has gone well beyond my expectations, I would like to recollect one. In 2013, I needed a letter verifying my status and intent to secure a Schengen VISA without which I could not secure one. Through just a few email interactions with Becky, I could communicate my situation to her and she wrote up a letter worded exactly how the stringent VISA authorities expected it and I landed with a VISA. That letter Becky wrote made my 2 week vacation in Schengen zone a reality. My UMICH experience would have been a lot less enjoyable without Becky's support.

Shelly (Michelle Feldkamp) has responded to several of my requests for booking conference rooms, filing my expense reports etc., very promptly. Even with just three or four conference rooms at her disposal, it is amazing that she has always found one for me for exactly the period I wanted, and always kept me informed of the rules and regulations that I might accidentally straddle. Elizabeth Zaenger, Dr. Don Winsor, Joel VanLaven, Linda Randolph and Kyle Banas at EECS DCO have been very helpful with my computing needs. Not having administrative privileges with `kaveri`, I have requested them every time I have had some issue with the system, and they have been exceptional in getting `kaveri` up and running flawlessly. They have also very graciously lent me laptops for extended periods of time which has been invaluable to me. Ann Pace has met many of my travel related questions and has aided me in completing my travel reports.

My office mates Aria Sahebi, Deepanshu Vasal, Curtis Jin, Raj Tejas, Farhad Shirani, David Hong have been very good friends and tolerated all my idiosyncrasies without ever raising an eyebrow. I have constantly interacted with Aria about my thesis problems and greatly benefited through these interactions. Aria and I have worked closely on the findings presented in chapter 4. He worked on employing codes built over groups while I focused on codes built over finite fields. These and other interactions have been very fruitful. I have also been able to simplify certain proofs here by employing techniques that Aria and Sandeep developed in the context of codes built over groups. Raj has helped me solve many of my tex and MATLAB® questions. I am indebted to him for teaching me the way to generate MATLAB® plots in eps and pdf formats through export_fig. He spent quite a good deal of time with me to get my figures 6.2, 6.3, 6.4, 6.5 up and I very thankful to him for this.

My roommates Phani Motamarri, Alok Talekar, Ajay Gopalakrishnan, Neeraj Gupta, Kishan Somayaji have made it easy for me to share my living space. I have stayed with Phani for more than 3 years and it has been a nice

experience. Phani and I have known each other since our IISc days and have shared some of our evolving thoughts and perceptions. I have lived with Alok for 15 months or more and he has been kind and gracious to my needs. He has always been very welcoming.

Vinodh Kanakadass, my friend from IISc Bangalore, has been a very good counselor to me. Being a little senior to me, he had been through the phase a little before me and was therefore able to provide very good advise. Even beyond my PhD questions, I have benefited from his advise on other matters such as job search, interview cracking etc. I have found my interactions with my seniors Ramji Venkataramanan, Dinesh Krithivasan and Ali Nazari very useful. First and foremost, I must thank Ramji and Dinesh for advising me to join Sandeep's research group for my PhD. I got the idea of the second moment method that I have used extensively in my thesis while talking to Ali very early in my PhD. I thank him for those interactions.

I have reserved the last to express my gratitude to the people that have meant the most to me through this period and my entire life. Anna and Sindhu have meant the world to me and I will never me able to convey in words how much I owe them. After the untimely demise of Amma back in 2000, Anna has single-handedly brought me up and instilled in me the principles of hard work, honesty, sincerity, patience, commitment etc. These principles have made all the difference in my life. Anna is the person I love the most. Anna is the person I have the most regard and respect for. I owe every success, everything I possess today to Anna. Anna has always encouraged me to express my opinions and I have had several discussions during the last 10 to 15 years, some of which we have slightly differed. Through all these and much more, the one constant has been my deep sense of regard for him, his accomplishments and his contribution to my upbringing. Through these years, he has been constantly talking to me, lifting my spirits, encouraging me, and teaching me principles that I have adhered to. This has made all the difference. In spite of my mother not being there, and Sindhu being at IISc for long periods, Anna has always let me pursue my dreams even if that meant I have to stay away from him. He has never let his desires or needs come in the way of my ambitions and has always encouraged and directed me towards my dreams. Anna has given me everything - love, affection, encouragement - unconditionally and I will only pray to live the rest of my years closely adhering to the principles he has taught me.

My sister Sindhu's constant love and encouragement have been the reasons for my success. Sindhu has shared with me everything she possessed and has always taken care of my needs and met my demands. She has done all of this with loads of love and affection. She has been constantly speaking to me and encouraging me through the long and frustrating periods of my PhD. Every time I have gone home for a vacation, she has taken charge of me, set me right, and stitched me back into one piece again. She has done all of this with so much joy and love and expected nothing back in return. Sindhu has taken care of my father during these years so wonderfully well. I am deeply indebted to her for this. This enabled me not worry about home and concentrate on my work. I am so lucky to have such a wonderful and affectionate sister. Sindhu is the dearest girl whom I owe all my accomplishments to. She means the world to me.

Contents

List of Tables

List of Figures

List of Appendices

Notation

We employ notation that is now widely employed in information theory literature supplemented by the following.

- $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ denote the set of natural numbers, integers, rational numbers and real numbers respectively.
- Calligraphic letters such as $\mathcal{X}$, $\mathcal{Y}$ are employed exclusively to denote finite sets. $\mathcal{F}_q$ denotes the finite field of cardinality q. For any set $A \subseteq \mathbb{R}^k$, $\mathrm{cl}\,(A), \mathrm{cocl}\,(A)$ denote closure of A and closure of the convex hull of A respectively. If A is a finite set, $|A|$ denotes cardinality of A.
- For positive integers $i \leq j$, $[i:j] := \{i, i+1, \cdots, j\}$. We let $[j] := [1:j]$.
- We denote (i) random variables using upper case letters, (ii) specific realization of random variables and elements of a set using lower case letters. For example, U is a random variable taking values in $\mathcal{U}$ and $u \in \mathcal{U}$ represents a realization of U. Vectors are distinguished from scalars using a superscript that indicates the length of the vector. For example, U^n is an $n-$length random vector taking values in $\mathcal{U}^n := \underbrace{\mathcal{U} \times \cdots \times \mathcal{U}}_{n \text{ times}}$ and $u^n \in \mathcal{U}^n$ denotes a realization of U^n.
- For $\alpha, \beta \in [0,1]$, $\alpha * \beta := (1-\alpha)\beta + \alpha(1-\beta)$ denotes binary convolution.
- While $+$ denotes addition in $\mathbb{R}$, we let $\oplus$ denote addition in a finite field. The particular finite field, which is uniquely determined (up to an isomorphism) by it's cardinality, is clear from context. When ambiguous, or to enhance clarity, we specify addition in $\mathcal{F}_q$ using $\oplus_q$. For $a, b \in \mathcal{F}_q$, $a \ominus b := a \oplus (-b)$, where $(-b)$ is the additive inverse of b.
- If $f : \mathcal{U} \to \mathcal{X}$ is a map, the n-letter extension of f denoted $f^n : \mathcal{U}^n \to \mathcal{X}^n$ is defined $f^n(u^n) := (f(u_i) : i \in [n])$.
- We employ standard notation for probability mass functions (pmf). For example, if p_{UXSY} is a pm on $\mathcal{U} \times \mathcal{X} \times \mathcal{S} \times \mathcal{Y}$, then p_{UY} is the corresponding marginal on $\mathcal{U} \times \mathcal{Y}$. p_{UY}^n is the pm on $\mathcal{U}^n \times \mathcal{Y}^n$ obtained as an $n-$fold product of p_{UY} i.e., $p_{UY}^n(u^n, y^n) = \prod_{i=1}^n p_{UY}(u_i, y_i)$. $p_{Y|U}(y|u)$, defined whenever $p_U(u) \neq 0$, is conditional probability of observing $y \in \mathcal{Y}$ given $u \in \mathcal{U}$ is observed. We write $U \sim p_U$ if p_U is the pmf of U.

- The log and exp functions are taken with respect to the same base. For concreteness, the base may be assumed to be 2, in which case units for information theoretic quantities such as entropy and mutual information would be bits/symbol.
- Let $h_b : [0,1] \to [0,1]$ defined as $h_b(x) := -x \log x - (1-x)\log(1-x)$ denote binary entropy function.
- We employ standard notation for information theoretic quantities such as entropy and mutual information. For example, $H(UY) := -\sum_{(u,y)\in\mathcal{U}\times\mathcal{Y}} p_{UY}(u,y)\log p_{UY}(u,y)$ denotes entropy of p_{UY}, $H(U|Y) := H(U,Y) - H(Y), I(U;Y) := H(U) - H(U|Y)$ and $I(U;Y|S) := I(U;YS) - I(U;S)$.
- The probability of an event A is denoted $P(A)$, and whenever B is an event with non-zero probability, $P(A|B)$ denotes conditional probability of event A given event B.
- We write $U - (X,S) - Y$ if $U, (X,S)$ and Y forms a Markov chain, i.e., U and Y are conditionally independent given (X,S).
- For any $r \in \mathbb{R}$, $\lceil r \rceil := \min\{k \in \mathbb{Z} : k \geq r\}$ and $\lfloor r \rfloor := \max\{k \in \mathbb{Z} : k \leq r\}$.
- For $a \in \mathbb{N}$, $\pi(a) := \min\{k \in \mathbb{N} : k \geq a, k \text{ is a prime power}\}$.
- For a pmf p_{UXSY} defined on $\mathcal{U}\times\mathcal{X}\times\mathcal{S}\times\mathcal{Y}$, let

$$\mathcal{R}(p_{UXSY}, U) := \{u \in \mathcal{U} : \text{ there exists } (x,s,y) \in \mathcal{X}\times\mathcal{S}\times\mathcal{Y} : p_{UXSY}(u,x,s,y) > 0\}$$

  denote essential range of U. When clear from context, we omit the underlying pmf and let $\mathcal{R}(U)$ denote $\mathcal{R}(p_{UXSY}, U)$.

Abstract

We consider the problem of developing coding techniques and characterizing information-theoretic achievable rate regions for the following three multi-terminal communication channels. Firstly, we study an interference channel with three transmitter receiver pairs (3-IC). Secondly, we consider a broadcast channel with three receivers (3-BC), wherein three independent information streams are to be communicated to the three receivers. Thirdly, we consider a two user multiple access channel (MAC) with channel state information distributed at the transmitters (MAC-DSTx). The above channels are assumed discrete, memoryless and used without feedback.

The current known coding technique for a general instance of these channels are based on independent unstructured codes. Recognizing the need for codes endowed with algebraic closure properties, we identify three ensembles of coset codes. We propose coding techniques based on these ensembles that exploit their algebraic closure property. We develop tools to characterize the information-theoretic performance of the proposed coding techniques. These enable us derive achievable rate regions for a general instance of the above channels. The current known achievable rate regions can be enlarged by gluing together current known coding techniques and the ones proposed herein. Moreover, such an enlargement, as indicated below, is proven to be strict for certain instances.

We identify additive and non-additive instances of 3-IC for which the derived achievable rate region is analytically proven to be strictly larger than current known largest. Moreover, for these channels, the proposed coding techniques based on coset codes are optimal, i.e., capacity achieving. We also identify a vector additive 3-BC for which the achievable rate region derived herein is analytically proven to be strictly larger than the current known largest. This vector additive 3-BC is the first known broadcast channel, for which superposition and binning of unstructured independent codes, proposed over three decades ago, can be strictly improved upon. We also identify non-additive and non-symmetric instances of MAC-DSTx for which the proposed coding technique is verified, through computation, to yield strictly larger achievable rate regions.

Finally, we develop a coding technique based on nested coset codes to characterize a weaker set of sufficient conditions for the problem of computing sum of sources over a discrete memoryless MAC.

Chapter 1

Introduction

In his magnum opus [1], Shannon developed an elegant mathematical theory to model the problem of communication. He formalized the notion of reliable communication over a noisy channel and precisely quantified the object of interest - *capacity region* - as the set of rates at which information can be reliably communicated from a transmitter (Tx) to a receiver (Rx). For the particular scenario of communicating an information source from a single Tx to a single Rx, henceforth referred to as a PTP, Shannon provided a comprehensive solution, i.e., a *single-letter* characterization[1] of the capacity region.

Following the publication of [1], it was recognized, that the characterization of the capacity region of a communication system was fundamental to our understanding of it's performance limits. This led to the information theoretic study[2] of multi-terminal systems. In spite of some comprehensive solutions ([3], [4], [5], [6], [7] among others) and ingenious techniques such as [5], [8], [9], among several others, the problem of characterizing the capacity region of several multi-terminal systems, such as interference and broadcast channels, remain open. In this thesis, we address this problem of four multi-terminal systems which are described in the following. Throughout, we assume the multi-terminal systems are *discrete*, *memoryless* and *used without feedback*.[3]

(i) **Three user interference channel (3−IC):** Consider an interference channel (IC) with three transmitter-receiver (Tx-Rx) pairs as depicted in figure 1.1. The symbol input on the channel by each Tx influences the symbols observed by every Rx and this is modeled through the joint channel transition probabilities $W_{Y_1Y_2Y_3|X_1X_2X_3}$. Each Tx wishes to communicate a specific information stream[4] to it's corresponding Rx.

[1] In simple terms, a characterization of a set is said to be single-letter if it is obtained through the result of an optimization over a finite number of parameters. For a detailed description, please refer to [2, Chapter 13].

[2] The mathematical theory proposed by Shannon which seeks, among others, a characterization of capacity region of communication systems is referred to as information theory.

[3] These assumptions are well established in information theory and the reader is referred to [10, Section 4.1] for a lucid description of the same in the context of a PTP. These assumptions in the context of the four multi-terminal systems will be precisely stated in the corresponding chapters.

[4] The information streams being specific to corresponding Tx-Rx pairs, are assumed to be mutually statistically independent.

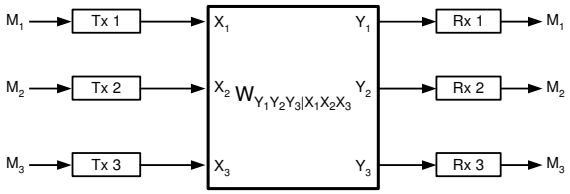

Figure 1.1: Three user interference channel (3−IC)

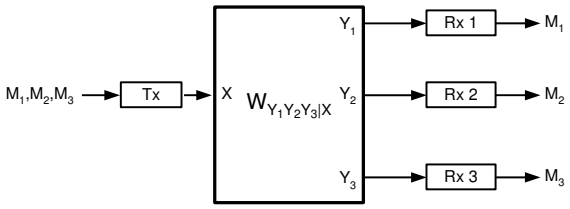

Figure 1.2: Three user broadcast channel (3−BC)

(ii) **Three user broadcast channel (3−BC):** As depicted in figure 1.2, a 3−BC consists of a single Tx and three Rxs. Each Rx demands a specific information stream and it is assumed that the three information streams are mutually statistically independent. The objective is therefore to multiplex three information streams through a single input terminal.

(iii) **Multiple access channel with distributed states (MAC-DSTx):** Consider a two user multiple access channel (MAC) depicted in figure 1.3. The channel transition probabilities $W_{Y|X_1S_1X_2S_2}$ of the MAC depend on a random parameter $\mathbf{S} = (S_1, S_2)$ called *state*. The evolution of the state is independent and identically distributed across time. Tx j is provided with the entire realization of component S_j even before communication begins and the Rx is oblivious to the evolution of the state. As in a MAC, the Txs wish to communicate a pair of independent messages to the Rx.

(iv) **Computation of sum of sources over an arbitrary multiple access channel (MAC):** Consider a MAC with two Txs as depicted in figure 1.4. Each encoder observes one component of a pair of sources that take values over a common finite field. The Rx is interested in reconstructing the sum of sources. The problem of interest is to characterize the maximum number of digits of the sum that can reliably be reconstructed at the Rx per channel use.

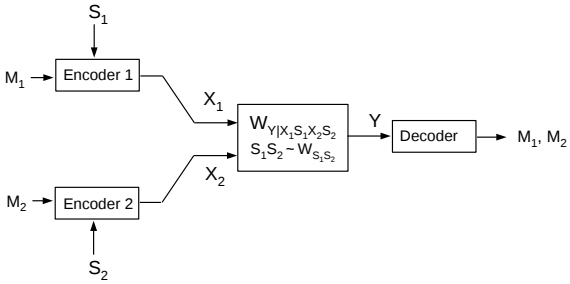

Figure 1.3: MAC with distributed states

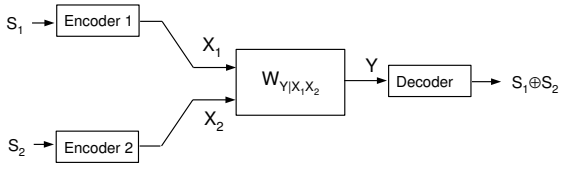

Figure 1.4: Computation sum of sources over a MAC

In this thesis, we focus attention on characterizing inner bounds to the capacity region, i.e., *achievable rate regions*, for the above multi-terminal systems. Our main contribution is a characterization of new achievable rate regions that are strictly larger than current known largest. In the following two sections, we lead the reader to the motivating principles that have shaped this thesis.

1.1 The technique of random coding

The most common technique of proving achievability of rate regions in information theory is random coding[5]. As against to identifying a particular code, a random code of a particular rate R is defined for each block length n. The reliability of this random code, i.e, the probability of incorrect decoding, is evaluated, as a function of block length $n \in \mathbb{N}$. The rate R is characterized as achievable, if this function decays to zero with increasing block length.

Conventionally, the letters of the random code are independently and identically distributed (iid) according to a particular single-letter distribution. What is the effect of choosing the distribution of the random code to be iid in characterizing an achievable rate region? Since we are interested in the limiting performance of the random code, as a function of the block length n, we may employ ideas from typicality to answer this question.[6] For large block lengths n, an iid distribution places an exponentially larger weight on a particular sub-collection of codebooks whose codewords have an empirical distribution close to the single-letter distribution, and moreover, every such codebook is weighed almost equally. In fact, it can be shown that average probability of incorrect decoding of the above sub-collection of codebooks,[7] is equal, at least in the exponent, to the probability of incorrect decoding of the random codebook. Therefore, choosing the distribution of the random code to be iid according to a particular single-letter distribution has the effect of characterizing the reliability of a typical codebook whose codewords have an empirical distribution close to the single-letter distribution.

Let us now consider the case of a multi-terminal system. Since communication over multi-terminal systems employs a multi-terminal code consisting of several constituent codes, defining a random multi-terminal code requires one to specify joint distribution of the constituent codes. Conventionally, the constituent random codes are chosen to be independent.[8] Moreover, as before, the letters of each constituent random code are iid according to a particular single-letter distribution. Let us study the effect of choosing this distribution for the random multi-terminal code in characterizing an achievable rate region. The rate regions proved achievable for a multi-terminal system via this approach is essentially that achievable using a typical multi-terminal code wherein the codewords of each constituent code has an empirical distribution close to a particular single-letter distribution. We emphasize that the constituent

[5]The other known techniques are based on Feinstein's lemma [11] and graph decomposition [12].

[6]The subject of chapter 2 is typicality and covers all the material required to provide the kind of answers we are seeking. A reader not familiar with typicality is encouraged to read through the following without getting bogged down by the technicalities.

[7]averaged with respect to a uniform distribution on this sub-collection

[8]The informed reader might point to the technique of superposition, wherein the satellite and cloud center codebooks are not independent. However, we point out that the same rate region can be proved achievable using independent codes as done in [13].

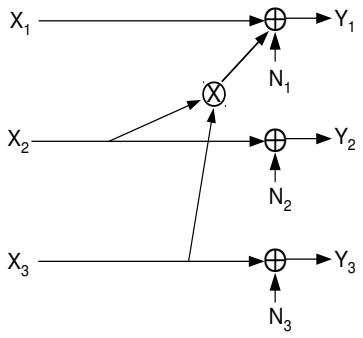

Figure 1.5: A binary 3−IC.

codes of such a typical multi-terminal code do not possess any joint relationship. In particular, (i) individually, the constituent codes do not possess any properties other than the empirical properties mentioned above and (ii) jointly do not possess any particular relationship. We will henceforth refer to such a collection of constituent codes that make a multi-terminal code as *independent unstructured* codes. In the sequel, we illustrate through a simple example, how constituent codebooks (i) possessing additional properties, and (ii) jointly related facilitate efficient communication over multi-terminal systems.

1.2 Interaction of codes in a multi-terminal system

Consider a three user binary interference channel as depicted in figure 1.5.[9] Three distributed Txs wish to communicate specific information to corresponding Rxs over a shared communication medium. Each Tx can input symbols in $\{0,1\}$. If X_j denotes symbol input by Tx j and Y_j denotes the symbol observed by Rx j, we have $Y_1 = X_1 \oplus (X_2 \otimes X_3) \oplus N_1, Y_2 = X_2 \oplus N_2$ and $Y_3 = X_3 \oplus N_3$, where (i) N_1, N_2 and N_3 are independent Bernoulli processes with $P(N_1 = 1) = \delta_1$ and $P(N_j = 1) = \delta$ for $j = 2, 3$, and (ii) $\otimes$ is any particular binary operation such as binary addition $\oplus$ or logical OR $\vee$. Observe that users 2 and 3 enjoy interference free PTPs and can therefore communicate at their respective capacities simultaneously. If C_2 and C_3 denote codebooks employed by users 2 and 3 respectively, then note that user 1 has to deal with the interference patterns in $C_2 \otimes C_3 := \{x_2^n \otimes x_3^n : x_j^n \in C_j : j = 2, 3\}$. Clearly, smaller the cardinality of $C_2 \otimes C_3$, larger the rate at which user 1 can communicate. If C_2 and C_3 were arbitrary capacity achieving codes possessing no joint relationship, $|C_2 \otimes C_3|$ could be large, thereby severely limiting user 1's rate. On the contrary, C_2 and C_3 could be carefully chosen capacity achieving codes such that cardinality of $C_2 \otimes C_3$ is limited, thus facilitating higher rate of communication for user 1. For example, suppose $\otimes$ is binary addition $\oplus$, then users 2 and 3 can achieve capacity by employing cosets of a common linear code.[10] Thereby, $C_2 \oplus C_3$ is

[9] As mentioned earlier, we assume the channel is memoryless and used without feedback

[10] There exists cosets of a linear code that achieve capacity of a binary symmetric channel. This has been proved in [14, Section 6.2].

another coset of the same linear code, and therefore of the same rate. In contrast, if users 2 and 3 employed arbitrary codes of rate R, then $C_2 \oplus C_3$ could potentially be of rate $\min\{1, 2R\}$!!!

We highlight the key elements observed in the above example. Firstly, codes of users 2 and 3 interact through the binary operation $\otimes$. Secondly, in order to enable higher rates for user 1, their codes must be jointly designed to keep the number of interference patterns low. Thirdly, in addition to be being jointly designed, the codes must individually possess algebraic closure properties.[11] These elements, observed in the context of a particular multi-terminal system, being the ***motivating principles for the theory developed in this thesis***, are restated for the sake of emphasis. ***Constituent codes when employed in a multi-terminal system interact. Constituent codes (i) individually possessing certain algebraic closure properties, and (ii) jointly designed enable favorable interaction, and thereby efficient communication.***

1.3 What do we seek and what do we accomplish?

In contrast to independent unstructured codes, a multi-terminal code whose constituent codes possessing the above properties are henceforth referred to as *structured codes*. In this thesis, our aim is to leverage illustrative examples, such as the one above, to develop coding techniques based on structured codes for *arbitrary instances* of the four multi-terminal systems (figures 1.3-1.4) that can exploit properties of the code to enable efficient communication. Thus far, linear and nested linear codes have been employed to develop coding techniques for particular additive and symmetric instances of certain multi-terminal systems, as in [15], [16], [17] etc, that exploit the (algebraic closure) property of the code to yield strictly larger achievable rate regions than that achievable using coding techniques based on independent unstructured codes. However, these coding techniques do not generalize to arbitrary instances of the multi-terminal systems studied therein. This raises, among others, the following three questions.

Firstly, are linear and nested linear codes applicable only for additive and symmetric multi-terminal systems, and if not, how does one go about developing coding techniques based on these examples that is applicable in a wider context? Secondly, if such coding techniques exist, would they provide strict improvement in achievable rate regions even for non-additive scenarios? Thirdly, for multi-terminal systems such as broadcast channel (BC), we are unaware of any example, including additive instances, for which structured codes yield strictly larger achievable rate regions than that based on independent unstructured codes. Can we develop new coding techniques based on structured codes for such multi-terminal systems, in particular the BC, and derive strictly larger achievable rate regions than current known largest, and thereby enable us inch closer to solutions for these long standing open problems?[12]

[11] For example, when the binary operation $\otimes$ is addition $\oplus$, individually codes of users 2 and 3 must be cosets, which are algebraically closed. The notion of algebraically closed will be explained in due course.

[12] An achievable rate region was derived by Marton [9] in the context of a discrete two user BC three decades ago. The current known largest achievable rate region for any BC, including the larger class of BC's with any number of Rxs and arbitrary alphabet sets, is obtained by 'appropriate stitching together' the coding techniques proposed in [9]. We are unaware whether this is the capacity region even for the discrete two user BC.

The central theme of this thesis is ***codes endowed with algebraic closure properties, such as linear and nested linear codes, enable efficient communication over multi-terminals in general, not just particular additive and symmetric instances***. This motivates us to develop a framework based on codes possessing algebraic closure properties for communication over arbitrary instances of the four multi-terminal systems. We illustrate the central theme and prove the utility of this framework by identifying non-additive and non-symmetric instances for which the proposed framework yields strictly more efficient communication over current known techniques. We develop a new coding technique based on codes possessing algebraic closure property for communicating over a 3−DBC that enables us derive a strictly larger achievable rate region than current known largest. A significant element of our findings is an identification of the first BC for which precoding and superposition coding using independent unstructured codes can be strictly improved upon. It maybe noted that even within the wider class of discrete or continuous valued BCs with any number of Rxs we have been unaware of any such example since 1980.

1.4 Characterizing achievable rate regions using coset codes

The motivating principles recognized in section 1.2 lead us to step beyond independent unstructured codes. In this thesis, we study the use of *coset codes* built over finite fields in characterizing new achievable rate regions for the four multi-terminal systems depicted in figures 1.3 - 1.4. Coset codes over finite fields are simply cosets of a linear code. Coset codes are *algebraically closed*[13]. Any two cosets[14] of a linear code when added, result in another coset of the same linear code. As against to adding two arbitrary collections of codewords (over the finite field), the addition of two cosets of a common linear code, results in a collection of codewords[15] of the same size.[16] This property of coset codes motivates their choice and will play a central role throughout this thesis. We exploit this property of coset codes by proposing new coding techniques.[17] It is the analysis of these proposed coding techniques coupled with the use of coset codes that yield new achievable rate regions for the four multi-terminal communication systems studied herein. In the following, we describe the key challenges in characterizing achievable rate regions using coset codes. The theory and techniques developed to overcome these challenges are some of the key contributions of this thesis.

Quite naturally, we employ the technique of random coding to analyze the performance of proposed coding

[13]Consider a linear code over a finite field. The sum of any two codewords is another codeword in the same linear code. This property of the linear code is usually referred to as algebraic closure. In this thesis, we employ a slightly generalized version of this property. Note that any two codewords in a particular coset of a linear code, when added, result in a codeword in another coset. Here, we refer to this property as algebraic closure.

[14]We will use the words coset and coset code interchangeably. In this context, coset is preferred to a coset code since we wish to address coset shifts of the same linear code.

[15]Indeed, this collection is another coset of the same linear code.

[16]Consider the scenario depicted in figure 1.5 with the binary operation $\otimes$ being the binary addition $\oplus$. In this case, codes of users 2 and 3 interact through binary addition $\oplus$. Since the sum of user 2 and 3's codebooks is the collection of interference patterns Rx 1 has to put up with, we favor a reduction in the size of the interference patterns. Choosing user 2 and 3 codes to be cosets of the same linear code accomplishes this.

[17]For the 3−IC depicted in figure 1.5, having employed coset codes to restrict the number of interference patterns, Rx 1 can potentially decode the same. This suggests that we enhance the current decoding technique to exploit the property of coset code and decode the sum interference pattern.

techniques based on coset codes. Since a coset code is completely characterized by a generator matrix and a vector that specifies the coset shift, a random coset code can be defined by specifying the distribution of the generator matrix and the vector. Let the generator matrix and the vector be independent and uniformly distributed. This defines a random coset code. The first challenge lies in characterizing the information theoretic performance of the proposed coding technique using this random coset code over an *arbitrary* instance of the multi-terminal system studied herein.[18]

In this thesis, we develop a mathematical framework based on joint typicality encoding and decoding to analyze the performance of random coset codes. This framework enables us characterize achievable rate regions for *arbitrary* instances of the four multi-terminal systems studied herein. Developing this framework has involved several new elements. Note that codewords in the above defined random coset code are statistically correlated. Moreover, our coding techniques rely on employing jointly correlated random coset codes.[20] An informed reader will note that the analysis of joint typicality encoding and decoding of statistically correlated coset codes will involve several new proof elements. The reader is encouraged to peruse the proofs which are detailed in the appendices.

It can be proved that the codewords of a random coset code, as defined earlier, are uniformly distributed. In contrast to the conventional technique[21], wherein the codewords, of the constituent code, can be chosen to possess any empirical distribution,[22] the codewords of the above random coset code possess only the uniform empirical distribution. A random coset code will therefore enable us achieve rates corresponding to a uniform distribution. How do we achieve rates corresponding to non-uniform distributions?[23] Since constituent codes employed over an arbitrary multi-terminal system must achieve rates corresponding to non-uniform distributions, the second challenge is therefore to find a technique that enables us induce the same using coset codes.

We overcome the second challenge via the technique of *binning* which is best illustrated in the context of a PTP. Consider a finite field input alphabet $\mathcal{X}$ and suppose the capacity achieving distribution p_X is non-uniform. Consider a random coset code of block length n and rate $\frac{k}{n}$ whose generator matrix and coset shifts are uniform and independently distributed. Since we seek codewords of this random coset code whose empirical distribution is close to p_X, we ask the following question. What is the expected number of codewords of this random coset code whose empirical distribution is close to p_X? A reader familiar with the notions of typicality will be able to ascertain this to be close to $|\mathcal{X}|^{k+nH(p_X)-n} = |\mathcal{X}|^{n[\frac{k}{n}-(1-H(p_X))]}$, where the entropy $H(p_X)$ of the distribution p_X is evaluated with

[18]Currently, random coset codes have been employed to derive achievable rate regions only for additive and symmetric instances of multi-terminal communication systems ([18], [15], [16], [19][19]). These works rely on analyzing syndrome decoding which does not generalize for an arbitrary problem instance.

[20]For example, we equip users 2 and 3 of 3−IC depicted in figure 1.5 are equipped with cosets of the same linear code.

[21]By conventional technique, we mean random independent unstructured codes wherein codewords of each constituent code is picked letter by letter iid with respect to a particular single-letter distribution.

[22]This is done by choosing the appropriate single letter distribution of the random iid codebook.

[23]The import of this question can be understood by studying the earlier case of a PTP with a finite field input alphabet. Employing a random coset code, we can prove achievability of mutual information corresponding to a uniform input distribution. This would be strictly sub-optimal if the capacity achieving distribution were non-uniform.

respect to base $|\mathcal{X}|$. One can prove[24] that for large n the probability of the actual number deviating significantly from this expected value is very small. In other words, with high probability, a random coset code of rate $\frac{k}{n} > 1 - H(p_X)$ contains $|\mathcal{X}|^{n[\frac{k}{n}-(1-H(p_X))]}$ codewords whose empirical distribution is close to p_X. By using only these codewords over the channel, one can induce a distribution p_X on the channel. In the sequel, we provide an alternate view of this technique which motivates the term binning.

Suppose we partition the random coset code of rate $\frac{k}{n} > 1 - H(p_X)$ by throwing each codeword uniformly and independently into $|\mathcal{X}|^{n[\frac{k}{n}-(1-H(p_X))]}$ bins. It can be proved that a uniformly chosen bin with high probability contains (i) $|\mathcal{X}|^{n(1-H(p_X))}$ codewords and moreover (ii) at least one codeword whose empirical distribution is close to p_X. If we were to use the message[25] to index a bin, then with high probability, we can choose a codeword within this bin whose empirical distribution is close to p_X.[26] The informed reader will recognize that this is akin to the technique of binning proposed by Gel'fand and Pinsker [7].

Coset codes, joint typical encoding and decoding, and the technique of binning are the building blocks for the theory developed in this thesis. Via binning, we are able to induce non-uniform distribution over the input and auxiliary input alphabets. Joint typical encoding and decoding will enable us analyze performance over arbitrary instances of the multi-terminal systems studied herein. Coset codes will enable us shrink the range of the sum when applied on codebooks. These building blocks will enable us characterize new achievable rate regions. Before we describe our contributions in particular to each of the four multi-terminal systems, let us formally state the modeling assumptions applicable throughout this thesis.

1.5 Modeling assumptions

Throughout, we are concerned with communication channels and information sources that evolve over discrete time. The sources and channels are assumed to be discrete, i.e., the sources take values over finite sets and the channels provide finite input and output alphabet sets. Sources are assumed memoryless, i.e., their distribution across time is assumed to be independent and identical. We assume the channels are (i) memoryless, i.e., conditioned on the input at time n, the output at time n is independent of past inputs, past outputs, (ii) time-invariant i.e., the channel transition probabilities do not vary with time, and (iii) used without feedback, i.e., the inputs have no information of the symbols received at the output. Please refer to the specific chapters for a precise statement of these assumptions

[24] This can be established using the second moment method that employs Cheybyshev inequality. The pairwise independence of codewords aids evaluating the second moment.

[25] Recall that we wish to communicate over a PTP and need to assign codewords to messages.

[26] An informed reader might go the next step and question whether this will enable us achieve capacity over this PTP. Indeed, to achieve capacity using this technique, we need $\frac{k}{n} - (1 - H(p_X)) = I(p_X; W_{Y|X})$, where $W_{Y|X}$ denotes the channel transition probabilities and $I(p_X; W_{Y|X})$ is the mutual information of the joint distribution $p_X W_{Y|X}$. This implies the rate of the complete code $\frac{k}{n} = I(p_X; W_{Y|X}) + 1 - H(p_X)$ is in general larger than the mutual information $I(p_X; W_{Y|X})$. Owing to the sparsity of codewords in the code, whose empirical distribution is close to p_X, we can achieve capacity. This is proven in chapter 3 which forms an important element of the theory developed in this thesis.

in the particular context.

1.6 Contributions of this thesis

In the following, we briefly list our contributions particular to the four multi-terminal systems.

1.6.1 Three user interference channel (3−IC)

(i) Recognizing that interference over a 3−IC is, in general, a bivariate function of the two interfering signals, we develop a framework based on a specific ensemble of coset codes - *partitioned coset codes* (PCC) (definition 3.4.2 - to enable efficient decoding of the relevant bivariate interfering component. A key element of this framework - new encoding and decoding rules based on joint typicality - lends it applicable to a general 3−IC. The other key element - binning of coset codes into PCC - enables us achieve rates corresponding to arbitrary single-letter distributions.

(ii) Analyzing the performance of this framework, we derive a new achievable rate region for a general 3−IC that subsumes the current known largest and is strictly enlarges the same for particular instances.

(iii) We identify additive and non-additive instance of 3−IC for which the derived achievable rate region is *analytically* proven to be (i) capacity achieving and (ii) strictly larger than the current known largest. The non-additive example (example 4.6.7) illustrates the utility of this framework and validates the central theme of this thesis.

1.6.2 Three user broadcast channel(3−BC)

(i) One of the techniques for communicating over a BC involves decoding the interfering signal, or a part thereof. Moreover, the other technique - precoding - being, in general, less efficient,[27] motivates decoding as large a part of the interfering signal as possible. The interfering signal over a 3−DBC being a pair of signals, we propose a framework based on PCC to decode the bivariate interfering component efficiently. This framework is analogous to the one developed for communicating over a 3−IC with certain new elements.

(ii) As in the case of a 3−IC, we analyze the performance of the proposed framework to derive a new achievable rate region for 3−DBC that subsumes the current known largest, and moreover, strictly enlarges the same for particular instances.

(iii) We identify a vector additive 3−DBC and *analytically* prove that the derived achievable rate region is strictly larger than the current known largest.

[27] This is due to the presence of a *rate loss*. In other words, if there is a choice between decoding and precoding, the former is generally preferred, as it yields higher rates of communication. However, it must be noted that decoding the interfering signal constrains the rate of the interfering signal, which is in general undesirable.

1.6.3 Multiple access channel with distributed states (MAC-DSTx)

(i) The current known coding technique for communicating over a MAC-DSTx is a natural generalization of Gel'fand and Pinsker's technique of precoding via binning [7], proposed in the context of a single Tx. In particular, for the MAC-DSTx, the two codes are independently and uniformly partitioned into bins, and the pair of chosen codewords is decoded via a joint typicality decoder. Following Philosof and Zamir [15], we develop a framework based on *nested coset codes* (NCC) (section 3.4.1) and *union coset codes* (UCC) (section 3.4.3) that facilitates favorable interaction of the bins of two codes, and thereby develop a new coding technique. In contrast to [15], this framework enables (i) exploit the structure of the coset codes for communicating over an arbitrary MAC-DSTx, and (ii) achieve rates corresponding to arbitrary single-letter distributions. Furthermore, the framework incorporates UCC built over groups (group UCC) to enable more efficient communication over a larger class of MAC-DSTx.

(ii) We analyze the performance of the proposed framework to derive a new achievable rate region for MAC-DSTx that subsumes the current known largest, and strictly enlarges the same for particular instances.

(iii) We identify several non-additive and non-symmetric instances of MAC-DSTx for which the proposed framework yields strictly larger achievable rate regions.[28] The utility of incorporating group UCC is indicated through an example.

1.6.4 Computation of sum of sources over an *arbitrary* MAC

(i) Following [16], we develop an interface, based on NCC, between the source coding module and channel coding module that enables the Rx decode the sum of sources by decoding the sum of transmitted codewords. The proposed interface, in conjunction with separation based strategy, yields a more efficient coding technique to compute the sum of sources at the Rx of a MAC. In contrast to the findings presented in [16], the interface developed herein enables computing the sum of sources over an *arbitrary* MAC.

(ii) Analyzing the performance of the proposed coding technique, we derive a new set of sufficient conditions for computing the sum of sources reliably over an *arbitrary* MAC, that are weaker than current known conditions. The utility of this framework is demonstrated through examples involving non-additive MAC.

1.7 Significance of our contribution

This thesis presents new achievable rate regions for multi-terminal systems, including the broadcast and interference channels. Since the characterization of capacity regions plays a fundamental role in our understanding of performance

[28] The examples being non-additive, it is significantly harder to provide analytical comparisons, and hence we resort to direct computation of rate regions achievable using current and proposed coding techniques.

limits, the contributions of this thesis cannot be overemphasized. Secondly, we propose new coding techniques for communicating over these multi-terminal systems. With ever increasing processing power and information theoretic techniques finding their way into practice, these coding techniques can better harness available degrees of freedom and enable efficient utilization of resources.

As observed in the past, characterizing new achievable rate regions involves a balance of new examples and new theory, with the former preceding the latter for some problems. For example, in the case of degraded and general broadcast channels, ingenious coding techniques [20], [8] for particular examples, preceded the theory [21], [9]. While for the lossless distributed source coding problem, examples and theory were concurrently revealed in [5]. Notwithstanding the order, the significance of either cannot be undervalued. While the examples have shone the light in the right direction, generalizing these ingenious coding techniques have resulted in fundamentally new ideas. For example, in his quest to generalize Cover's superposition technique for the binary additive broadcast channel, Bergmans [21] developed the fundamental technique of characterizing achievable rate regions using an auxiliary random variable.

This thesis contributes a good balance of examples and theory. For the broadcast channel, we identify the first example for which linear codes yield strictly larger achievable rate region than that using current known techniques based on superposition and binning of independent unstructured codes.[29] We build on this to develop a coding framework for communication over an arbitrary discrete three user broadcast channel. For the interference channel, while lattices [19] and interference alignment techniques [17] have been employed for continuous valued *additive* channels, we identify the first discrete $3-$IC and *analytically* prove that linear codes strictly outperform independent unstructured code based superposition coding [13].[30] Of particular significance is the identification of non-additive interference channels, such as example 4.6.7, for which linear codes built over suitably larger fields strictly outperform[31] current known techniques based on independent unstructured codes.[32] We leverage these examples to derive a new achievable rate region for an arbitrary $3-$IC involving all valid test channels. For the other two problems, MAC with state and computation over MAC, we build on novel coding techniques proposed for particular additive examples in [15] and [16] respectively. While their techniques are applicable only to symmetric and additive scenarios, we generalize the same using the machinery developed herein to derive new achievable rate regions for arbitrary problem instances. As described in section 1.4, this has involved several new elements. We validate our generalization by identifying non-additive and non-symmetric examples (sections 6.2.3 and examples 7.2.4 - 7.2.7) for which the

[29]Superposition coding as proposed by Bergmans [21] involves a conditional coded satellite codebook. However, this coding technique, and the corresponding achievable rate region can be realized using independent unstructured codebooks via the technique of Han and Kobayashi [13]. We are therefore justified in saying that conventional coding techniques for arbitrary problem instances are based on independent unstructured codes.

[30]Moreover, we note that [19], [17] prove strict sub-optimality of only Gaussian test channels.

[31]We provide an *analytical* proof of this statement in section 4.6.1.

[32]This example demonstrates the underlying theme of this thesis - codes endowed with algebraic closure properties yields strictly larger achievable rate regions even for non-additive problem instances - and thereby validates all the machinery - characterizing and analyzing performance of coset codes over arbitrary problem instances using binning and joint typicality encoding, decoding - developed in this thesis.

proposed generalization is strictly more efficient.

In the case of 3−BC and 3−IC, we have provided *analytical* proofs of strict sub-optimality of current known techniques based on independent unstructured codes. It maybe noted that description of the current known achievable rate regions for these problems involve more than 7 auxiliary random variables with loose cardinality bounds. Moreover, even a tractable characterization of these regions not involving parameters other than the three rates are not available, thus lending our task considerable difficulty.[33] We leverage (i) the structure of the identified instances and (ii) alternate converse proof techniques (sections 5.10 and 4.5, 4.6.1) to provide analytical proofs. Indeed, the examples are carefully chosen to amplify the interaction of codes that we are after, and yet simple enough, to enable us prove strict sub-optimality of current known technique. We highlight the *analytical* proof of strict suboptimality of independent unstructured codes based techniques for *non-additive* instances presented in section 4.6.1.[34]

The theory developed herein relies on characterizing performance of random multi-terminal codes whose constituent codes are statistically correlated coset codes. This builds in statistical dependence between (i) codewords of the same code, and (i) different constituent codebooks. Traditionally, analyzing performance of joint typicality based coding techniques crucially relies on statistical independence of these elements. To accommodate statistical correlation among constituent codebooks, we develop several new proof techniques to characterize the average performance of the proposed coding technique.[35]

1.8 The role of coset codes in multi-terminal information theory

We conclude this chapter by mentioning relevant prior work. The use of coset codes in deriving achievable rate regions began with Körner and Marton's [18] ingenious coding technique proposed for the particular problem of computing modulo−2 sum of distributed binary sources. Studied in the context of a source coding problem, they proposed partitioning the two quantizers using cosets of a common linear code. This was in contrast to the conventional technique of uniformly and independently partitioning the quantizers. Exploiting the coset structure of the partitions, Körner and Marton proposed a coding technique that outperformed all techniques based on unstructured codes.

Körner and Marton's technique [18], in spite of yielding strictly better performance, was not pursued upon. For over twenty five years following their work, it was unaware how to generalize their techniques to an arbitrary instance of the problem studied therein. Moreover, it was generally believed that Körner and Marton's technique was only applicable for particular symmetric and additive instances. Naturally, there were much fewer attempts at characterizing performance of other multi-terminal communication systems using coset codes.

[33] We note that proof of strict sub-optimality of independent unstructured codes for continuous valued channels restrict attention to Gaussian test channels.

[34] On a similar note, we commend Philosof and Zamir's [15] proof of strict sub-optimality of independent unstructured binning for the problem studied therein. Our attempt to generalize their proof to derive an upper bound for the mod−4 additive MAC with state (section 6.4) has been unsuccessful and we are forced to resort computation based technique.

[35] Since we employ new code ensembles, we have detailed all the proof elements. Please refer to the appropriate appendices for the same.

Recently, there has been renewed interest in the use of coset codes for characterizing performance limits of multi-terminal communication systems. Philosof and Zamir [15] propose a technique of structured precoding via correlated binning at the inputs of a particular symmetric additive binary doubly dirty MAC, followed by a new decoding technique that outperforms all earlier known techniques based on unstructured codes. Sridharan et. al. [19] employ the ensemble of lattice codes to effect interference alignment and exploit this to derive strictly better performance over a three user Gaussian IC. Nazer and Gastpar [16] employ linear codes to develop a new interface between source and channel coding modules that enable very efficient decoding of sum of sources over an additive MAC. Bresler, Parekh and Tse [22] prove achievability of strictly larger degrees of freedom over a three user Gaussian interference channel using lattice codes. For all of the above problems, we are unaware of any technique that replicate the same performance using unstructured codes.[36]

While the above works demonstrate the utility of algebraic properties in codes, their study is limited to particular symmetric additive instance of the problem studied therein. For example, the technique proposed in [15] is not applicable for a arbitrary instance of a MAC with distributed states. Similarly, Nazer and Gastpar's technique [16] heavily relies on a structural match between the sources and the channel. These, and other works, therefore do not address the reason for the long period of skepticism that followed [18], and the question whether coset codes are applicable only for particular additive and symmetric problem instances, or have a more fundamental role to play in multi-terminal information theory has remained.

More than three decades following the publication of [18], Krithivasan and Pradhan [23] develop a framework for generalizing the ingenious coding technique of Kórner Marton to an arbitrary instance of the distributed source coding (DSC) problem. In particular, they propose an ensemble of codes possessing algebraic closure properties and a coding technique that exploits these properties to derive an achievable rate region for an arbitrary instance of the DSC problem. [23] demonstrates that coset codes have a role to play in a general instance of the DSC problem, not just an additive and symmetric case as that studied in [18].

Krithivasan and Pradhan [23] provided the first leads in unravelling the role of coset codes in multi-terminal information theory. DSC being just one multi-terminal communication problem, it is natural to ask whether coset codes aid more efficient communication over other multi-terminal scenarios such as broadcast and interference channels. Motivated by these questions, this thesis continues the pursuit to unravel the role of coset codes in multi-terminal information theory.

[36]Moreover, for certain problems, such as [15], the authors therein prove strict sub-optimality of all *known* coding techniques based on unstructured codes.

Chapter 2

Typicality

In this chapter we compile together results from typicality that form the basis for most proofs in this thesis. We adopt the notion of typicality as proposed by Sundaresan in [24]. This notion of typicality is based on robust typicality proposed by Orlitsky and Roche [25] and subsequently adopted in [26]. Though slightly different from that adopted in [2], it is functionally equivalent. In the sequel, we provide definitions and state the results in their simplest form. Since the following results have been well documented in books such as [2], [26], [27] among others, we omit proofs, and allude to one of the above references for the same. Where appropriate, we supplement with additional references.

2.1 Definitions

Let $\mathcal{X}_1$, $\mathcal{X}_2$ be finite sets and $X := (X_1, X_2)$, a pair of random variables taking values in $\mathcal{X} := \mathcal{X}_1 \times \mathcal{X}_2$ with pmf $p_X := p_{X_1X_2}$. Let $X^n := (X_1^n, X_2^n)$ be n independent and identically distributed copies of X. For a pair $a = (a_1, a_2) \in \mathcal{X}$, and an $n-$tuple $x^n := (x_1^n, x_2^n) \in \mathcal{X}^n$, let $N(a|x^n) = \sum_{i=1}^n 1_{\{x_i=a\}}$ be the number of occurrences of a in x^n. Lastly, for $j \in \{1,2\}$, let $\dot{j} \in \{1,2\} \setminus \{j\}$ denote the element in it's complement. We are now set to define typical set. For any $\delta > 0$, let

$$T_\delta := \left\{ x^n \in \mathcal{X}^n : \left| \frac{N(a|x^n)}{n} - p_X(a) \right| \leq \frac{\delta p_X(a)}{\log |\mathcal{X}|} \text{ for all } a \in \mathcal{X} \right\}$$

be the typical set on $\mathcal{X}$ with respect to pmf p_X and parameter $\delta > 0$. For $j = 1, 2$, the projection

$$T_\delta(X_j) := \left\{ x_j^n \in \mathcal{X}_j^n : \text{ there exists } x_{\dot{j}}^n \in \mathcal{X}_{\dot{j}}^n \text{ such that } (x_1^n, x_2^n) \in T_\delta \right\}$$

is the typical set on $\mathcal{X}_j$ with respect to pmf p_X and parameter $\delta > 0$. For $j = 1, 2$ and any $x_{\not j}^n \in \mathcal{X}_{\not j}^n$,

$$T_\delta(X_j|x_{\not j}^n) := \left\{x_j^n \in \mathcal{X}_j^n \text{ such that } (x_1^n, x_2^n) \in T_\delta\right\}$$

is the typical set on $\mathcal{X}_j$ conditioned on $x_{\not j}^n$ with respect to distribution p_X and parameter $\delta > 0$. Before we state the basic results, the following remarks are worth noting.

2.2 Simple consequences

Remark 2.2.1 *If for any $a \in \mathcal{X}$, $p_X(a) = 0$, and $x^n \in T_\delta$, then $N(a|x^n) = 0$.*

Remark 2.2.2 *If $x_j^n \in T_\delta(X_j)$, then $\left|\frac{N(a_j|x_j^n)}{n} - p_{X_j}(a_j)\right| \leq \frac{\delta p_{X_j}(a_j)}{\log|\mathcal{X}|}$. Since $x_j^n \in T_\delta(X_j)$, there exists $x_{\not j}^n \in \mathcal{X}_{\not j}^n$ such that $(x_j^n, x_{\not j}^n) \in T_\delta$, and for this $x_{\not j}^n$, we have*

$$\left|\frac{N(a_j|x_j^n)}{n} - p_{X_j}(a_j)\right| = \left|\sum_{a_{\not j} \in \mathcal{X}_{\not j}} \frac{N(a_j, a_{\not j}|x_j^n, x_{\not j}^n)}{n} - \sum_{a_{\not j} \in \mathcal{X}_{\not j}} p_{X_j X_{\not j}}(a_j, a_{\not j})\right| \leq \sum_{a_{\not j} \in \mathcal{X}_{\not j}} \left|\frac{N(a_j, a_{\not j}|x_j^n, x_{\not j}^n)}{n} - p_{X_j X_{\not j}}(a_j, a_{\not j})\right|$$
$$\leq \sum_{a_{\not j} \in \mathcal{X}_{\not j}} \frac{\delta p_{X_j X_{\not j}}(a_j, a_{\not j})}{\log|\mathcal{X}|} = \frac{\delta p_{X_j}(a_j)}{\log|\mathcal{X}|}.$$

Lemma 2.2.3 *If $x^n \in T_\delta$, then for every $n \in \mathbb{N}$, we have*

(i) $|\frac{1}{n}\log p_{X^n}(x^n) + H(X)| \leq \delta$,

(ii) $|\frac{1}{n}\log p_{X_j^n}(x_j^n) + H(X_j)| \leq \delta$ *for* $j \in [2]$ *and therefore*

(iii) $|\frac{1}{n}\log p_{X_j^n|X_{\not j}^n}(x_j^n|x_{\not j}^n) + H(X_j|X_{\not j})| \leq 2\delta$.

□

Proof: Observe that

$$\left|\frac{1}{n}\log p_{X^n}(x^n) + H(X)\right| \quad = \quad \left|\frac{1}{n}\sum_{i=1}^n \log p_X(x_i) + H(X)\right| = \left|\sum_{a \in \mathcal{X}} \frac{N(a|x^n)}{n}\log p_X(a) + H(X)\right|.$$

Substituting upper and lower bounds $p_X(a) - \frac{\delta p_X(a)}{\log|\mathcal{X}|} \leq \frac{N(a|x^n)}{n} \leq p_X(a) + \frac{\delta p_X(a)}{\log|\mathcal{X}|}$ on $\frac{N(a|x^n)}{n}$ and employing the definition of $H(X)$, we have

$$\left|\frac{1}{n}\log p_{X^n}(x^n) + H(X)\right| \quad \leq \quad \sum_{a \in \mathcal{X}} \frac{\delta p_X(a)\log p_X(a)}{\log|\mathcal{X}|} \leq \delta,$$

where the last inequality follows from $H(X) \leq \log|\mathcal{X}|$. This proves (i). In order to prove (ii), it suffices to prove $\left|\frac{N(a_j|x_j^n)}{n} - p_{X_j}(a_j)\right| \leq \frac{\delta p_{X_j}(a_j)}{\log|\mathcal{X}_j|}$ for each $a_j \in \mathcal{X}_j$. We may then employ a sequence of steps analogous to the one

above. This has been established in remark 2.2.2. We are left to argue statement (iii). Statement (iii) is a simple consequence of $\frac{1}{n}\log p_{X_j^n|X_{\not j}^n}(x_j^n|x_{\not j}^n) = \frac{1}{n}\log p_X(x^n) - \frac{1}{n}\log p_{X_{\not j}}(x_{\not j}^n)$, the bounds in statements (i) and (ii) and the triangular inequality. ■

2.3 Typical sets are large and highly probable

Lemma 2.3.1 *For every $\epsilon > 0$, $\delta > 0$, there exists $N(\epsilon,\delta) \in \mathbb{N}$, such that for every $n \geq N(\epsilon,\delta)$, $P(X^n \in T_\delta) \geq 1-\epsilon$, and therefore, $P(X_j^n \in T_\delta(X_j)) \geq 1-\epsilon$, for each $j \in [2]$. Moreover,*

$$P\left(X^n \notin T_\delta\right) \leq 2\exp\left\{-n^3\delta^2\lambda\right\} \text{ where } \lambda = \min\left\{\frac{2p_X^2(a)\log e}{\log|\mathcal{X}|} : p_X(a) > 0, a \in \mathcal{X}\right\} \tag{2.1}$$

□

Proof: Note that (2.1) reiterates the first statement of the lemma with a tighter bound. While the first statement can be proved using Cheybyshev inequality, the second statement, due to Hoeffding [28], Sanov [29], requires a finer analysis. We begin with the proof of the first statement.

Note that $N(a|X^n)$ is a binomial random variable with $P(N(a|X^n) = k) = \binom{n}{k} p_X(a)^k(1-p_X(a))^{n-k}$. For every $a \in \mathcal{X}$ such that $p_X(a) = 0$, we have

$$P\left(\left|\frac{N(a|X^n)}{n} - p_X(a)\right| > \frac{\delta p_X(a)}{\log|\mathcal{X}|}\right) = P\left(N(a|X^n) > 0\right) = 0$$

As a consequence of this, union bound and the Cheybyshev inequality, we have

$$\begin{aligned}
P(X^n \notin T_\delta) &= P\left(\bigcup_{a\in\mathcal{X}}\left\{\left|\frac{N(a|X^n)}{n} - p_X(a)\right| > \frac{\delta p_X(a)}{\log|\mathcal{X}|}\right\}\right) \leq \sum_{\substack{a\in\mathcal{X}:\\ p_X(a)>0}} P\left(\left|\frac{N(a|X^n)}{n} - p_X(a)\right| > \frac{\delta p_X(a)}{\log|\mathcal{X}|}\right) \qquad (2.2)\\
&\leq \sum_{\substack{a\in\mathcal{X}:\\ p_X(a)>0}} \frac{\mathrm{Var}\left\{\frac{N(a|X^n)}{n}\right\}(\log|\mathcal{X}|)^2}{\delta^2 p_X(a)^2} = \sum_{\substack{a\in\mathcal{X}:\\ p_X(a)>0}} \frac{np_X(a)(1-p_X(a))(\log|\mathcal{X}|)^2}{n^2\delta^2 p_X(a)^2}\\
&= \frac{(\log|\mathcal{X}|)^2}{n\delta^2}\left[\sum_{\substack{a\in\mathcal{X}:\\ p_X(a)>0}} \frac{(1-p_X(a))}{p_X(a)}\right] \leq \frac{\theta|\mathcal{X}|(\log|\mathcal{X}|)^2}{n\delta^2}, \text{ where } \theta = \min\left\{\frac{(1-p_X(a))}{p_X(a)} : p_X(a) > 0\right\}.
\end{aligned}$$

Given $\epsilon > 0$ and $\delta > 0$, choose $N(\epsilon,\delta) = \frac{1}{\epsilon}\left\lceil\frac{\theta|\mathcal{X}|(\log|\mathcal{X}|)^2}{\delta^2}\right\rceil$ and note that for all $n \geq N(\epsilon,\delta)$, $P(X^n \notin T_\delta) < \epsilon$. By definition, $x^n \in T_\delta$ implies $x_j^n \in T_\delta(X_j)$ for each $j = 1,2$. Therefore, for $n \geq N(\epsilon,\delta)$, we have $1-\epsilon \leq P(X_n \in T_\delta) \leq P(X_j^n \in T_\delta(X_j))$ for each $j = 1,2$.

We now provide a sketch of the argument that proves the tighter upper bound stated in (2.1). The argument is based on the following lemma found in [2, Problem 3.18(b), page 44].

Lemma 2.3.2 *If $Z_1, Z_2, \cdots$ are independent and identically distributed Bernoulli random variables taking values in $\{0,1\}$ with $P(Z_i = 1) = p$, then*

$$P\left(\left|\sum_{i=1}^{n} Z_i - np\right| > \eta\right) \leq 2e^{-2\eta^2 n}$$

. □

Substituting $Z_i = 1_{\{X_i = a\}}, \eta = \frac{n\delta p_X(a)}{\log|\mathcal{X}|}$, we recognize $\sum_{i=1}^{n} Z_i = N(a|X^n)$, and therefore lemma 2.3.2 implies

$$P\left(\left|\frac{N(a|X^n)}{n} - p_X(a)\right| > \frac{\delta p_X(a)}{\log|\mathcal{X}|}\right) \leq 2\exp\left\{-\frac{2n^3\delta^2 p_X^2(a)\log e}{(\log|\mathcal{X}|^2)}\right\}. \tag{2.3}$$

Substituting (2.3) in (2.2), we have

$$P(X^n \notin T_\delta) \leq \sum_{\substack{a\in\mathcal{X}:\\ p_X(a)>0}} 2\exp\left\{-\frac{2n^3\delta^2 p_X^2(a)\log e}{(\log|\mathcal{X}|^2)}\right\} \leq 2\exp\left\{-n^3\delta^2\lambda\right\} \text{ where } \lambda \text{ is as defined in (2.1).} \tag{2.4}$$

■

Lemma 2.3.3 *For every $\delta > 0$, there exists $N_1(\delta), N_2(\delta) \in \mathbb{N}$, such that,*

(i) for every $n \geq N_1(\delta)$, $\exp\{n(H(X) - 2\delta)\} \leq |T_\delta| \leq \exp\{n(H(X) + 2\delta)\}$, and

(ii) for every $n \geq N_2(\delta)$, $\exp\{n(H(X_j) - 2\delta)\} \leq |T_\delta(X_j)| \leq \exp\{n(H(X_j) + 2\delta)\}$.

□

Proof: From lemma 2.2.3(i), we have $p_{X^n}(x^n) \geq \exp\{-n(H(X) + \delta)\}$ for every $x^n \in T_\delta$. We therefore have

$$1 \geq P(X^n \in T_\delta) = \sum_{x^n\in T_\delta} p_{X^n}(x^n) \geq \sum_{x^n\in T_\delta} \exp\{-n(H(X) + \delta)\} \geq |T_\delta|\exp\{-n(H(X) + \delta)\}$$

which gives us the upper bound on $|T_\delta|$. We employ the lower bound on the probability of the typical set derived in lemma 2.3.1 for establishing the lower bound on $|T_\delta|$. For $n \geq N(\delta, \delta)$, we have

$$1 - \delta \leq P(X^n \in T_\delta) = \sum_{x^n\in T_\delta} p_{X^n}(x^n) \leq \sum_{x^n\in T_\delta} \exp\{-n(H(X) - \delta)\} \leq |T_\delta|\exp\{-n(H(X) - \delta)\}$$

which implies $|T_\delta| \geq (1-\delta)\exp\{n(H(X) - \delta)\}$. For $n \geq \max\{N(\delta,\delta), \lceil\frac{1}{\delta}\log\frac{1}{1-\delta}\rceil\}$, we have $|T_\delta| \geq \exp\{n(H(X) - 2\delta)\}$. Statement (ii) can be proved following an analogous sequence of steps. ■

2.4 And so are conditional typical sets

Lemma 2.4.1 *For every $\epsilon > 0$, $\delta > 0$, there exists $N(\epsilon, \delta) \in \mathbb{N}$, such that for every $n \geq N(\epsilon, \delta)$, $x_{\not j}^n \in T_\delta(X_{\not j})$, implies $P(X_j^n \in T_{2\delta}(X_j|x_{\not j}^n)|X_{\not j}^n = x_{\not j}^n) \geq 1 - \epsilon$ and therefore $P\left(X_{\not j} \in T_\delta(X_{\not j}), X^n \notin T_{2\delta}\right) \leq \epsilon$.* □

Proof: We prove the statement in the lemma for $j = 2$, i.e, $\not j = 1$. We begin with an alternate characterization of $\mathcal{X}_2^n \setminus T_{2\delta}(X_2^n|x_1^n)$, i.e., the complement of $T_{2\delta}(X_2^n|x_1^n)$. Note that if $x_2^n \in \mathcal{X}_2^n \setminus T_{2\delta}(X_2^n|x_1^n)$, then for some $a \in \mathcal{X}$, we have $\left|\frac{N(a|x^n)}{n} - p_X(a)\right| > \frac{2\delta p_X(a)}{\log|\mathcal{X}|}$. Also note that if $N(a_1|x_1^n) = 0$, then $p_{X_1}(a_1) \leq \frac{\delta p_{X_1}(a_1)}{\log|\mathcal{X}|}$.[1] For any $a_2 \in \mathcal{X}_2$, we have

$$p_X(a) = p_{X_1}(a_1)p_{X_2|X_1}(a_2|a_1) \leq \frac{\delta p_{X_1}(a_1)p_{X_2|X_1}(a_2|a_1)}{\log|\mathcal{X}|} = \frac{\delta p_X(a)}{\log|\mathcal{X}|}$$

and therefore

$$\left|\frac{N(a|x^n)}{n} - p_X(a)\right| = |-p_X(a)| = p_X(a) \leq \frac{\delta p_X(a)}{\log|\mathcal{X}|}.$$

In characterizing $\mathcal{X}_2^n \setminus T_{2\delta}(X_2^n|x_1^n)$, we only need to consider $\mathfrak{P}(x_1^n) := \{a_1 \in \mathcal{X}_1 : N(a_1|x_1^n) > 0\}$. For $a_1 \in \mathfrak{P}(x_1^n)$, we have

$$\begin{aligned}
\left|\frac{N(a|x^n)}{n} - p_X(a)\right| &= \left|\frac{N(a_1|x_1^n)}{n}\left(\frac{N(a|x^n)}{N(a_1|x_1^n)} - p_{X_2|X_1}(a_2|a_1)\right) + p_{X_2|X_1}(a_2|a_1)\left(\frac{N(a_1|x_1^n)}{n} - p_{X_1}(a_1)\right)\right| \\
&\leq \frac{N(a_1|x_1^n)}{n}\left|\frac{N(a|x^n)}{N(a_1|x_1^n)} - p_{X_2|X_1}(a_2|a_1)\right| + p_{X_2|X_1}(a_2|a_1)\left|\frac{N(a_1|x_1^n)}{n} - p_{X_1}(a_1)\right| \\
&\leq \frac{N(a_1|x_1^n)}{n}\left|\frac{N(a|x^n)}{N(a_1|x_1^n)} - p_{X_2|X_1}(a_2|a_1)\right| + \frac{\delta p_X(a)}{\log|\mathcal{X}|}
\end{aligned}$$

This implies that if $x_2^n \in \mathcal{X}_2^n \setminus T_{2\delta}(X_2^n|x_1^n)$, then for some $a_1 \in \mathfrak{P}(x_1^n)$, we have $\left|\frac{N(a|x^n)}{N(a_1|x_1^n)} - p_{X_2|X_1}(a_2|a_1)\right| > \frac{n\delta p_X(a)}{N(a_1|x_1^n)\log|\mathcal{X}|}$. This enables us conclude

$$\mathcal{X}_2^n \setminus T_{2\delta}(X_2^n|x_1^n) \subseteq \left[\bigcup_{a_1 \in \mathfrak{P}(x_1^n)} \bigcup_{\substack{a_2 \in \mathcal{X}_2: \\ p_{X_2|X_1}(a_2|a_1) > 0}} \left\{x_2^n \in \mathcal{X}_2^n : \left|\frac{N(a|x^n)}{N(a_1|x_1^n)} - p_{X_2|X_1}(a_2|a_1)\right| > \frac{n\delta p_X(a)}{N(a_1|x_1^n)\log|\mathcal{X}|}\right\}\right] \bigcup$$
$$\left[\bigcup_{a_1 \in \mathfrak{P}(x_1^n)} \bigcup_{\substack{a_2 \in \mathcal{X}_2: \\ p_{X_2|X_1}(a_2|a_1) = 0}} \{x_2^n \in \mathcal{X}_2^n : N(a|x^n) > 0\}\right].$$

With an intent of employing the union bound, we provide upper bounds on the probability of each set in the above union. We begin with the following observation. Conditioned on $X_1^n = x_1^n$, note that $N(a|x_1^n, X_2^n)$ is a binomial

[1] This follows from $x_1^n \in T_\delta(X_1)$.

random variable with parameters $N(a_1|x_1^n), p_{X_2|X_1}(a_2|a_1)$, i.e.,

$$P(N(a|x_1^n, X_2^n) = k|X_1^n = x_1^n) = \binom{N(a_1|x_1^n)}{k} p_{X_2|X_1}(a_2|a_1)^k (1 - p_{X_2|X_1}(a_2|a_1))^{N(a_1|x_1^n)-k}.$$

For $a_1 \in \mathfrak{P}(x_1^n)$ and $a_2 \in \mathcal{X}_2$ such that $p_{X_2|X_1}(a_2|a_1) = 0$, we have $P(N(a|x_1^n, X_2^n) > 0|X_1^n = x_1^n) = 0$. For any $a_1 \in \mathfrak{P}(x_1^n)$ and $a_2 \in \mathcal{X}_2$ such that $p_{X_2|X_1}(a_2|a_1) > 0$, we have

$$P\left(\left|\frac{N(a|x_1^n, X_2^n)}{N(a_1|x_1^n)} - p_{X_2|X_1}(a_2|a_1)\right| > \alpha\right) \leq \mathbb{E}\left\{\left(\frac{N(a|x_1^n, X_2^n)}{N(a_1|x_1^n)} - p_{X_2|X_1}(a_2|a_1)\right)^2\right\} \frac{1}{\alpha^2},$$

where the intended value for α is $\frac{n\delta p_X(a)}{N(a_1|x_1^n)\log|\mathcal{X}|}$. It maybe verified that

$$\mathbb{E}\left\{\left(\frac{N(a|x_1^n, X_2^n)}{N(a_1|x_1^n)} - p_{X_2|X_1}(a_2|a_1)\right)^2\right\} = \frac{p_{X_2|X_1}(a_2|a_1)(1 - p_{X_2|X_1}(a_2|a_1))}{N(a_1|x_1^n)},$$

and therefore,

$$P\left(\left|\frac{N(a|x_1^n, X_2^n)}{N(a_1|x_1^n)} - p_{X_2|X_1}(a_2|a_1)\right| > \frac{n\delta p_X(a)}{N(a_1|x_1^n)\log|\mathcal{X}|}\right) \leq \frac{p_{X_2|X_1}(a_2|a_1)(1 - p_{X_2|X_1}(a_2|a_1))}{N(a_1|x_1^n)} \frac{(N(a_1|x_1^n)\log|\mathcal{X}|)^2}{n^2\delta^2 p_X^2(a)},$$
$$\leq \frac{1 - p_{X_2|X_1}(a_2|a_1)}{\delta^2 p_{X_2|X_1}(a_2|a_1)} \frac{(\log|\mathcal{X}|)^2}{n}.$$

Substituting this in the probability of the desired event, we have

$$P(X_2^n \notin T_{2\delta}(X_2|x_1^n)|X_1^n = x_1^n) \leq |\mathcal{X}| \frac{1 - p^*}{\delta^2 p^*} \frac{(\log|\mathcal{X}|)^2}{n},$$

where $p^* = \min\{p_{X_2|x_1}(a_2|a_1) : p_{X_1}(a_1) > 0, p_{X_2|x_1}(a_2|a_1) > 0\}$. Given $\epsilon > 0$ and $\delta > 0$, choose $N(\epsilon, \delta) = \frac{1}{\epsilon}\lceil|\mathcal{X}|\frac{1-p^*}{\delta^2 p^*}(\log|\mathcal{X}|)^2\rceil$ and note that for $n \geq N(\epsilon, \delta)$, we have $P(X_2^n \notin T_{2\delta}(X_2|x_1^n)|X_1^n = x_1^n) \leq \epsilon$ whenever $x_1^n \in T_\delta(X_1)$. ■

Lemma 2.4.2 *For every $\delta > 0$, there exists $N(\delta) \in \mathbb{N}$, such that, for every $n \geq N(\delta)$, $x_{\not j}^n \in T_\delta(X_{\not j})$ we have* $\exp\{n(H(X_j|X_{\not j}) - 4\delta)\} \leq \left|T_{2\delta}(X_j|x_{\not j}^n)\right| \leq \exp\{n(H(X_j|X_{\not j}) + 4\delta)\}$. □

Proof: Quite naturally, the proof mimics that of lemma 2.3.3. We have

$$1 \geq P(X_2^n \in T_{2\delta}(X_2|x_1^n)|X_1^n = x_1^n) = \sum_{x_2^n \in T_{2\delta}(X_2|x_1^n)} P(X_2^n = x_2^n|X_1^n = x_1^n)$$
$$\geq \sum_{x_2^n \in T_{2\delta}(X_2|x_1^n)} \exp\{-n(H(X_2|X_1) + 3\delta)\} = |T_{2\delta}(X_2^n|x_1^n)|\{-n(H(X_2|X_1) + 3\delta)\}.$$

which gives us the upper bound. For $n \geq N(\delta, \delta)$, we have

$$1 - \delta \leq P(X_2^n \in T_{2\delta}(X_2|x_1^n)|X_1^n = x_1^n) = \sum_{x_2^n \in T_{2\delta}(X_2|x_1^n)} P(X_2^n = x_2^n|X_1^n = x_1^n)$$
$$\leq \sum_{x_2^n \in T_{2\delta}(X_2|x_1^n)} \exp\left\{-n(H(X_2|X_1) - 3\delta)\right\} = |T_{2\delta}(X_2^n|x_1^n)| \left\{-n(H(X_2|X_1) - 3\delta)\right\}.$$

For $n \geq \max\{N(\delta, \delta)\frac{1}{\delta}\log\frac{1}{1-\delta}\}$, we have $|T_{2\delta}(X_2^n|x_1^n)| \geq \exp\left\{n(H(X_2|X_1) - 4\delta)\right\}$. ■

Chapter 3

Coset codes achieve capacity of general point-to-point channels

The three objectives of this chapter are the following. Firstly, we intend to describe the three ensembles of coset codes that will be employed for deriving new achievable rate regions in this thesis. Secondly, we wish to characterize the performance of these ensembles for communicating in the presence of noise. Thirdly, we desire to present the two other building blocks - joint typicality encoding, decoding and binning - in a simple, yet non-trivial setting, that enables the reader absorb the underlying idea and study the proofs without getting bogged down by too many technicalities. An interested reader is therefore strongly encouraged to read through this chapter carefully.

We can satisfy our second objective by characterizing the performance of coset codes in communicating over a general PTP. However, communicating over multi-terminal systems necessitates codes to possess *covering* properties, in addition to *packing* properties.[1] We therefore characterize the performance of coset codes in communicating over a general point-to-point channel with knowledge of channel state at transmitter (PTP-STx). PTP-STx being the simplest communication channel that employs a code possessing both packing and covering properties, motivates our choice.

As depicted in figure 3.1, a PTP-STx is a PTP whose channel transition probabilities depend on a random parameter S called *state*. The evolution of the state is assumed to be iid across time with respect to distribution W_S, and moreover, the encoder is provided the entire realization of the state sequence before communication begins. The channel is assumed discrete, memoryless, time-invariant and used without feedback.[2] The objective is to design an optimal strategy that enables the encoder utilize the state information to efficiently communicate an information stream to the decoder and thereby characterize the capacity region of PTP-STx.

[1] The packing properties of a code determine it's ability to communicate in the presence of noise.
[2] These will be precisely defined in section 3.2.

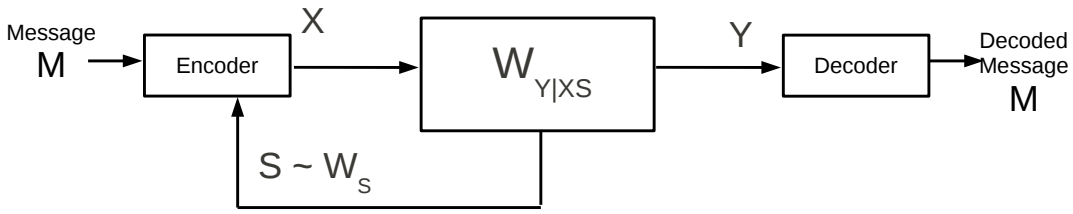

Figure 3.1: A point-to-point channel with knowledge of channel state at transmitter (PTP-STx).

In a celebrated result [7], Gel'fand and Pinsker derived a single-letter characterization for the capacity of PTP-STx. Their proof of achievability employs a code possessing both packing and covering properties and a coding technique that exploits the same. In order to achieve capacity, the code must simultaneously possess optimal packing and covering properties. Gel'fand and Pinsker prove existence of such a code via the random coding technique wherein the letters of random code is iid according to a single letter distribution. Clearly, the capacity achieving code is not guaranteed to possess any additional properties, such as algebraic closure that is of interest herein.

In this chapter, our goal is to characterize the performance of the three ensembles of codes for communicating over a PTP-STx. In particular, we would like to know whether the three ensembles of coset codes possess optimal covering and packing properties that enable them achieve capacity of PTP-STx? If not, what rates are achievable over an arbitrary PTP-STx by restricting to these coset codes?

We prove the three ensembles of coset codes achieve capacity of an *arbitrary* PTP-STx.[3] In other words, the property of algebraic closure and optimal packing, covering properties are *not* mutually exclusive. We wish to note that these three ensembles of coset codes are currently the only ensemble of codes possessing an algebraic structure that has been proven to achieve capacity of an arbitrary PTP-STx. This assumes significance in the light of Ahlswede's finding [30] that linear codes do *not* achieve capacity of an arbitrary PTP.

This chapter is organized as follows. Sections 3.1, 3.2 and 3.3 state the preliminaries - notation, definitions and a single-letter characterization of capacity of PTP-STx. In section 3.4.1, we describe the three ensembles of coset codes. We prove that these ensembles achieve capacity of PTP-STx in section 3.5.

3.1 Notation

We employ notation that is now widely employed in information theory literature supplemented by the following.

- We let $\mathbb{N}, \mathbb{R}$ denote the set of natural numbers and real numbers respectively. Calligraphic letters such as $\mathcal{X}$, $\mathcal{Y}$ are employed exclusively to denote finite sets. $\mathcal{F}_q$ denotes the finite field of cardinality q. For any set A,

[3]Since PTP-STx is a generalization of a PTP, this also proves the three ensembles of coset codes achieve capacity of an *arbitrary* PTP.

$\mathrm{cl}(A), \mathrm{cocl}(A)$ denote closure of A and closure of the convex hull of A respectively. If A is a finite set, $|A|$ denotes cardinality of A.

- For positive integers $i \leq j$, $[i:j] := \{i, i+1, \cdots, j\}$. We let $[j] := [1:j]$.
- While $+$ denotes addition in $\mathbb{R}$, we let $\oplus$ denote addition in a finite field. The particular finite field, which is uniquely determined (up to an isomorphism) by it's cardinality, is clear from context. When ambiguous, or to enhance clarity, we specify addition in $\mathcal{F}_q$ using $\oplus_q$. For $a, b \in \mathcal{F}_q$, $a \ominus b := a \oplus (-b)$, where $(-b)$ is the additive inverse of b.
- If $f : \mathcal{U} \to \mathcal{X}$ is a map, the n-letter extension of f denoted $f^n : \mathcal{U}^n \to \mathcal{X}^n$ is defined $f^n(u^n) := (f(u_i) : i \in [n])$.
- We employ the standard notation for probability mass functions (pmf). For example, if p_{UXSY} is a pmf on $\mathcal{U} \times \mathcal{X} \times \mathcal{S} \times \mathcal{Y}$, then p_{UY} is the corresponding marginal on $\mathcal{U} \times \mathcal{Y}$. p_{UY}^n is the pmf on $\mathcal{U}^n \times \mathcal{Y}^n$ obtained as an $n-$fold product of p_{UY} i.e., $p_{UY}^n(u^n, y^n) = \prod_{i=1}^n p_{UY}(u_i, y_i)$. We write $U \sim p_U$ if p_U is the pmf of U.
- The log and exp functions are taken with respect to the same base. For concreteness, the base may be assumed to be 2, in which case units for information theoretic quantities such as entropy and mutual information would be bits/symbol.
- For $a \in \mathbb{N}$, $\pi(a) := \min\{k \in \mathbb{N} : k \geq a, k \text{ is a prime power}\}$.
- For a pmf p_{UXSY} defined on $\mathcal{U} \times \mathcal{X} \times \mathcal{S} \times \mathcal{Y}$, let

$$\mathcal{R}(p_{UXSY}, U) := \{u \in \mathcal{U} : \exists (x, s, y) \in \mathcal{X} \times \mathcal{S} \times \mathcal{Y} : p_{UXSY}(u, x, s, y) > 0\}$$

  denote the essential range of U. When clear from context, we omit the underlying pmf and let $\mathcal{R}(U)$ denote $\mathcal{R}(p_{UXSY}, U)$.

3.2 Definitions - PTP-STx, achievability and capacity

Consider a point-to-point channel with knowledge of channel state at transmitter (PTP-STx) studied by Gel'fand and Pinsker [7]. Let $\mathcal{X}$ and $\mathcal{Y}$ denote finite input and output alphabet sets respectively. Transition probabilities depend on a random parameter, called state, that takes values in a finite set $\mathcal{S}$. The discrete time channel is (i) time invariant, i.e., pmf of Y_i, the output at time i, conditioned on (X_i, S_i), the input and state at time i, is invariant with i, (ii) memoryless, i.e., Y_i is conditionally independent of $(X_t, S_t) : 1 \leq t < i$ given (X_i, S_i), and (iii) used without feedback, i.e., encoder has no knowledge of outputs observed by decoder. Let $W_{Y|XS}(y|x, s)$ be the probability of observing $y \in \mathcal{Y}$ at the output given $x \in \mathcal{X}$ is input to PTP-STx in state $s \in \mathcal{S}$. The state at time i, S_i is (i) independent of $(X_t, S_t, Y_t) : 1 \leq t < i$, and (ii) identically distributed for all i. Let $W_S(s)$ be probability

of PTP-STx being in state $s \in \mathcal{S}$. We assume the sequence of states is non-causally available at the encoder. The input is constrained with respect to a cost function $\kappa : \mathcal{X} \times \mathcal{S} \to [0, \infty)$. We assume that the cost is time-invariant and additive i.e., cost of input X^n to the channel in state S^n is $\bar{\kappa}^n(X^n, S^n) := \frac{1}{n}\sum_{i=1}^n \kappa(X_i, S_i)$. We refer to this channel as PTP-STx $(\mathcal{S}, W_S, \mathcal{X}, \kappa, \mathcal{Y}, W_{Y|XS})$.

Definition 3.2.1 *A PTP-STx code $(n, \mathscr{M}, e, d)$ consists of (i) an index set $\mathcal{M}$ of messages, of cardinality $\mathscr{M}$, (ii) an encoder map $e : \mathcal{M} \times \mathcal{S}^n \to \mathcal{X}^n$, and (iii) a decoder map $d : \mathcal{Y}^n \to \mathcal{M}$.*

Assuming a uniform pmf on the set of messages, we define the average error probability and the cost of a PTP-STx code.

Definition 3.2.2 *The error probability of PTP-STx code $(n, \mathscr{M}, e, d)$ conditioned on message $m \in \mathcal{M}$ is*

$$\xi(e, d|m) := \sum_{s^n \in \mathcal{S}^n} \sum_{\substack{y^n : d(y^n) \\ \neq m}} W_{S^n}(s^n) W_{Y^n|X^n, S^n}(y^n | e(m, s^n), s^n).$$

The average error probability of PTP-STx code $(n, \mathscr{M}, e, d)$ is $\bar{\xi}(e, d) := \sum_{m=1}^{\mathscr{M}} \frac{1}{\mathscr{M}} \xi(e, d|m)$. The average cost of transmitting message $m \in \mathcal{M}$ is $\tau(e|m) := \sum_{s^n \in \mathcal{S}^n} W_{S^n}(s^n) \bar{\kappa}^n(e(m, s^n), s^n)$ and the average cost of PTP-STx code $(n, \mathscr{M}, e, d)$ is $\tau(e) := \frac{1}{\mathscr{M}} \sum_{m=1}^{\mathscr{M}} \tau(e|m)$.

Definition 3.2.3 *A rate cost pair $(R, \tau) \in [0, \infty)^2$ is achievable if for every $\eta > 0$, there exists $N(\eta) \in \mathbb{N}$ such that for all $n > N(\eta)$, there exists a PTP-STx code $(n, \mathscr{M}^{(n)}, e^{(n)}, d^{(n)})$ such that (i) $\frac{\log \mathscr{M}^{(n)}}{n} \geq R - \eta$, (ii) $\bar{\xi}(e^{(n)}, d^{(n)}) \leq \eta$, and (iii) average cost $\tau(e^{(n)}) \leq \tau + \eta$. The capacity region is $\mathbb{C}(\tau) := cl\{R \geq 0 : (R, \tau)$ is achievable$\}$.*

In a celebrated result, Gel'fand and Pinsker [7] derived a single letter characterization of $\mathbb{C}(\tau)$. In the next section, we state this characterization.

3.3 Capacity of PTP-STx

Definition 3.3.1 *Let $\overline{\mathbb{D}}(\tau)$ be the collection of pmfs p_{UXSY} on $\mathcal{U} \times \mathcal{X} \times \mathcal{S} \times \mathcal{Y}$ such that (i) $\mathcal{U}$ is a finite set, (ii) $p_S = W_S$, (iii) $p_{Y|XSU} = p_{Y|XS} = W_{Y|XS}$, (iv) $p_{X|SU}(x|s, u) \in \{0, 1\}$ for all $(u, x, s) \in \mathcal{U} \times \mathcal{X} \times \mathcal{S}$ and (v) $\mathbb{E}\{\kappa(X, S)\} \leq \tau$. Let*

$$\mathbb{D}(\tau) = \left\{ p_{UXSY} \in \overline{\mathbb{D}}(\tau) : |\mathcal{R}(p_{UXSY}, U)| \leq \min\{(|\mathcal{X}| \cdot |\mathcal{S}|)^2, (|\mathcal{X}| + |\mathcal{S}| + |\mathcal{Y}| - 2) \cdot |\mathcal{X}| \cdot |\mathcal{S}|\} \right\}.$$

For any pmf p_{UXSY} defined on $\mathcal{U} \times \mathcal{X} \times \mathcal{S} \times \mathcal{Y}$, let $\alpha(p_{UXSY}) := [0, I(U;Y) - I(U;S)]$, and

$$\overline{\alpha}(\tau) := cocl\left(\bigcup_{p_{UXSY} \in \overline{\mathbb{D}}(\tau)} \alpha(p_{UXSY}) \right), \alpha(\tau) := cocl\left(\bigcup_{p_{UXSY} \in \mathbb{D}(\tau)} \alpha(p_{UXSY}) \right).$$

Theorem 3.3.2 $\mathbb{C}(\tau) = \alpha(\tau) = \overline{\alpha}(\tau)$. □

Gel'fand and Pinsker [7] proved theorem 3.3.2 for channels without a cost constraint. While the central elements of their proof can be adopted for cost constrained channels, the sufficiency of restricting to test channels p_{USXY} satisfying condition (iv) in definition 3.3.1 is established in [31, Lemma 2], which is attributed to Cohen. A cardinality bound on $|\mathcal{U}|$ can be established using Fenchel-Eggleston strengthening of Carathéodory's theorem [26, Appendix C] as done in [32, Lemma 9]. In particular, one can first prove the upper bound $\min\{|\mathcal{X}| \cdot |\mathcal{S}|, |\mathcal{X}| + |\mathcal{S}| + |\mathcal{Y}| - 2\}$ on $|\mathcal{U}|$ for test channels p_{USXY} that do not satisfy condition (iv) in definition 3.3.1. Any such test channel p_{USXY} can be mapped to a test channel $p_{\tilde{U}SXY}$ that satisfies condition (iv) in definition 3.3.1 without compromising on the achievable rate for which $|\tilde{\mathcal{U}}| \leq |\mathcal{X}| \cdot |\mathcal{S}| \cdot |\mathcal{U}|$.

3.4 Nested, partitioned and union coset PTP-STx codes

Gel'fand and Pinsker prove achievability of $\mathbb{C}(\tau)$ by averaging error probability over an ensemble of PTP-STx codes. A code in this ensemble is specified by a corresponding auxiliary code λ_O built over an auxiliary set and a mapping. An ingenious technique of partitioning (binning) λ_O into $\mathcal{M}$ bins, one for each message $m \in \mathcal{M}$, is the key feature of the coding technique. In the following, we consider PTP-STx codes which are endowed with a coset code structure. Note that if the auxiliary set is a finite field, then one can visualize λ_O and/or λ_I possessing certain algebraic closure properties. For example, λ_O could be coset of a linear code, or the bins of λ_O could be cosets of sub-linear code λ_I. In the sequel, we characterize PTP-STx codes possessing these algebraic closure properties.

3.4.1 Nested coset PTP-STx codes

We begin with a brief review of coset and nested coset codes. An (n, k) coset code is a collection of vectors in $\mathcal{F}_q^n$ obtained by adding a bias vector to a $k-$dimensional subspace of $\mathcal{F}_q^n$. If $\lambda_O \subseteq \mathcal{F}_q^n$ and $\lambda_I \subseteq \lambda_O$ are $(n, k+l)$ and (n, k) coset codes respectively, then q^l cosets λ_O/λ_I that partition λ_O is a nested coset code. We refer to this as nested coset code $(n, k, l, g_I, g_{O/I}, b^n)$ where b^n is the bias vector, $g_I \in \mathcal{F}_q^{k\times n}$ and $g_O^T = \begin{bmatrix} g_I^T & g_{O/I}^T \end{bmatrix} \in \mathcal{F}_q^{(k+l)\times n}$ are generator matrices of λ_I and λ_O respectively.

The structure of a nested coset PTP-STx code must now be apparent to an informed reader. The bins are cosets of the smaller linear code λ_I. The entire collection of bins forms a coset of the larger linear code λ_O. The message to be sent to the decoder indexes the bins. For this nested coset code, we let $v^n(a^k, m^l) := a^k g_I \oplus m^l g_{O/I} \oplus b^n$ denote a generic codeword in coset $c(m^l) := \left\{v^n(a^k, m^l) \in \mathcal{F}_q^n : a^k \in \mathcal{F}_q^k\right\}$. We refer to $c(m^l)$ as the coset corresponding to message m^l. The following is therefore a natural characterization of a nested coset PTP-STx code.

Definition 3.4.1 *A PTP-STx code $(n, \mathscr{M}, e, d)$ is a nested coset PTP-STx code over $\mathcal{F}_q$ if there exists (i) a nested coset code $\left(n, k, l, g_I, g_{O/I}, b^n\right)$ over $\mathcal{F}_q$, ii) map $f : \mathcal{F}_q \times \mathcal{S} \to \mathcal{X}$ and, (iii) a $1:1$ onto map $\iota : \mathcal{M} \to \mathcal{F}_q^l$ such that $e(m, s^n) \in \left\{f^n\left(a^k g_I \oplus \iota(m) g_{O/I} \oplus b^n, s^n\right) : a^k \in \mathcal{F}_q^k\right\}$ for every $m \in \mathcal{M}$.*

3.4.2 Partitioned coset PTP-STx codes

Let us now describe partitioned coset codes and define partitioned coset PTP-STx codes. As mentioned earlier, an (n,k) coset code $\lambda_O \subseteq \mathcal{F}_q^n$ is a collection of vectors obtained by adding a bias vector to a $k-$dimensional subspace of $\mathcal{F}_q^n$. The coset code λ_O is completely specified through it's generator matrix $g \in \mathcal{F}_q^{k\times n}$ and bias vector $b^n \in \mathcal{F}_q^n$. Consider a partition of λ_O into q^l bins. Each codeword $v^n(a^k) := a^k g \oplus b^n$ is assigned an index $i(a^k) \in \mathcal{F}_q^l$. This coset code λ_O with it's partitions is called a *partitioned coset code* and denoted (n,k,l,g,b^n,i). For each $m^l \in \mathcal{F}_q^l$, let $c(m^l) := \left\{a^k \in \mathcal{F}_q^k : i(a^k) = m^l\right\}$ denote the indices of codewords in the m^l bin.

The structure of a partitioned coset code forms the essential building block for the coding techniques proposed in chapters 4, 5. We therefore formalize the same through the following definition for easy reference.

Definition 3.4.2 *Recall that a coset code $\lambda \subseteq \mathcal{F}_\pi^n$ is a coset of a linear code $\overline{\lambda} \subseteq \mathcal{F}_\pi^n$. The coset code is completely specified by the generator matrix $g \in \mathcal{F}_\pi^{k\times n}$ and a bias vector $b_j^n \in \mathcal{F}_\pi^n$. Consider a partition of λ into π^l bins. Each codeword $a^k g \oplus b^n$ is assigned an index $i(a^k) \in [\pi^l]$. This coset code λ with it's partitions is referred to as partitioned coset code (PCC) (n,k,l,g,b^n,i) or succinctly as an (n,k,l) PCC. For each $m \in [\pi^l]$, let $c(m) := \left\{a^k \in \mathcal{F}_\pi^k : i(a^k) = m\right\}$.*

The reader will now be able to visualize the structure of a partitioned coset PTP-STx code. The auxiliary code is obtained by partitioning a coset code $\lambda_O \in \mathcal{F}_q^n$ into q^l bins. The following characterization makes this precise.

Definition 3.4.3 *A PTP-STx code $(n,\mathscr{M},e,d)$ is a partitioned coset PTP-STx code over $\mathcal{F}_q$ if there exists (i) a partitioned coset code (n,k,l,g,b^n,i) over $\mathcal{F}_q$, ii) a map $f : \mathcal{F}_q \times \mathcal{S} \to \mathcal{X}$ and, (iii) a $1:1$ onto map $\iota : \mathcal{M} \to \mathcal{F}_q^l$ such that $e(m,s^n) \in \left\{f^n\left(a^k g \oplus b^n, s^n\right) : i(a^k) = \iota(m)\right\}$ for every $m \in \mathcal{M}$.*

3.4.3 Union coset PTP-STx codes

Consider a linear code $\overline{\lambda}_I \subseteq \mathcal{F}_q^n$ with generator matrix $g \in \mathcal{F}_q^{k\times n}$. For each $m^l \in \mathcal{F}_q^l$, let $b^n(m^l) \in \mathcal{F}_q^n$. The union of q^l cosets of $\overline{\lambda}_I$ corresponding to each of the shifts $b^n(m^l) : m^l \in \mathcal{F}_q^l$ is termed a union coset code. Letting $\underline{b}^n := (b^n(m^l) : m^l \in \mathcal{F}_q^l)$, a union coset code is completely specified by the generator matrix $g \in \mathcal{F}_q^{k\times n}$ and $\underline{b}^n$. In particular, the union coset code is the union of cosets $(a^k g_I \oplus b^n(m^l) : a^k \in \mathcal{F}_q^k)$ corresponding to each of the shifts $\underline{b}^n$. We refer to this as the $(n,k,l,g,\underline{b}^n)$ union coset code. Following is a natural characterization of a union coset PTP-STx code.

Definition 3.4.4 *A PTP-STx code $(n,\mathscr{M},e,d)$ is a union coset PTP-STx code over $\mathcal{F}_q$ if there exists (i) a union coset code $(n,k,l,g,\underline{b}^n)$ over $\mathcal{F}_q$, ii) a map $f : \mathcal{F}_q \times \mathcal{S} \to \mathcal{X}$ and, (iii) a $1:1$ onto map $\iota : \mathcal{M} \to \mathcal{F}_q^l$ such that $e(m,s^n) \in \left\{f^n\left(a^k g \oplus b^n(\iota(m)), s^n\right) : a^k \in \mathcal{F}_q^k\right\}$ for every $m \in \mathcal{M}$.*

Before we conclude this section, we make a simple observation. Note that an $(n,k,l,g_I,g_{O/I},b^n)$ nested coset code is (i) a $(n,k+l,l,g_O,b^n,i)$ partitioned coset code where $i(a^{k+l}) = a_{k+1}a_{k+2}\cdots a_{k+l}$ and (ii) a $(n,k,l,g_I,\underline{b})$

union coset code where $b^n(m^l) = m^l g_{O/I} \oplus b^n$. If we therefore prove nested coset PTP-STx codes achieve capacity of an arbitrary PTP-STx, we can conclude that all the above ensembles of coset PTP-STx codes - nested, union and partitioned - achieve capacity of an arbitrary PTP-STx. The following section is dedicated to proving nested coset PTP-STx codes achieve capacity of an arbitrary PTP-STx.

3.5 Coset codes achieve capacity of arbitrary PTP-STx

We now state and prove our first main finding - nested coset PTP-STx codes achieve $\mathbb{C}(\tau)$.

Theorem 3.5.1 *For a PTP-STx $(\mathcal{S}, W_S, \mathcal{X}, \kappa, \mathcal{Y}, W_{Y|XS})$, if $R \in \mathbb{C}(\tau)$, i.e., R is achievable, then there exists a sequence $(n, \mathscr{M}^{(n)}, e^{(n)}, d^{(n)}) : n \geq 1$ of nested coset PTP-STx codes over $\mathcal{F}_q$ that achieves (R, τ), where $q = \pi(\min\{(|\mathcal{X}| \cdot |\mathcal{S}|)^2, (|\mathcal{X}| + |\mathcal{S}| + |\mathcal{Y}| - 2) \cdot |\mathcal{X}| \cdot |\mathcal{S}|\})$.* □

Proof: Consider any pmf $p_{VXSY} \in \mathbb{D}(\tau)$ and $\eta > 0$. We prove the existence of a nested coset PTP-STx code $(n, \mathscr{M}^{(n)}, e^{(n)}, d^{(n)})$ of rate $\frac{\log \mathscr{M}^{(n)}}{n} \geq I(V;Y) - I(V;S) - \eta$, average cost $\tau(e^{(n)}) \leq \tau + \eta$ and average probability of error $\overline{\xi}(e^{(n)}, d^n) \leq \eta$ for every $n \in \mathbb{N}$ sufficiently large. The underlying finite field is of cardinality $\pi(\min\{(|\mathcal{X}| \cdot |\mathcal{S}|)^2, (|\mathcal{X}| + |\mathcal{S}| + |\mathcal{Y}| - 2) \cdot |\mathcal{X}| \cdot |\mathcal{S}|\})$ referred to as π for short.

We prove the existence by averaging the error probability over a specific ensemble of nested coset PTP-STx codes. We begin with a description of a generic code in this ensemble.

Consider a nested coset PTP-STx code $(n, k, l, g_I, g_{O/I}, b^n)$, denoted λ_O/λ_I with parameters

$$k := \lceil n\left(1 - \frac{H(V|S)}{\log \pi} + \frac{\eta}{8 \log \pi}\right)\rceil \tag{3.1}$$

$$l := \lfloor n\left(1 - \frac{H(V|Y)}{\log \pi} - \frac{\eta}{8 \log \pi}\right)\rfloor - k. \tag{3.2}$$

The reader is advised to bear in mind our notation is not reflective of k and l being functions of n. This abuse of notation reduces clutter. We specify encoding and decoding rules that map λ_O/λ_I into a corresponding nested coset PTP-STx code.

The encoder is provided with nested coset code λ_O/λ_I. The message is used to index one among π^l cosets of λ_O/λ_I. For simplicity, we assume that the set of messages $\mathcal{M}$ is $\mathcal{V}^l$, and $M^l \in \mathcal{V}^l$ to be the uniformly distributed random variable representing user's message. The encoder observes the state sequence S^n and populates the list $L(M^l, S^n) = \left\{v(a^k, M^l) : (v(a^k, M^l), S^n) \in T_{\frac{\delta}{2}}(V, S), a^k \in \mathcal{F}_q^k\right\}$ of codewords in the coset corresponding to the message that are jointly typical with the state sequence, where $\delta := \frac{1}{2}\min\left\{\frac{\eta}{48}, \frac{\eta \log(|\mathcal{V}||\mathcal{X}||\mathcal{S}||\mathcal{Y}|)}{\kappa_{\max}}\right\}$, $\kappa_{\max} := \max\{\kappa(x,s) : (x,s) \in \mathcal{X} \times \mathcal{S}\}$. If $L(M^l, S^n)$ is empty, it picks a codeword uniformly at random from coset $c(M^l)$. Otherwise, it picks a codeword uniformly at random from $L(M^l, S^n)$. Let $V(A^k, M^l)$ denote the picked codeword in either case. The encoder computes $X^n(M^l, S^n) := f^n(V^n(A^k, M^l), S^n)$, where $f : \mathcal{V} \times \mathcal{S} \to \mathcal{X}$ is any

map that satisfies $p_{X|VS}(f(v,s)|v,s)=1$ for all pairs $(v,s)\in\mathcal{V}\times\mathcal{S}$. $X^n(M^l,S^n)$ is fed as input to the channel.

The decoder observes the received vector Y^n and populates the list

$$D(Y^n):=\left\{m^l\in\mathcal{V}^l:\exists v^n(a^k,m^l) \text{ such that } (v^n(a^k,m^l),Y^n)\in T_\delta(V,Y)\right\}.$$

If $D(Y^n)$ is a singleton, the decoder declares the content of $D(Y^n)$ as the decoded message pair. Otherwise, it declares an error.

The above encoding and decoding rules map λ_O/λ_I into a corresponding nested coset PTP-STx code $(n,\mathscr{M}^n,e^{(n)},d^{(n)})$ of rate $\frac{\log\mathscr{M}^{(n)}}{n}=\frac{l\log\pi}{n}$. Observe that, for $n\geq N_1(\eta):=\lceil\frac{8\log\pi}{\eta}\rceil$, we have

$$\begin{aligned} n\left(1-\frac{H(V|S)}{\log\pi}+\frac{\eta}{8\log\pi}\right)\leq k &\leq n\left(1-\frac{H(V|S)}{\log\pi}+\frac{\eta}{8\log\pi}\right)+1 &(3.3)\\ &\leq n\left(1-\frac{H(V|S)}{\log\pi}+\frac{\eta}{4\log\pi}\right), &(3.4)\end{aligned}$$

and similarly,

$$\begin{aligned} n\left(1-\frac{H(V|Y)}{\log\pi}-\frac{\eta}{8\log\pi}\right)\geq k+l &\geq n\left(1-\frac{H(V|Y)}{\log\pi}-\frac{\eta}{8\log\pi}\right)-1 &(3.5)\\ &\geq n\left(1-\frac{H(V|Y)}{\log\pi}-\frac{\eta}{4\log\pi}\right). &(3.6)\end{aligned}$$

Combining the upper bound for k in (3.4) and the lower bound for $k+l$ in (3.6), we get

$$\frac{l\log\pi}{n}\geq H(V|S)-H(V|Y)-\frac{\eta}{2}=I(V;Y)-I(V;S)-\frac{\eta}{2}. \tag{3.7}$$

Since λ_O/λ_I was a generic nested coset code satisfying (3.1), (3.2), we have characterized, through our encoding and decoding maps, an ensemble of nested coset PTP-STx codes, one for each $n\in\mathbb{N}$, $n\geq N_1(\eta)$ of rate at least $I(V;Y)-I(V;S)-\frac{\eta}{2}$. It suffices to prove existence of a PTP-STx code $(n,\mathscr{M}^{(n)},e^{(n)},d^{(n)})$ in this ensemble, one for each $n\in\mathbb{N}$ sufficiently large, with average probability of error $\xi(e^{(n)},d^{(n)})\leq\eta$ and average cost constraint $\tau(e^{(n)})\leq\tau+\eta$. This is done by averaging $\xi(e^{(n)},d^{(n)})$ over the ensemble.

Consider a random nested coset code $(n,k,l,G_I,G_{O/I},B^n)$, denoted Λ_O/Λ_I, with parameters n,k,l satisfying (3.1) and (3.2). Let $G_I\in\mathcal{V}^{k\times n}$, $G_{O/I}\in\mathcal{V}^{l\times n}$ and bias vector $B^n\in\mathcal{V}^n$ be mutually independent and uniformly distributed on their respective range spaces. In the sequel, we study the average probability of error $\xi(e^{(n)},d^{(n)})$ of the corresponding random nested coset PTP-STx code. Towards this end, we begin with a few remarks on notation. Let $V^n(a^k,m^l):=a^kG_I\oplus m^lG_{O/I}\oplus B^n$ denote a generic codeword in coset $C(m^l):=\left\{V^n(a^k,m^l):a^k\in\mathcal{V}^k\right\}$ corresponding to message m^l.

In order to study $\xi(e^{(n)}, d^{(n)})$, we need to characterize the error events associated with the random nested coset PTP-STx code corresponding to Λ_O/Λ_I. If $\epsilon_1 := \{S^n \notin T_{\frac{\delta}{4}}(S)\}$, $\epsilon_2 := \{\phi_{\frac{\delta}{2}}(S^n, M^l) = 0\}$, where $\phi_{\frac{\delta}{2}}(s^n, m^l) := \sum_{a^k \in \mathcal{V}^k} 1_{\{(V^n(a^k, m^l), s^n) \in T^n_{\frac{\delta}{2}}(VS)\}}$, then the error event at the encoder is contained in $\epsilon_1 \cup \epsilon_2$. The error event at the decoder is contained in $\epsilon_3^c \cup \epsilon_4$, where $\epsilon_3 := \cup_{a^k \in \mathcal{V}^k} \{(V^n(a^k, M^l), Y^n) \in T^n_\delta(V, Y)\}$ and $\epsilon_4 := \cup_{\hat{m}^l \neq M^l} \cup_{a^k \in \mathcal{V}^k} \left\{\left(V^n(a^k, \hat{m}^l), Y^n\right) \in T^n_\delta(V, Y)\right\}$. It suffices to derive an upper bound on $P(\epsilon_1) + P(\epsilon_1^c \cap \epsilon_2) + P((\epsilon_1 \cup \epsilon_2)^c \cap \epsilon_3^c) + P(\epsilon_4)$. In the sequel, we derive an upper bound on each term of the above sum.

Lemma 2.3.1 guarantees the existence of $N_2(\eta) \in \mathbb{N}$[4] such that $\forall n \geq N_2(\eta)$, $P(\epsilon_1) \leq \frac{\eta}{16}$. In appendix A, we prove the existence of $N_3(\eta) \in \mathbb{N}$, such that $\forall n \geq N_3(\eta)$,

$$P(\epsilon_1^c \cap \epsilon_2) \leq \exp\left\{-n \log \pi \left(\frac{k}{n} - \left(1 - \frac{H(V|S)}{\log \pi} + \frac{3\delta}{4 \log \pi}\right)\right)\right\}. \tag{3.8}$$

Substituting the lower bound in (3.3) for k in (3.8), for all $n \geq \max\{N_1(\eta), N_3(\eta)\}$, we have

$$P(\epsilon_1^c \cap \epsilon_2) \leq \exp\left\{-n\left(\frac{\eta}{8} - \frac{3\delta}{4}\right)\right\} \leq \exp\left\{-n\left(\frac{7\eta}{64}\right)\right\}, \tag{3.9}$$

where the last inequality follows from the choice of δ.

We now consider $P((\epsilon_1 \cup \epsilon_2)^c \cap \epsilon_3^c)$. An informed reader will recognize that an upper bound on this term can be derived using a typical application of conditional frequency typicality lemma 2.4.1. For the sake of completeness we state the arguments. The encoding rule ensures, $(\epsilon_1 \cup \epsilon_2)^c \subseteq \{(V^n(M^l, S^n), S^n) \in T^n_{\frac{\delta}{2}}(V, S)\}$, and thus

$$\begin{aligned}
P((\epsilon_1 \cup \epsilon_2)^c \cap \epsilon_3^c) &\leq P\left(\left\{(V^n(M^l, S^n), S^n) \in T^n_{\frac{\delta}{4}}(V, S)\right\} \cap \epsilon_3^c\right) \\
&\leq \sum_{(v^n, s^n) \in T^n_{\frac{\delta}{2}}(V,S)} P((V^n(M^l, S^n), S^n) = (v^n, s^n)) P\left(\epsilon_3^c | (V^n(M^l, S^n), S^n) = (v^n, s^n)\right) \\
&\leq \sum_{(v^n, s^n) \in T^n_{\frac{\delta}{2}}(V,S)} P((V^n(M^l, S^n), S^n) = (v^n, s^n)) P\left(Y^n \notin T_\delta(Y|v^n, s^n) | (V^n(M^l, S^n), S^n) = (v^n, s^n)\right).
\end{aligned} \tag{3.10}$$

For any $(v^n, s^n) \in T^n_{\frac{\delta}{2}}(V, S)$, note that,

$$P\left(\begin{smallmatrix} Y^n = y^n, \\ X^n(M^l, S^n) = x^n \end{smallmatrix} | \begin{smallmatrix} (V^n(M^l, S^n), S^n) \\ = (v^n, s^n) \end{smallmatrix}\right) = \prod_{i=1}^n P\left(X_i = x_i, Y_i = y_i | V_i = v_i, S_i = s_i\right)$$

where the second equality follows from Markov chain $V - (X, S) - Y$. By lemma 2.4.1, there exists $N_4(\eta) \in \mathbb{N}$ such that for all $n \geq N_4(\eta)$

$$P((Y^n, X^n(M^l, S^n)) \notin T^n_\delta(X, Y|v^n, s^n) | (V^n(M^l, S^n), S^n) = (v^n, s^n)) \leq \frac{\eta}{8}. \tag{3.11}$$

[4]Since δ is a function of η, the dependence of $N_2(\eta)$ on δ is captured through η.

Substituting (3.11) in (3.10), we have $P((\epsilon_1 \cup \epsilon_2)^c \cap \epsilon_3) \leq \frac{\eta}{8}$ for all $n \geq N_4(\eta)$. It remains to provide an upper bound on $P(\epsilon_4)$. In appendix B, we prove the existence of $N_5(\eta) \in \mathbb{N}$ such that $\forall n \geq N_5(\eta)$, $P((\epsilon_1 \cup \epsilon_2 \cup \epsilon_2)^c \cap \epsilon_4) \leq \exp\left\{-n \log \pi \left(1 - \frac{H(V|Y)}{\log \pi} - \frac{3\delta}{2 \log \pi} - \frac{k+l}{n}\right)\right\}$. For $n \geq \max\{N_1(\eta), N_5(\eta)\}$, the upper bound for $k+l$ derived in (3.6) is substituted to yield, $P((\epsilon_1 \cup \epsilon_2 \cup \epsilon_2)^c \cap \epsilon_4) \leq \exp\left\{-n\left(\frac{\eta}{8} - \frac{3\delta}{2}\right)\right\} \leq \exp\left\{-n\left(\frac{3\eta}{32}\right)\right\}$.

We have therefore proved that for every $n \geq \max\{N_i(\eta) : i \in [5]\}$, there exists at least one nested coset PTP-STx code (n, π^l, e, d) over $\mathcal{F}_\pi$ for which $\bar{\xi}(e,d) \leq \frac{\eta}{8} + \exp\left\{-n\frac{7\eta}{64}\right\} + \frac{\eta}{8} + \exp\left\{-n\frac{3\eta}{32}\right\}$. For $n \geq \max\{N_i(\eta) : i \in [6]\}$, where $N_6(\eta) = \lceil \frac{32}{3\eta} \log \frac{8}{\eta} \rceil$, $\bar{\xi}(e,d) \leq \frac{\eta}{2}$. It only remains to prove this code satisfies the average cost constraint. It can be verified that $\tau(e) \leq \frac{\eta}{2}\kappa_{\max} + (1 - \frac{\eta}{2})(\tau + \frac{\delta \kappa_{\max}}{2\log(|\mathcal{X}||\mathcal{S}|)})$. The choice of δ ensures that $\tau(e) \leq \frac{\eta}{2}\kappa_{\max} + (\tau + \frac{\eta}{2})$. Since $\kappa_{\max} \in \mathbb{R}$ is bounded, this proves the existence of a sequence $(n, \pi^{l(n)}, e^{(n)}, d^{(n)}) : n \geq 1$ of nested coset PTP-STx codes that achieve (R, τ) for every $R \in \mathbb{C}(\tau)$. ■

The codewords of Λ_O being uniformly distributed over $\mathcal{F}_\pi^n$ (c.f. Lemma A.0.1(i)), the probability of it being jointly typical with a typical state sequence s^n is $\frac{|T_\delta(U|S)|}{\pi^n} = \exp\{n(H(U|S) - \log \pi)\}$. This indicates that each coset must contain roughly $\frac{q^n}{|T_\delta(U|S)|} = \frac{q^n}{q^{n(H(U|S))}} = q^{n(\log \pi - H(U|S))}$ codewords. Indeed, it suffices to partition Λ_O with a coset of rate $\frac{k}{n} > 1 - \frac{H(U|S)}{\log \pi}$. $1 - \frac{H(U|S)}{\log \pi}$ being in general larger than $\frac{I(U;S)}{\log \pi}$, we conclude that the constraint of linearity forces us to increase the rate of the binning code.

However, the sparsity of typical vectors in a random linear code comes to our rescue when we attempt to pack cosets. The decoder looks for all vectors in the auxiliary code that are jointly typical with the received vector Y^n. In unstructured random coding, since each codeword is individually typical with high probability, the rate of auxiliary code is bounded from above by $\frac{I(U;Y)}{\log \pi}$. The typical vectors being sparse in random linear code, a similar argument as above enables us to enlarge the auxiliary code to a rate $1 - \frac{H(U|Y)}{\log \pi}$. The rate of the code is thus $(1 - \frac{H(U|Y)}{\log \pi}) - (1 - \frac{H(U|S)}{\log \pi}) = \frac{I(U;Y) - I(U;S)}{\log \pi}$.

We have thus proved nested coset codes achieve the capacity of arbitrary PTP-STx. The interested reader is referred to [33] wherein nested lattice codes are proved to achieve capacity of arbitrary continuous point to point channels. In order to achieve capacity of arbitrary continuous PTP-STx, it is necessary to construct lattices which result in non-uniform distribution of error when employed for source quantization.

The following corollaries are a direct consequence of nested coset PTP-STx codes being both partitioned coset and union coset PTP-STx codes.

Corollary 3.5.2 *For a PTP-STx $(\mathcal{S}, W_S, \mathcal{X}, \kappa, \mathcal{Y}, W_{Y|XS})$, if $R \in \mathbb{C}(\tau)$, i.e., R is achievable, then there exists a sequence $(n, \mathscr{M}^{(n)}, e^{(n)}, d^{(n)}) : n \geq 1$ of partitioned coset PTP-STx codes over $\mathcal{F}_q$ that achieves (R, τ), where $q = \pi(\min\{(|\mathcal{X}| \cdot |\mathcal{S}|)^2, (|\mathcal{X}| + |\mathcal{S}| + |\mathcal{Y}| - 2) \cdot |\mathcal{X}| \cdot |\mathcal{S}|\})$.*

Corollary 3.5.3 *For a PTP-STx $(\mathcal{S}, W_S, \mathcal{X}, \kappa, \mathcal{Y}, W_{Y|XS})$, if $R \in \mathbb{C}(\tau)$, i.e., R is achievable, then there exists a sequence $(n, \mathscr{M}^{(n)}, e^{(n)}, d^{(n)}) : n \geq 1$ of union coset PTP-STx codes over $\mathcal{F}_q$ that achieves (R, τ), where $q = \pi(\min\{(|\mathcal{X}| \cdot |\mathcal{S}|)^2, (|\mathcal{X}| + |\mathcal{S}| + |\mathcal{Y}| - 2) \cdot |\mathcal{X}| \cdot |\mathcal{S}|\})$.*

Chapter 4

Three user interference channel

We begin with a brief description of a three user interference channel (3−IC) and state the problem of interest. A 3−IC, depicted in figure 4.3, consists of three transmitter receiver (Tx-Rx) pairs that share a common communication medium. Let $\mathcal{X}_j$ denote the input alphabet available to transmitter j. Receiver j observes symbols in output alphabet $\mathcal{Y}_j$. The symbol observed by receiver j depends on the input of the *three* transmitters. This is modelled through the channel transition probabilities $W_{Y_1Y_2Y_3|X_1X_2X_3}$. In particular, conditioned on x_1, x_2, x_3 being the symbols input by transmitters $1, 2$ and 3 respectively, the probability of receivers $1, 2$ and 3 observing symbols y_1, y_2, y_3 respectively, is $W_{Y_1Y_2Y_3|X_1X_2X_3}(y_1, y_2, y_3|x_1, x_2, x_3)$. As always, we assume the channel to be discrete, memoryless and used without feedback.

Transmitter j wishes to reliably communicate a specific information stream to it's corresponding receiver j. The problem of interest is to characterize the capacity region of a 3−IC. Please refer to section 4.3 for a precise statement of this problem. The main contributions of this chapter are (i) characterization of a new achievable rate region for a general discrete 3−IC and (ii) identification of 3−ICs for which the proposed achievable rate region strictly enlarges upon the current known largest. In the following, we provide a discussion of current known coding techniques and the key elements of our contribution.

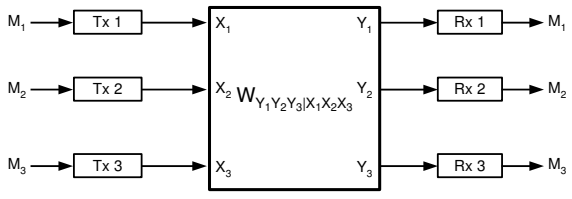

Figure 4.1: Three user interference channel (3−IC)

Evidently, an interference channel (IC) is a model for communication between multiple transmitter receiver (Tx-Rx) pairs that share a common communication medium. Since the Tx-Rx pairs share a common communication medium, every user's transmission causes interference to every other user. Communication over an IC is therefore facilitated by a coding technique that manages interference efficiently, in addition to channel noise.

The quest for designing an efficient coding technique for managing interference was initiated in the context of an IC with two Tx-Rx pairs [34] [35] [36], henceforth referred to as $2-$IC. Over a $2-$IC, the source of interference is the transmission of the *lone* interfering transmitter. Based on his findings in [37], Carleial proposed the technique of each receiver decoding a part of the interfering transmitter's transmission. To enable this, Carleial employed superposition coding [20] [21]. Each transmitter splits it's message and transmission into two parts - public and private. Cloud center and satellite codebooks encode the public and private parts of the message respectively. In addition to both parts of the corresponding transmitter, each receiver decodes the public part, i.e., the cloud center codeword, of the interfering transmitter.

In characterizing the performance of his coding technique via random coding, Carleial employed, quite naturally, random unstructured codebooks for each pair of cloud center and satellite codebooks. Moreover the two pairs were statistically independent. Subsequently, Han and Kobayashi [13] strictly enlarged Carleial's achievable rate region by (i) replacing the successive decoder he employed by a more powerful joint decoder, and (ii) incorporating a time sharing random variable.

The above coding technique of message splitting via superposition coding and employing unstructured cloud and satellite codebooks, henceforth referred to as CHK-technique, remains to be the best known coding technique for communication over a $2-$IC. The interfering transmitter's transmission being the only source of interference, decoding a part of the same amounts to decoding a part of the interference. This coding technique is in general more efficient than either ignoring or decoding the entire interference. Moreover, superposition coding using unstructured codes enables efficient decoding of a part of the interfering transmitter's transmission [21]. Whether the rate region proved achievable in [13], henceforth referred to as the CHK rate region, is the capacity region of a $2-$IC has remained a long standing open problem in information theory.

In this chapter, we consider the problem of communicating over a $3-$IC. In a $3-$IC, transmission by two transmitters contribute to interference. The nature of interference over a $3-$IC being richer, we develop a technique based on *coset codes* for interference management. Coset codes built over finite fields, as introduced in section 1.4, are algebraically closed. The sum of any two codewords of a coset lies in another coset. Moreover, two cosets of a linear code, when added result in another coset. As against to adding two random codebooks whose codewords are statistically independent, we emphasize that the sum of two random cosets of a random linear code yields a collection of the same size. This property of coset codes behaving nicely under addition - a bivariate operation - is exploited for managing interference, wherein, interference over a $3-$IC is in general a compressive bivariate function of the transmissions of the two interfering transmitters.

The use of lattice codes [19], [22] and interference alignment techniques [17] have been proposed for efficient interference management over Gaussian IC's with three or more Tx-Rx pairs. While these works are restricted to *additive* IC's, the key contribution herein is the development of a framework based on coset codes for efficient communication over an *arbitrary* discrete 3−IC. The framework involves (i) a new ensemble of coset codes - *partitioned coset codes* (PCC) - possessing algebraic and empirical properties, (coupled with) (ii) efficient joint typicality based encoding and decoding rules that exploit algebraic properties of PCC and moreover, enable us achieve rates corresponding to *arbitrary* single-letter distributions, (iii) mathematical tools and proof techniques to characterize the performance of the proposed coding technique over arbitrary 3−ICs. This framework enables us characterize *PCC rate region* - a new achievable rate region for an arbitrary discrete 3−IC. We demonstrate the utility of this framework by identifying additive as well as non-additive 3−IC's for which the proposed technique enables efficient communication.

Conventionally, the random codebooks employed in characterizing achievable rate regions are unstructured and independent, i.e., codewords of each random codebook, and the random codebooks themselves, are statistically independent. Since our findings are based on a fundamentally different philosophy - use of statistically correlated codes possessing algebraic closure properties - it is natural to enquire the need for the same. Indeed, one can employ **u**nstructured codes for communication over an arbitrary 3−IC and optimally stitch together all current known relevant coding techniques - message splitting, **b**inning and **s**uperposition - to derive the current known largest achievable rate region for communication over an arbitrary 3−IC. How does this rate region, henceforth referred to as $\mathscr{USB}-$region, compare to the PCC rate region?

An important element of our findings is the strict sub-optimality of the $\mathscr{USB}-$technique[1] for communicating over 3−IC's, including non-additive instances. In particular, we identify (i) an additive 3−IC, and (ii) a non-additive 3−IC for which we *analytically* prove strict containment of the $\mathscr{USB}-$region in it's corresponding capacity region. Moreover, for these 3−IC's the PCC rate region is the capacity region. This justifies the need for the framework developed herein. The reader will now wonder whether PCC rate region strictly subsumes $\mathscr{USB}-$region for an arbitrary 3−IC.[2]

In addition to efficiently decoding a bivariate function of the two interfering transmitters' transmission, which the proposed coding technique based on PCC accomplishes, it is necessary to enable receivers efficiently decode individual parts of interfering transmitters' transmissions. The coding technique based on statistically correlated PCC proposed herein, is tuned to exploit the algebraic properties of coset codes in decoding a bivariate function - field addition - of transmissions of the two interfering transmitters. Such a technique is strictly sub-optimal for the purpose of

[1]The above coding technique that employs **u**nstructured codes and optimally stitches together all current known relevant coding techniques - message splitting, **b**inning and **s**uperposition is the current known best coding technique for communicating over an arbitrary 3−IC. We refer to this as the $\mathscr{USB}-$technique. We state the $\mathscr{USB}-$technique in section 4.4.2. This yields the current known largest achievable rate region for a general 3−IC which is referred to herein as $\mathscr{USB}-$region. We provide a characterization of the $\mathscr{USB}-$region for a sub-class of 3−IC's in section 4.4.2.

[2]A little thought will convince an alert reader, that is this were true, the PCC rate region should particularize or enlarge the CHK rate region for a 2−IC. Indeed, this is not true, as will be indicated in the sequel.

decoding individual parts of interfering transmitters' transmissions, when compared to traditional technique based on unstructured independent codes. This leads us to enhance the PCC coding technique by incorporating the $\mathscr{USB}-$technique. This enables us characterize a new achievable rate region for an arbitrary discrete $3-$IC that contains PCC rate region and strictly enlarges the $\mathscr{USB}-$region.

4.1 Outline

We state the preliminaries - notation, definitions and the precise statement of the problem - in section 4.3. In section 4.4, we provide a characterization of the CHK rate region for a $2-$IC. The first main finding of this chapter is the strict sub-optimality of current known coding techniques based on unstructured codes for communication over $3-$IC. In order to present this finding, we characterize a sub-class of $3-$IC's called $3-$to-1 IC (section 4.3), and derive, in section 4.4.2, an achievable rate region for the same, called $\mathscr{USB}-$region, that employs current known coding techniques based on unstructured codes. In section 4.5, we identify an additive $3-$to-1 IC and propose a strategy based on correlated linear codes that is analytically proven to strictly outperform $\mathscr{USB}-$technique.

Our second main finding - a new achievable rate region for an arbitrary discrete $3-$IC - is presented in section 4.6 in three pedagogical steps. In section 4.6.1, we define partitioned coset codes (PCC) and present the first step that describes all the new elements of our framework in a simple setting. Here, we employ PCC to manage interference seen by only one receiver. For this step, we furnish a complete and elaborate proof of achievability. In this section, we also identify a non-additive $3-$to-1 IC (Example 4.6.7) for which $\mathscr{USB}-$technique is strictly sub-optimal and moreover, the coding technique based on PCC is capacity achieving. This example illustrates the central theme of this thesis - codes endowed with algebraic closure properties enable efficient communication over arbitrary general multi-terminal systems, not just additive, symmetric instances - and thereby justifies the framework developed herein. In the second step, presented in section 4.6.2, we employ PCC to manage interference seen by all three receivers. Finally, in section 4.6.3, we indicate how to enlarge the $\mathscr{USB}-$region by incorporating the framework based on PCC.

4.2 Prior work

An IC has been the subject of considerable interest since Shannon's study [34] of the two way channel. Carleial [38] made the key observation that the technique of superposition [20], [21] could be employed to split each user's transmission and thereby enable each receiver decode a part of the interfering transmitter's transmission. While Carleial derived a rate region by analyzing a sequential decoder, Han and Kobayashi [13] employed the joint decoder to enlarge upon the rate region proved achievable in [38].[3] The technique developed by Carleial [38], and furthered

[3]Moreover, they included a time sharing random variable in it's characterization and argue that a time sharing random variable provides a strict enlargement over the then common practice of convex hull operation

by Han and Kobayashi [13] is the only coding technique known to counter interference in a 2−IC. This coding technique is optimal under strong interference [37], [39]. El Gamal and Costa [40] prove that CHK-technique is optimal for a class of deterministic IC's. Recently Etkin, Tse and Wang [41] prove the CHK-technique is within 1 bit of the optimal for the Gaussian IC. Following a period of reduced activity, there has been renewed interest in developing strategies for managing interference in an IC setting. Cadambe and Jafar [17] propose the technique of interference alignment for the Gaussian IC and thereby harness the available of degrees of freedom in an IC with several Tx-Rx pairs more efficiently. Bresler, Parekh and Tse [22] employ lattice codes to align interference and thereby characterize capacity of Gaussian ICs within a constant number of bits. While our findings appear similar to the idea of interference alignment, we would like to reiterate the following key elements. Our work provides a technique of aligning interference over arbitrary channels even while achieving rates corresponding to non-uniform distributions.[4] Example 4.6.7 illustrates the utility of this technique.

4.3 Definitions: 3−IC, 3−to−1 IC, achievability, capacity region

A 3−IC consists of three finite input alphabet sets $\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3$ and three finite output alphabet sets $\mathcal{Y}_1, \mathcal{Y}_2, \mathcal{Y}_3$. The discrete time channel is (i) time invariant, i.e., the pmf of $\underline{Y}_t := (Y_{1t}, Y_{2t}, Y_{3t})$, the output at time t, conditioned on $\underline{X}_t := (X_{1t}, X_{2t}, X_{3t})$, the input at time t, is invariant with t, (ii) memoryless, i.e., conditioned on present input $\underline{X}_t$, the present output $\underline{Y}_t$ is independent of past inputs $\underline{X}_1, \cdots, \underline{X}_{t-1}$, past outputs $\underline{Y}_1, \cdots, \underline{Y}_{t-1}$ and (iii) used without feedback, i.e., encoders have no information of the symbols received by decoders. Let $W_{\underline{Y}|\underline{X}}(\underline{y}|\underline{x}) = W_{Y_1Y_2Y_3|X_1X_2X_3}(y_1, y_2, y_3|x_1, x_2, x_2)$ denote probability of observing symbol $y_j \in \mathcal{Y}_j$ at output j, given $x_j \in \mathcal{X}_j$ is input by encoder j. Inputs are constrained with respect to cost functions $\kappa_j : \mathcal{X}_j \to [0, \infty) : j \in [3]$. The cost function is assumed additive, i.e., cost of transmitting vector $x_j^n \in \mathcal{X}_j^n$ is $\bar{\kappa}_j^n(x_j^n) := \frac{1}{n}\sum_{t=1}^n \kappa_j(x_{jt})$. We refer to this 3−IC as $(\underline{\mathcal{X}}, \underline{\mathcal{Y}}, W_{\underline{Y}|\underline{X}}, \underline{\kappa})$.

Definition 4.3.1 *A 3−IC code $(n, \underline{\mathcal{M}}, \underline{e}, \underline{d})$ consist of (i) index sets $\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3$ of messages, (ii) encoder maps $e_j : \mathcal{M}_j \to \mathcal{X}_j^n : j \in [3]$, and (iii) three decoder maps $d_j : \mathcal{Y}_j^n \to \mathcal{M}_j : j \in [3]$.*

Definition 4.3.2 *The error probability of a 3−IC code $(n, \underline{\mathcal{M}}, \underline{e}, \underline{d})$ conditioned on message triple $(m_1, m_2, m_3) \in \underline{\mathcal{M}}$ is*

$$\xi(\underline{e}, \underline{d}|\underline{m}) := 1 - \sum_{\underline{y}^n : \underline{d}(\underline{y}^n) = \underline{m}} W_{\underline{Y}|\underline{X}}(\underline{y}^n | e_1(m_1), e_2(m_2), e_3(m_3)).$$

The average error probability of a 3−IC code $(n, \underline{\mathcal{M}}, \underline{e}, \underline{d})$ is $\bar{\xi}(\underline{e}, \underline{d}) := \sum_{\underline{m} \in \underline{\mathcal{M}}} \frac{1}{|\mathcal{M}_1||\mathcal{M}_2||\mathcal{M}_3|} \xi(\underline{e}, \underline{d}|\underline{m})$. Average cost per symbol of transmitting message $\underline{m} \in \underline{\mathcal{M}}$ is $\underline{\tau}(\underline{e}|\underline{m}) := \big(\bar{\kappa}_j^n(e_j(m_j)) : j \in [3]\big)$ and average cost per symbol of 3−IC code $(n, \underline{\mathcal{M}}, \underline{e}, \underline{d})$ is $\underline{\tau}(\underline{e}) := \frac{1}{|\mathcal{M}_1||\mathcal{M}_2||\mathcal{M}_3|} \sum_{\underline{m} \in \underline{\mathcal{M}}} \underline{\tau}(\underline{e}|\underline{m})$.

[4] We note that the technique of interference alignment proposed by Cadambe and Jafar is restricted to Gaussian channels and achieve rates corresponding Gaussian input distributions.

Definition 4.3.3 *A rate-cost sextuple* $(R_1, R_2, R_3, \tau_1, \tau_2, \tau_3) \in [0,\infty)^6$ *is achievable if for every* $\eta > 0$*, there exists* $N(\eta) \in \mathbb{N}$ *such that for all* $n > N(\eta)$*, there exists a* $3-$*IC code* $(n, \underline{\mathcal{M}}^{(n)}, \underline{e}^{(n)}, \underline{d}^{(n)})$ *such that (i)* $\frac{\log |\mathcal{M}_j^{(n)}|}{n} \geq R_j - \eta : j \in [3]$*, (ii)* $\bar{\xi}(\underline{e}^{(n)}, \underline{d}^{(n)}) \leq \eta$*, and (iii) average cost* $\underline{\tau}(e^{(n)})_j \leq \tau_j + \eta$*. The capacity region is* $\mathbb{C}(\underline{\tau}) := \big\{\underline{R} \in \mathbb{R}^3 : (\underline{R}, \underline{\tau})$ *is achievable*$\big\}$.

We now introduce the class of $3-$to-1 IC that enables us prove strict sub-optimality coding techniques based on unstructured codes. A $3-$to-1 IC is an $3-$IC wherein two of the users enjoy interference free point-to-point links. Formally, a $3-$IC $(\underline{\mathcal{X}}, \underline{\mathcal{Y}}, W_{\underline{Y}|\underline{X}}, \underline{\tau})$ is a $3-$to-1 IC if (i) $W_{Y_2|\underline{\mathcal{X}}}(y_2|\underline{x}) := \sum_{(y_1,y_3)\in\mathcal{Y}_1\times\mathcal{Y}_3} W_{\underline{Y}|\underline{X}}(\underline{y}|\underline{x})$ is independent of $(x_1, x_3) \in \mathcal{X}_1 \times \mathcal{X}_3$, and (ii) $W_{Y_3|\underline{\mathcal{X}}}(y_3|\underline{x}) := \sum_{(y_1,y_2)\in\mathcal{Y}_1\times\mathcal{Y}_2} W_{\underline{Y}|\underline{X}}(\underline{y}|\underline{x})$ is independent of $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$ for every collection of input output symbols $(\underline{x}, \underline{y}) \in \underline{\mathcal{X}} \times \underline{\mathcal{Y}}$. For a $3-$to-1 IC, the channel transition probabilities factorize as $W_{\underline{Y}|\underline{X}}(\underline{y}|\underline{x}) = W_{Y_1|\underline{X}}(y_1|\underline{x}) W_{Y_2|X_2}(y_2|x_2) W_{Y_3|X_3}(y_3|x_3)$ for some conditional pmfs $W_{Y_1|\underline{X}}$, $W_{Y_2|X_2}$ and $W_{Y_3|X_3}$. We also note that $X_1X_3 - X_2 - Y_2$ and $X_1X_2 - X_3 - Y_3$ are Markov chains for any distribution $p_{X_1}p_{X_2}p_{X_3}W_{\underline{Y}|\underline{X}}$.[5]

In the following section, we describe the coding technique of message splitting and superposition using unstructured codes, in the context of a $2-$IC, and employ the same in deriving the $\mathscr{USB}-$region for $3-$to-1 IC.

4.4 Message splitting and superposition using unstructured codes

4.4.1 CHK-technique for $2-$IC

The main impediment to communicating efficiently over a $2-$IC is interference. As against to treating the interfering transmitters' transmission as noise, CHK-technique enables each decoder decode a part of the same to enhance it's capability to decode the desired signal. In order for encoder j to make available one part of it's transmission to the decoder $\not{j}$, it's transmission is split into two parts - public and private. Decoder $\not{j}$ decodes public part of encoder j's transmission, peels it off, and thereby enhance it's capability to decode the intended signal - public and private transmissions of encoder $\not{j}$.

Encoder j builds codebooks over two layers - public and private. The public layer contains a cloud center codebook built over $\mathcal{W}_j$. For each codeword in the cloud center codebook, a corresponding satellite codebook is built over $\mathcal{X}_j$. The satellite codebooks form the private layer. The user's message is split into two parts - public and private. The cloud center codeword is the codeword in the cloud center codebook indexed by the public part of the message. In the satellite codebook corresponding to the cloud center codeword, the codeword indexed by the private part of the message forms the satellite codeword. The satellite codeword is input on the channel. Decoder j decodes into codebooks built over $\mathcal{W}_1, \mathcal{W}_2$ and $\mathcal{X}_j$, i.e., the two cloud center codebooks and it's satellite codebook. A standard information theoretic analysis of probability of error yields an achievable rate region referred to herein as CHK rate

[5]Any interference channel wherein only one of the users is subjected to interference is a $3-$to-1 IC by a suitable permutation of the user indices.

region for 2−IC.

Definition 4.4.1 and theorem 4.4.2 provide a characterization of rate pairs achievable using CHK-technique. We omit restating the definitions analogous to definitions 4.3.1, 4.3.2, 4.3.3 for a 2−IC.

Definition 4.4.1 *Let $\mathbb{D}_{HK}(\underline{\tau})$ denote the collection of pmfs $p_{QW_1W_2X_1X_2Y_1Y_2}$ defined on $\mathcal{Q}\times\mathcal{W}_1\times\mathcal{W}_2\times\mathcal{X}_1\times\mathcal{X}_2\times\mathcal{Y}_1\times\mathcal{Y}_2$, where $\mathcal{Q},\mathcal{W}_1,\mathcal{W}_2$ are finite sets of cardinality at most $7, |\mathcal{X}_1|+4, |\mathcal{X}_2|+4$ respectively, such that (i) $p_{\underline{Y}|\underline{X}\underline{W}} = W_{\underline{Y}|\underline{X}}$, (ii) (W_1, X_1) is conditionally independent of (W_2, X_2) given Q, (iii) $\mathbb{E}\{\kappa_j(X_j)\} \leq \tau_j$. For $p_{Q\underline{W}\underline{X}\underline{Y}} \in \mathbb{D}_{HK}(\underline{\tau})$, let $\alpha_{HK}(p_{Q\underline{W}\underline{X}\underline{Y}})$ denote the set of rate pairs $(R_1, R_2) \in [0,\infty]^2$ that satisfy*

$$\begin{aligned} R_j &< \min\{I(X_j;Y_j|QW_{\not j}), I(X_j;Y_j|Q\underline{W}) + I(W_jX_{\not j};Y_{\not j}|QW_{\not j})\} : j\in[2] \\ R_1+R_2 &< \min\left\{ I(X_j;Y_j|Q\underline{W}) + I(W_jX_{\not j};Y_{\not j}|Q) : j\in[2], \sum_{j=1}^{2} I(W_jX_{\not j};Y_{\not j}|QW_{\not j}) \right\} \\ 2R_j+R_{\not j} &< I(X_j;Y_j|Q\underline{W}) + I(W_jX_{\not j};Y_{\not j}|QW_{\not j}) + I(W_{\not j}X_j;Y_j|Q) : j\in[2] \end{aligned}$$

and

$$\alpha_{HK}(\underline{\tau}) = cl\left(\bigcup_{\substack{p_{QW\underline{X}\underline{Y}}\in \\ \mathbb{D}_{HK}(\underline{\tau})}} \alpha_{HK}(p_{QW\underline{X}\underline{Y}}) \right).$$

Theorem 4.4.2 *For 2−IC $(\underline{\mathcal{X}},\underline{\mathcal{Y}},W_{\underline{Y}|\underline{X}},\underline{\kappa})$, $\alpha_{HK}(\underline{\tau})$ is achievable, i.e., $\alpha_{HK}(\underline{\tau}) \subseteq \mathbb{C}(\underline{\tau})$.* □

Remark 4.4.3 *Recently, several efforts [42], [43], [44] have yielded simplified descriptions [45] of $\alpha_{HK}(\underline{\tau})$. The description stated above involving fewer auxiliary random variables and tighter bounds on their cardinalities, is due to Chong et. al. [42].*

4.4.2 $\mathscr{USB}$−technique for 3−to−1 IC

Before we consider the case of a 3−to−1 IC, it is appropriate to state how does one optimally stitch together current known coding techniques - **m**essage splitting, **s**uperposition coding and precoding via **b**inning - for communicating over 3−IC? Each encoder must make available parts of it's transmission to each user it interferes with. Specifically, encoder j splits it's transmission into four parts - one public, two semi-private and one private. The corresponding decoder j decodes all of these parts. The other two decoders, say i and k, for which encoder j's transmission is interference, decode the public part of user j's transmission. The public part is decoded by all receivers, and is therefore encoded using a cloud center codebook at the base layer. Moreover, each semi-private part of encoder j's transmission is decoded by exactly one among the decoders i and k. The semi-private parts are encoded at the intermediate level using one codebook each. These codebooks, referred to as semi-satellite codebooks, are conditionally coded over the cloud center codebook. The semi-satellite codebooks are precoded for each other via

binning. The private part is encoded at the top layer using a satellite codebook. The satellite codebook is conditionally coded over the cloud center and semi-satellite codebooks. Each decoder decodes the seven parts using a joint typicality decoder. Finally, the encoders and decoders share a time sharing sequence to enable them synchronize the choice of codebooks at each symbol interval. We henceforth refer to the above coding technique as the $\mathscr{USB}-$technique.

One can characterize $\mathscr{USB}-$region - an achievable rate region corresponding to the above coding technique - via random coding. Indeed, such a characterization is quite involved. Since our objective is to illustrate sub-optimality of $\mathscr{USB}-$technique, it suffices to obtain a characterization of $\mathscr{USB}-$region for $3-$to-1 ICs.

For the case of $3-$to-1 IC, user 1's transmission does not cause interference to users 2 and 3, and therefore will not need it to split it's message. This can be proved using Markov chains $X_1X_3 - X_2 - Y_2$ and $X_1X_2 - X_3 - Y_3$. Moreover, transmission of user 2 does not interfere user 3's reception and vice versa. Therefore, users 2 and 3 will only need to split their messages into two parts - a private part and a semi-private part that is decoded by user 1. We now describe this coding technique.

Since encoder 1's transmission does not cause interference to any of the other users, it employs a simple PTP encoder. Specifically, encoder 1 builds a single codebook $(x_1^n(m_1) : m_1 \in \mathcal{M}_1)$ of rate T_1 over $\mathcal{X}_1$ and the codeword indexed by the message is input on the channel. The operations of encoder 2 and 3 are identical and we only describe the former. Moreover, since their transmissions cause interference only to user 1, their operations are identical to that of a generic encoder of a $2-$IC. In anticipation of a generalization to $3-$IC , we employ an alternate notation and therefore describe operation of encoder 2.

Encoder 2 splits it's message $M_2 \in \mathcal{M}_2$ into two parts - semi-private and private. We let message (i) $M_{21} \in \mathcal{M}_{21}$ of rate L_2 denote it's semi-private part and (ii) $M_{2X} \in \mathcal{M}_{2X}$ of rate T_2 denote it's private part. A single semi-private layer codebook $(u_2^n(m_{21}) : m_{21} \in \mathcal{M}_{21})$ is built over $\mathcal{U}_2$. For each message $m_{21} \in \mathcal{M}_{21}$, a codebook $(x_2(m_{21}, m_{2X}) : m_{2X} \in \mathcal{M}_{2X})$ is built over $\mathcal{X}_2$. The codebooks over $\mathcal{X}_2$ form the private layer. The codeword $x_2(M_{21}, M_{2X})$ corresponding to message $M_2 = (M_{21}, M_{2X})$ is input on the channel.

Decoders 2 and 3 enjoying interference free reception perform simple point to point joint typical decoding into the corresponding pair of semi-private and private codebooks. Decoder 1 looks for all messages $\hat{m}_1 \in \mathcal{M}_1$ for which there exists a pair $(u_2^n(\hat{m}_{21}), u_3^n(\hat{m}_3))$ such that $(x_1(\hat{m}_1), u_2^n(\hat{m}_{21}), u_3^n(\hat{m}_{31}), Y_1^n)$ is jointly typical, where Y_1^n is the vector received by decoder 1. If there is exactly, one such message $\hat{m}_1 \in \mathcal{M}_1$, this is declared as decoded message of user 1. Otherwise, an error is signaled.

A typical information theoretic analysis of probability of decoding error yields the $\mathscr{USB}-$region for $3-$to-1 IC. For the sake of completeness, we provide the details. A well versed reader may skip over to the characterization provided in definition 4.4.4 and theorem 4.4.5. Let Q, taking values over the finite alphabet $\mathcal{Q}$, denote the time sharing random variable. Let p_Q be a pmf on $\mathcal{Q}$ and $q^n \in \mathcal{Q}^n$ denote a sequence picked according to $\prod_{t=1}^n p_Q$. q^n is revealed to the encoders and decoders. The distribution induced on the ensemble of codebooks is such that,

conditioned on time sharing sequence being q^n, the three collections of codebooks, one corresponding to each user,[6] are mutually independent. Let $p_Q p_{X_1|Q} p_{U_2X_2|Q} p_{U_3X_3|Q} W_{\underline{Y}|\underline{X}}$ be a pmf on $\mathcal{Q} \times \mathcal{U}_2 \times \mathcal{U}_3 \times \underline{\mathcal{X}} \times \underline{\mathcal{Y}}$. The codewords in $\mathcal{X}_1-$codebook are independent and identically distributed according to $\prod_{t=1}^n p_{X_1|Q}(\cdot|q_t)$. The codewords in user 2's semi-private codebook are independent and identically distributed according to $\prod_{t=1}^n p_{U_2|Q}(\cdot|q_t)$. Conditioned on the entire U_2-codebook, codewords $(x_2(m_{21}, m_{2X}) : m_{2X} \in \mathcal{M}_{2X})$ in the private codebook corresponding to semi-private message m_2^U are independent and identically distributed according to $\prod_{t=1}^n p_{X_2|U_2Q}(\cdot|(u_2^n(m_2^U))_t, q_t)$. The distribution induced on user 3's codebook is analogous to that of user 2 and a description is therefore omitted.

We now average probability of decoding error over the ensemble of codebooks. The probability of either decoder 2 or 3 decoding erroneously decays exponentially if

$$L_j + T_j < I(U_jX_j; Y_j|Q) \qquad \text{and} \qquad T_j < I(X_j; Y_j|Q, U_j) : j = 2, 3.$$

The probability of decoder 1 decoding erroneously decays exponentially if

$$\begin{aligned} T_1 < I(X_1; U_2, U_3, Y_1|Q), \quad L_2 + T_1 &< I(U_2X_1; U_3Y_1|Q), \quad L_3 + T_1 < I(U_3X_1; U_2Y_1|Q), \quad \text{and} \\ L_2 + L_3 + T_1 &< I(U_2U_3X_1; Y_1|Q). \end{aligned}$$

Incorporating non-negativity constraints, $T_j \geq 0 : j \in [3], L_j \geq 0 : j = 2, 3$, substituting R_1, R_2, R_3 for $T_1, L_2 + T_2, L_3 + T_3$ respectively, and eliminating all variables except $R_j : j \in [3]$ using the technique of Fourier-Motzkin yields the following achievable rate region.

Definition 4.4.4 *Let $\mathbb{D}_u(\underline{\tau})$ denote the collection of pmfs $p_{QU_2U_3\underline{XY}}$ defined on $\mathcal{Q} \times \mathcal{U}_2 \times \mathcal{U}_3 \times \underline{\mathcal{X}} \times \underline{\mathcal{Y}}$, where $\mathcal{Q}, \mathcal{U}_2, \mathcal{U}_3$ are finite sets, such that (i) $p_{\underline{Y}|\underline{X}U_2U_3Q} = W_{\underline{Y}|\underline{X}}$, (ii) the triplet $X_1, (U_2, X_2)$ and (U_3, X_3) are conditionally mutually independent given Q, (iii) $\mathbb{E}\{\kappa_j(X_j)\} \leq \tau_j : j \in [3]$. For $p_{QU_2U_3\underline{XY}} \in \mathbb{D}_u(\underline{\tau})$, let $\alpha_u(p_{QU_2U_3\underline{XY}})$ denote the set of rate triples $(R_1, R_2, R_3) \in [0, \infty)^3$ that satisfy*

$$0 \leq R_1 < I(X_1; Y_1|Q, U_2, U_3), \quad 0 \leq R_j < I(U_jX_j; Y_j|Q) : j = 2, 3 \tag{4.1}$$

$$R_1 + R_2 < I(U_2X_1; Y_1|QU_3) + I(X_2; Y_2|QU_2), \quad R_1 + R_3 < I(U_3X_1; Y_1|QU_2) + I(X_3; Y_3|QU_3)$$

$$R_1 + R_2 + R_3 < I(U_2U_3X_1; Y_1|Q) + I(X_2; Y_2|QU_2) + I(X_3; Y_3|QU_3), \tag{4.2}$$

and

$$\alpha_u(\underline{\tau}) = cl\left(\bigcup_{\substack{p_{QU_2U_3\underline{XY}} \in \\ \mathbb{D}_u(\underline{\tau})}} \alpha_u(p_{QU_2U_3\underline{XY}})\right).$$

[6]Here, the collection of user j's codebooks refers to the entire collection of codebooks employed by encoder j.

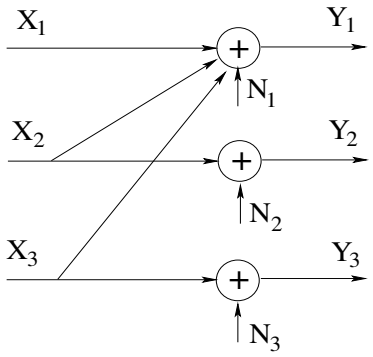

Figure 4.2: A binary additive 3−to−1 IC described in example 4.5.1.

Theorem 4.4.5 *For 3−to−1 IC* $(\underline{\mathcal{X}}, \underline{\mathcal{Y}}, W_{\underline{Y}|\underline{X}}, \underline{\kappa})$, $\alpha_u(\underline{\tau})$ *is achievable, i.e.,* $\alpha_u(\underline{\tau}) \subseteq \mathbb{C}(\underline{\tau})$. □

The reader will also recognize that $\alpha_u(\underline{\tau})$ is indeed achievable over an arbitrary 3−IC.[7] This is stated below.

Theorem 4.4.6 *For 3−IC* $(\underline{\mathcal{X}}, \underline{\mathcal{Y}}, W_{\underline{Y}|\underline{X}}, \underline{\kappa})$, $\alpha_u(\underline{\tau})$ *is achievable, i.e.,* $\alpha_u(\underline{\tau}) \subseteq \mathbb{C}(\underline{\tau})$. □

4.5 Strict sub-optimality of $\mathscr{USB}$−region for 3−to−1 IC

This section contains our first main finding of this chapter - strict sub-optimality of $\mathscr{USB}$−technique. In particular, we identify a binary additive 3−to−1 IC for which we prove strict sub-optimality of $\mathscr{USB}$−technique. We begin with the description of the 3−to−1 IC.

Example 4.5.1 *Consider a binary additive 3−to−1 IC illustrated in figure 4.2 with* $\mathcal{X}_j = \mathcal{Y}_j = \{0,1\} : j \in [3]$ *with channel transition probabilities* $W_{\underline{Y}|\underline{X}}(\underline{y}|\underline{x}) = BSC_{\delta_1}(y_1|x_1 \oplus x_2 \oplus x_3) BSC_{\delta_2}(y_2|x_2) BSC_{\delta_3}(y_3|x_3)$, *where* $BSC_\eta(0|1) = BSC_\eta(1|0) = 1 - BSC_\eta(0|0) = 1 - BSC_\eta(1|1) = \eta$ *denotes the transition probabilities of a BSC with cross over probability* $\eta \in [0, \frac{1}{2}]$. *Inputs of users 2 and 3 are not costed, i.e.,* $\kappa_j(0) = \kappa_j(1) = 0$ *for* $j = 2, 3$. *User 1's input is constrained with respect to a Hamming cost function, i.e.,* $\kappa_1(x) = x$ *for* $x \in \{0,1\}$ *to an average cost of* $\tau \in (0, \frac{1}{2})$ *per symbol. Let* $\mathbb{C}(\tau)$ *denote the capacity region of this 3−to−1 IC.*

Clearly, $\mathbb{C}(\tau) \subseteq \beta(\tau, \frac{1}{2}, \frac{1}{2}, \underline{\delta})$, where

$$\beta(\underline{\tau}, \underline{\delta}) := \left\{(R_1, R_2, R_3) \in [0, \infty)^3 : R_j \leq h_b(\delta_j * \tau_j) - h_b(\delta_j) : j = 1, 2, 3\right\}. \tag{4.3}$$

Let us focus on achievability. We begin with a few simple observations for the above channel. Let us begin with the assumption $\delta := \delta_2 = \delta_3$. As illustrated in figure 4.2, users 2 and 3 enjoy interference free unconstrained binary symmetric channels (BSC) with cross over probability $\delta = \delta_2 = \delta_3$. They can therefore communicate at their

[7] Unless the 3−IC $(\underline{\mathcal{X}}, \underline{\mathcal{Y}}, W_{\underline{Y}|\underline{X}}, \underline{\kappa})$ is a 3−to−1IC, $\alpha_u(\underline{\tau})$ is *not* it's $\mathscr{USB}$−region.

respective capacities $1 - h_b(\delta)$. Constrained to average Hamming weight of τ, user 1 cannot hope to achieve a rate larger than $h_b(\tau * \delta_1) - h_b(\delta_1)$.[8] What is the maximum rate achievable by user 1 while users 2 and 3 communicate at their respective capacities?

User 1 cannot hope to achieve rate $h_b(\tau * \delta_1) - h_b(\delta_1)$ *and* decode the pair of codewords transmitted by user 2 and 3 if $h_b(\tau * \delta_1) - h_b(\delta_1) + 2(1 - h_b(\delta)) > 1 - h_b(\delta_1)$ or equivalently $1 + h_b(\tau * \delta_1) > 2h_b(\delta)$. Under this condition, $\mathscr{USB}-$technique forces decoder 1 to be contented to decoding univariate components - represented through semi-private random variables U_2, U_3 - of user 2 and 3's transmissions. We state that as long as the univariate components leave residual uncertainty in the interfering signal, i.e., $H(X_2 \oplus X_3 | U_2, U_3) > 0$, the rate achievable by user 1 is strictly smaller than it's maximum $h_b(\tau * \delta_1) - h_b(\delta_1)$.[9] This claim and the proof of strict sub-optimality of $\mathscr{USB}-$technique is proved in theorem 4.5.3.

We now describe a simple linear coding technique that enables user 1 achieve it's maximum rate $h_b(\tau * \delta_1) - h_b(\delta_1)$ even under the condition $1 + h_b(\tau * \delta_1) > 2h_b(\delta)$! Let us assume $\tau * \delta_1 \leq \delta$. We choose a linear code, or a coset thereof, that achieves capacity of a BSC with cross over probability δ. We equip users 2 and 3 with the same code, thereby constraining the sum of their transmitted codewords to this linear code, or a coset thereof, of rate $1 - h_b(\delta)$. Since $\tau * \delta_1 \leq \delta$, decode 1 can first decode the interfering signal - sum of codewords transmitted by encoders 2 and 3 - treating the rest as noise, peel it off, and then decode the desired signal. User 1 can therefore achieve it's maximum rate $h_b(\tau * \delta_1) - h_b(\delta_1)$ if $\tau * \delta_1 \leq \delta$.

Are the two conditions $1 + h_b(\tau * \delta_1) > 2h_b(\delta)$ and $\tau * \delta_1 \leq \delta$ mutually exclusive? The two conditions are satisfied if $h_b(\tau * \delta_1) \leq h_b(\delta) < \frac{1+h_b(\tau*\delta_1)}{2}$. If $\tau * \delta_1 < \frac{1}{2}$, then $h_b(\tau * \delta_1) < \frac{1+h_b(\tau*\delta_1)}{2} < 1$ and δ can be chosen appropriately to ensure the two conditions are satisfied. For example, the choice $\delta_1 = 0.01$, $\tau = \frac{1}{8}$ and $\delta \in (0.1325, 0.21)$ proves these two conditions are indeed *not* mutually exclusive.

Let us now consider the general case with respect to δ_2, δ_3 and assume without loss of generality $\delta_2 \leq \delta_3$. The linear coding scheme generalizes naturally. We employ a capacity achieving linear code, or a coset thereof, that achieves capacity of BSC of user 2. This code, or a coset thereof, is sub-sampled uniformly at random to build a capacity achieving code for BSC of user 3. The sum of user 2 and user 3's transmissions is contained within a coset of user 2's code and can therefore be decoded by user 1 as long as $\tau * \delta_1 \leq \delta_2$. The above arguments are summarized in the following lemma.

Lemma 4.5.2 *Consider the* $3-to-1$ *IC in example 4.5.1. If* $\tau * \delta_1 \leq \min\{\delta_2, \delta_3\}$*, then* $\mathbb{C}(\tau) = \beta(\tau, \frac{1}{2}, \frac{1}{2}, \underline{\delta})$*, where* $\beta(\underline{\tau}, \underline{\delta})$ *is given by (4.3).* □

In theorem 4.5.3, we prove that if $1 + h_b(\delta_1 * \tau) > h_b(\delta_2) + h_b(\delta_3)$, then $(h_b(\tau * \delta_1) - h_b(\delta_1), 1 - h_b(\delta_2), 1 - h_b(\delta_3)) \notin \alpha_u(\tau)$. We therefore conclude in corollary 4.5.5 that if $\tau, \delta_1, \delta_2, \delta_3$ are such that $1 + h_b(\delta_1 * \tau) > h_b(\delta_2) + h_b(\delta_3)$ and

[8] If receiver 1 is provided with the codewords transmitted by users 2 and 3, the effective channel it sees is a BSC with cross over probability δ_1.

[9] An informed reader will be able to reason this by relating this situation to a point to point channel with partial state observed at the receiver.

$\min\{\delta_2, \delta_3\} \geq \delta_1 * \tau$, then $\mathscr{USB}-$technique is strictly suboptimal for the $3-$to-1 IC presented in example 4.5.1.

Theorem 4.5.3 *Consider the $3-to-1$ IC described in example 4.5.1. If $h_b(\delta_2) + h_b(\delta_3) < 1 + h_b(\tau * \delta_1)$, then $(h_b(\tau * \delta_1) - h_b(\delta_1), 1 - h_b(\delta_2), 1 - h_b(\delta_3)) \notin \alpha_u(\tau)$.* □

Proof: If $H(X_j|Q, U_j) = 0$ for $j = 2, 3$, then the upper bound in (4.2) reduces to $R_1 + R_2 + R_3 \leq I(X_2X_3X_1; Y_1|Q) \leq 1 - h_b(\delta_1)$. From the hypothesis, we have $h_b(\tau * \delta_1) - h_b(\delta_1) + 1 - h_b(\delta_2) + 1 - h_b(\delta_3) > 1 - h_b(\delta_1)$ which violates the above upper bound and hence the theorem statement is true.

Henceforth, we assume $H(X_j|Q, U_j) > 0$ for $j = 2$ or $j = 3$. Let us assume $j, \bar{j}$ are distinct elements in $\{2, 3\}$ and $H(X_j|Q, U_j) > 0$. Since (U_2, X_2) and (U_3, X_3) are conditionally independent given Q, we have

$$0 < H(X_j|Q, U_j) = H(X_j|X_{\bar{j}}, Q, U_2, U_3) = H(X_2 \oplus X_3|X_{\bar{j}}, Q, U_2, U_3) \leq H(X_2 \oplus X_3|Q, U_2U_3).$$

The univariate components U_2, U_3 leave residual uncertainty in the interfering signal and imply the existence of a $\tilde{q}^* = (q^*, u_2^*, u_3^*) \in \tilde{\mathcal{Q}} := \mathcal{Q} \times \mathcal{U}_2 \times \mathcal{U}_3$ for which $H(X_2 \oplus X_3|(Q, U_2U_3) = \tilde{q}^*) > 0$. Under this condition, we prove that the upper bound (4.1) on R_1 is strictly smaller than $h_b(\tau * \delta_1) - h_b(\delta_1)$. Towards that end, we prove a simple observation based on strict concavity of binary entropy function.

Lemma 4.5.4 *If $Z_j : j \in [3]$ are binary random variables such that (i) $H(Z_1) \geq H(Z_2)$, (ii) Z_3 is independent of (Z_1, Z_2), then $H(Z_1) - H(Z_2) \geq |H(Z_1 \oplus Z_3) - H(Z_2 \oplus Z_3)|$. Moreover, if $H(Z_1) > H(Z_2)$ and $H(Z_3) > 0$, then the inequality is strict, i.e., $H(Z_1) - H(Z_2) > |H(Z_1 \oplus Z_3) - H(Z_2 \oplus Z_3)|$.* □

Proof: Note that, if either $H(Z_1) = H(Z_2)$ or $H(Z_3) = 0$, then $H(Z_1) - H(Z_2) = H(Z_1 \oplus Z_3) - H(Z_2 \oplus Z_3)$. We therefore assume $H(Z_1) > H(Z_2)$ and $H(Z_3) > 0$ and prove the case of strict inequality. For $j \in [3]$, let $\{p_{Z_j}(0), p_{Z_j}(1)\} = \{\delta_j, 1 - \delta_j\}$ with $\delta_j \in [0, \frac{1}{2}]$, $\delta_3 > 0$. Define $f : [0, \frac{1}{2}] \to [0, 1]$ as $f(t) = h_b(\delta_1 * t) - h_b(\delta_2 * t)$. It suffices to prove $f(0) > f(\delta_3)$. By the Taylor series, $f(\delta_3) = f(0) + \delta_3 f'(\zeta)$ for some $\zeta \in [0, \delta_3]$ and therefore it suffices to prove $f'(t) < 0$ for $t \in (0, \frac{1}{2}]$.

It maybe verified that

$$f'(t) = (1 - 2\delta_1) \log \frac{1 - \bar{\delta}_1}{\bar{\delta}_1} - (1 - 2\delta_2) \log \frac{1 - \bar{\delta}_2}{\bar{\delta}_2}, \text{ where } \bar{\delta}_j = \delta_j + t(1 - 2\delta_j) : j \in [2].$$

Note that (i) $0 \leq (1 - 2\delta_1) < (1 - 2\delta_2) \leq 1$, (ii) $\bar{\delta}_j \leq \delta_j + \frac{1}{2}(1 - 2\delta_j) \leq \frac{1}{2}$, (iii) since $\delta_1 > \delta_2$ and $t \leq \frac{1}{2}$, $\bar{\delta}_1 - \bar{\delta}_2 = (\delta_1 - \delta_2)(1 - 2t) \geq 0$. We therefore have $0 \leq \bar{\delta}_2 \leq \bar{\delta}_1 \leq \frac{1}{2}$ and thus $\log \frac{1-\bar{\delta}_2}{\bar{\delta}_2} \geq \log \frac{1-\bar{\delta}_1}{\bar{\delta}_1}$. Combining this with the first observation, we conclude $(1 - 2\delta_2) \log \frac{1-\bar{\delta}_2}{\bar{\delta}_2} > (1 - 2\delta_1) \log \frac{1-\bar{\delta}_1}{\bar{\delta}_1}$ which implies $f'(t) < 0$ for $t \in (0, \frac{1}{2}]$. ■

We are now equipped to work with the upper bound (4.1) on R_1. Denoting $\tilde{Q} := (Q, U_2, U_3)$ and a generic element $\tilde{q} := (q, u_2, u_3) \in \tilde{\mathcal{Q}} := \mathcal{Q} \times \mathcal{U}_2 \times \mathcal{U}_3$, we observe that

$$
\begin{aligned}
I(X_1; Y_1|\tilde{Q}) &= H(Y_1|\tilde{Q}) - H(Y_1|\tilde{Q}X_1) \\
&= \sum_{\tilde{q}} p_{\tilde{Q}}(\tilde{q}) H(Y_1|\tilde{Q} = \tilde{q}) - \sum_{x_1, \tilde{q}} p_{\tilde{Q}X_1}(\tilde{q}, x_1) H(Y_1|X_1 = x_1, \tilde{Q} = \tilde{q}) \\
&= \sum_{\tilde{q}} p_{\tilde{Q}}(\tilde{q}) H(X_1 \oplus N_1 \oplus X_2 \oplus X_3|\tilde{Q} = \tilde{q}) - \sum_{x_1, \tilde{q}} p_{X_1\tilde{Q}}(x_1, \tilde{q}) H(x_1 \oplus N_1 \oplus X_2 \oplus X_3|X_1 = x_1, \tilde{Q} = \tilde{q}) \\
&= \sum_{\tilde{q}} p_{\tilde{Q}}(\tilde{q}) H(X_1 \oplus N_1 \oplus X_2 \oplus X_3|\tilde{Q} = \tilde{q}) - \sum_{x_1, \tilde{q}} p_{X_1\tilde{Q}}(x_1, \tilde{q}) H(N_1 \oplus X_2 \oplus X_3|\tilde{Q} = \tilde{q}) \qquad (4.4) \\
&= \sum_{\tilde{q}} p_{\tilde{Q}}(\tilde{q}) H(X_1 \oplus N_1 \oplus X_2 \oplus X_3|\tilde{Q} = \tilde{q}) - \sum_{\tilde{q}} p_{\tilde{Q}}(\tilde{q}) H(N_1 \oplus X_2 \oplus X_3|\tilde{Q} = \tilde{q}) \\
&\leq \sum_{\tilde{q}} p_{\tilde{Q}}(\tilde{q}) H(X_1 \oplus N_1|\tilde{Q} = \tilde{q}) - \sum_{\tilde{q}} p_{\tilde{Q}}(\tilde{q}) H(N_1|\tilde{Q} = \tilde{q}) = \sum_{q} p_Q(q) H(X_1 \oplus N_1|Q = q) - h_b(\delta_1) \qquad (4.5) \\
&= \sum_{q} p_Q(q) h_b(p_{X_1|Q}(1|q) * \delta_1) - h_b(\delta_1) \leq h_b(\mathbb{E}_Q \left\{ p_{X_1|Q}(1|q) * \delta_1 \right\}) - h_b(\delta_1) \leq h_b(\tau * \delta_1) - h_b(\delta_1), \qquad (4.6)
\end{aligned}
$$

where (i) (4.4) follows from independence of (N_1, X_2, X_3) and X_1 conditioned on realization of Q, (ii) (4.5) follows from existence of a $\tilde{q}^* \in \tilde{\mathcal{Q}}$ for which $H(X_2 \oplus X_3|\tilde{Q} = \tilde{q}^*) > 0$ and substituting $p_{X_1 \oplus N_1|\tilde{Q}}(\cdot|\tilde{q}^*)$ for p_{Z_1}, $p_{N_1|\tilde{Q}}(\cdot|\tilde{q}^*)$ for p_{Z_2} and $p_{X_2 \oplus X_3|\tilde{Q}}(\cdot|\tilde{q}^*)$ for p_{Z_3} in lemma 4.5.4, (iii) the first inequality in (4.6) follows from Jensen's inequality and the second follows from the cost constraint that any test channel in $\mathbb{D}_{3\text{-}1}(\tau)$ must satisfy.

The first inequality in (4.6) is an equality if and only if $p_{X_1|\tilde{Q}}(1|\tilde{q}) = \tau$ for all $\tilde{q} \in \tilde{\mathcal{Q}}$ For $\tilde{Q} = \tilde{q}^*$, we have $p_{X_1 \oplus N_1|\tilde{Q}}(1|\tilde{q}^*) > p_{N_1|\tilde{Q}}(1|\tilde{q}^*)$ and from lemma 4.5.4, we have $I(X_1; Y_1|\tilde{Q}) < h_b(\tau * \delta_1) - h_b(\delta_1)$. ∎

Corollary 4.5.5 *Consider the 3−to−1 IC in example 4.5.1 with $\delta = \delta_2 = \delta_3$. If $h_b(\tau * \delta_1) \leq h_b(\delta) < \frac{1+h_b(\delta_1 * \tau)}{2}$, then $(h_b(\tau * \delta_1) - h_b(\delta_1), 1 - h_b(\delta), 1 - h_b(\delta)) \notin \alpha_u(\tau)$ but $(h_b(\tau * \delta_1) - h_b(\delta_1), 1 - h_b(\delta), 1 - h_b(\delta)) \in \mathbb{C}(\tau)$ and thus $\alpha_u(\tau) \neq \mathbb{C}(\tau)$. In particular, if $\delta_1 = 0.01$ and $\delta_2 \in (0.1325, 0.21)$, then $\alpha_u(\frac{1}{8}) \neq \mathbb{C}(\frac{1}{8})$.*

4.5.1 The non-trivial role played by structured codes

In this section, we emphasize the role of algebraic closure properties of coset codes. The observant reader would have noted two new elements in the coding technique proposed for example 4.5.1. Firstly, we propose decoding a *bivariate* function of the codewords input by users 2 and 3. Secondly, we employ an ensemble of structured codes, linear codes in this case, to limit the number of interference patterns. An informed reader may note that the CHK-technique is based on decoding univariate functions of the other user's transmission. One might then claim that a natural extension of CHK-technique for the three user case must involve users decoding bivariate components of the interfering user's transmissions. Such a technique may be further enhanced by recognizing bivariate and univariate functions tend to 'saturate' and thereby enhancing the decoding rule. We refer the reader to Bandemer and El

Gamal [46] that presents an achievable rate region that exploits decoding bivariate functions and it's saturation. Our technique goes further through the use of structured code ensembles. When the rates of users are such that the number of interfering patterns does not saturate, then the structured codes play a non-trivial role in limiting the same. Example 4.5.1 with the parameters as stated in corollary 4.5.5 demonstrates this. In particular, we prove that for the choice of parameters as in corollary 4.5.5, the rate triple proved achievable using linear codes is not contained within the rate region presented in [46].

We refer the reader to [46, Section II.D] wherein the authors propose an achievable rate region for the three user deterministic interference channel with noisy observations. To avoid conflict in notation, we restate example 4.5.1 with a notation consistent with that employed in [46].

Example 4.5.6 *Consider a binary additive* $3-to-1$ *IC illustrated in figure 4.2 with* $\mathcal{X}_j = \mathcal{Z}_j = \{0,1\} : j \in [3]$ *with channel transition probabilities* $W_{\underline{Z}|\underline{X}}(\underline{z}|\underline{x}) = BSC_{\delta_1}(z_1|x_1 \oplus x_2 \oplus x_3) BSC_{\delta_2}(z_2|x_2) BSC_{\delta_3}(z_3|x_3)$*, where* $BSC_\eta(0|1) = BSC_\eta(1|0) = 1 - BSC_\eta(0|0) = 1 - BSC_\eta(1|1) = \eta$ *denotes the transition probabilities of a BSC's with cross over probability* $\eta \in [0, \frac{1}{2}]$*. Inputs of users* 2 *and* 3 *are not costed, i.e.,* $\kappa_j(0) = \kappa_j(1) = 0$ *for* $j = 2, 3$ *and user* 1*'s input is constrained by a Hamming cost function, i.e.,* $\kappa_1(x) = x$ *for* $x \in \{0,1\}$*.*

Let us describe the above example using the notation employed in [46]. It maybe verified that $X_{12}, X_{13}, X_{23}, X_{32}$, S_2, S_3 are trivial, $X_{j1} = X_j$ for $j = 1, 2, 3$, $Y_2 = X_{22} = X_2$, $Y_3 = X_{33} = X_3$, $S_1 = X_{21} \oplus X_{31}$, $Y_1 = X_1 \oplus S_1$, $Z_j = Y_j \oplus N_j$ for $j = 1, 2, 3$. N_1, N_2, N_3 are independent Bernoulli processes with $P(N_1 = 1) = \delta_1$ and $P(N_j = 1) = \delta$ for $j = 2, 3$. We now state the main elements in the argument that proves $(h_b(\tau * \delta_1) - h_b(\delta_1), 1 - h_b(\delta_2), 1 - h_b(\delta_3)) \notin \mathfrak{R}_{\text{ID}}$. Let (Q, X_1, X_2, X_3) be such that $(R_1, 1 - h_b(\delta_2), 1 - h_b(\delta_3)) \in \cap_{k=1}^3 \mathfrak{R}_k(Q, X_1, X_2, X_3)$. It can be proved that $p_{X_j|Q}(0|q) = p_{X_j|Q}(1|q) = \frac{1}{2}$ for every $q \in \mathcal{Q}$ and $j = 2, 3$ using standard information theoretic arguments[10]. We now employ the bound

$$R_1 + \min\{R_2 + H(X_{31}|Q), R_3 + H(X_{21}|Q), R_2 + R_3, H(S_1|Q)\} \leq I(X_1, S_1; Z_1|Q) \tag{4.7}$$

present in the description of $\mathfrak{R}_1(Q, X_1, X_2, X_3)$. Clearly, the right hand side of (4.7) is $1 - h_b(\delta_1)$. We also know $R_2 + R_3 \leq \min\{R_2 + H(X_{31}|Q), R_3 + H(X_{21}|Q)\}$. If $R_2 + R_3 \leq H(S_1|Q) = H(X_{21} \oplus X_{31}|Q) = H(X_2 \oplus X_3|Q) = 1$, then the above bound reduces to $R_1 + R_2 + R_3 \leq 1 - h_b(\delta_1)$. Therefore, if $(2 - 2h_b(\delta)) \leq 1$, or equivalently $h_b(\delta) > \frac{1}{2}$, we have $R_1 + R_2 + R_3 \leq 1 - h_b(\delta_1)$. Consider the choice $\delta_1 = 0.01, \tau = \frac{1}{8}$ and $\delta = 0.15$. We have $h_b(\tau * \delta_1) \leq h_b(\delta) < \frac{1 + h_b(\delta_1 * \tau)}{2}$ and therefore $(2 - 2h_b(\delta)) + (h_b(\delta_1 * \tau) - h_b(\delta_1)) > (1 - h_b(\delta_1 * \tau)) + (h_b(\delta_1 * \tau) - h_b(\delta_1)) = 1 - h_b(\delta_1)$ and moreover $h_b(\delta) = 0.6098 > \frac{1}{2}$. Therefore the rate triple $(h_b(\tau * \delta_1) - h_b(\delta_1), 1 - h_b(\delta_2), 1 - h_b(\delta_3)) \notin \mathfrak{R}_{\text{ID}}$ but is achievable using linear codes.

[10]This can be proved by employing the bound $R_j < I(X_j; Z_j|S_j, Q)$ involved in the description of $\mathfrak{R}_j(Q, X_1, X_2, X_3)$ for $j = 2, 3$ and noting that S_j if trivial for these j.

4.6 Achievable rate region for an arbitrary 3−IC

In this section, we present our second main finding - a new achievable rate region for a general discrete 3−IC based on partitioned coset codes (PCC). This rate region, referred to PCC rate region, in conjunction with the $\mathscr{USB}$−region strictly enlarges upon the latter, which is the current known largest for a general 3−IC. We derive PCC rate region in two pedagogical steps. In the first step, presented in section 4.6.1, we employ PCC to manage interference seen by only one of the receivers. This simplified setting aids the reader recognize and absorb all the key elements of the framework proposed herein. For this step, we provide a complete and elaborate proof of achievability. In this section, we also identify a *non-additive* 3−to−1 IC (Example 4.6.7) for which we *analytically* prove (i) strict sub-optimality of $\mathscr{USB}$−technique and (ii) optimality of PCC rate region. This example indeed illustrates the central theme of this thesis - codes endowed with algebraic closure properties enable efficient communication over arbitrary general multi-terminal communication channels, not just additive, symmetric instances - and thereby justifies the framework developed herein.

In the second step, presented in section 4.6.2, we employ PCC to manage interference seen by every receiver and thereby provide a characterization of PCC rate region. In section 4.6.3, we indicate a coding technique that incorporates PCC and unstructured independent codes for managing interference over a 3−IC. Any characterization of the corresponding rate region being quite involved, we refrain from providing the same.

4.6.1 Step I : Decoding sum of codewords chosen from PCC over an arbitrary 3−IC

The linear coding technique proposed for example 4.5.1 seems to hinge on the additive nature of the channel therein. One of our main contributions is in being able to generalize this technique to arbitrary channels. In this section, we present our generalization in a simple setting that elaborates on the structure of the codebooks and captures all the key elements.

Definition 4.6.1 *Let $\mathbb{D}_f(\underline{\tau})$ denote the collection of distributions $p_{QU_2U_3\underline{XY}} \in \mathbb{D}_f(\underline{\tau})$ defined over $\mathcal{Q}\times\mathcal{U}_2\times\mathcal{U}_3\times\underline{\mathcal{X}}\times\underline{\mathcal{Y}}$, where $\mathcal{U}_2 = \mathcal{U}_3$ is a finite field. For $p_{QU_2U_3\underline{XY}} \in \mathbb{D}_f^{3\text{-}1}(\underline{\tau})$, let $\alpha_f^{3\text{-}1}(p_{QU_2U_3\underline{XY}})$ be defined as the set of rate triples $(R_1, R_2, R_3) \in [0,\infty)^3$ that satisfy*

$$\begin{aligned}
R_1 &< \min\{0, H(U_j|Q) - H(U_2 \oplus U_3|QY_1) : j = 2,3\} + I(X_1; U_2 \oplus U_3, Y_1|Q),\\
R_j &< I(U_j, X_j; Y_j|Q) : j = 2,3,\\
R_1 + R_j &< I(X_j; Y_j|QU_j) + I(X_1; U_2 \oplus U_3, Y_1|Q) + H(U_j|Q) - H(U_2 \oplus U_3|QY_1) : j = 2,3,
\end{aligned}$$

and

$$\alpha_f^{3\text{-}1}(\underline{\tau}) = cocl\left(\bigcup_{\substack{p_{QU_2U_3\underline{XY}}\in\\ \mathbb{D}_f(\underline{\tau})}} \alpha_f^{3\text{-}1}(p_{QU_2U_3\underline{XY}})\right).$$

Theorem 4.6.2 *For 3−IC* $(\underline{\mathcal{X}}, \underline{\mathcal{Y}}, W_{\underline{Y}|\underline{X}}, \underline{\kappa})$, $\alpha_f^{3\text{-}1}(\underline{\tau})$ *is achievable, i.e.,* $\alpha_f^{3\text{-}1}(\underline{\tau}) \subseteq \mathbb{C}(\underline{\tau})$. □

Before we provide a proof, we describe the coding technique in a simplified setting that highlights the new elements and indicates achievability of promised rates. Towards that end, consider a pmf $p_{QU_2U_3\underline{XY}} \in \mathbb{D}_f(\underline{\tau})$ with $\mathcal{Q} = \phi$[11] and $\mathcal{U}_2 = \mathcal{U}_3 = \mathcal{F}_\pi$. Except for key differences, the coding technique proposed herein is identical to $\mathscr{USB}-$technique for the 3−to−1 IC (section 4.4.2). Let us revisit the linear coding technique proposed for example 4.5.1 to identify these key differences. Note that the structure and encoding rule of user 1 in example 4.5.1 and section 4.4.2 are identical. We therefore employ the same code structure and encoding rules for user 1. In particular, encoder 1 builds a single codebook $\mathcal{C}_1 = (x_1^n(m_1) : m_1 \in \mathcal{M}_1)$ of rate R_1 over $\mathcal{X}_1$ and the codeword indexed by the message is input on the channel.

The structure and encoding rules for users 2 and 3 are identical and we describe it using the generic index $j \in \{2, 3\}$. As in section 4.4.2, we employ a two layer - cloud center and satellite - code for user j and split it's message $M_j \in \mathcal{M}_j$ into two parts. Let (i) $M_{j1} \in \mathcal{M}_{j1} := [\pi^{t_j}]$ denote it's semi-private part, and (ii) $M_{jX} \in \mathcal{M}_{jX} := [\exp\{nL_j\}]$ denote it's private part. While in section 4.4.2 user 1 decoded the pair of cloud center codewords, the first key difference we propose is that user 1 decode the sum of user 2 and 3 cloud center codewords. Let coset $\lambda_j \subseteq \mathcal{U}_j^n$ of linear code $\overline{\lambda}_j \subseteq \mathcal{U}_j^n$ denote user j's cloud center codebook. In particular, let $g_j \in \mathcal{U}_j^{s_j \times n}$ denote generator matrix of $\overline{\lambda}_j$ and coset λ_j correspond to shift $b_j^n \in \mathcal{U}_j^n$. The second key difference we propose is that cloud center codebooks of users' 2 and 3 overlap, i.e., the larger of $\overline{\lambda}_2, \overline{\lambda}_3$ contains the other. For example, if λ_j contains π^{s_j} codewords[12] and $s_{j_1} \leq s_{j_2}$, then $\overline{\lambda}_{j_1} \subseteq \overline{\lambda}_{j_2}$. We therefore let $g_{j_2}^T = \begin{bmatrix} g_{j_1}^T & g_{j_2/j_1}^T \end{bmatrix}$.

Since codewords of a uniformly distributed coset code are uniformly distributed, we need to partition the coset code into bins to induce non-uniform distribution over the auxiliary alphabet $\mathcal{U}_j$. We therefore employ partitioned coset codes (section 3.4.2). The third key difference is therefore a partition of λ_j into π^{t_j} bins to enable induce a non-uniform distribution. For the benefit of a reader who has not studied through section 3.4.2, we describe and define partitioned coset codes again. In particular, for each codeword $u_j^n(a^{s_j}) := a^{s_j} g_j \oplus b_j^n$, where $a^{s_j} \in \mathcal{U}_j^{s_j}$, an index $i_j(a^{s_j}) \in [\pi^{t_j}]$ is defined that indexes the bin containing $u_j^n(a^{s_j})$. We let $c_{j1}(m_{j1}) = \{a^{s_j} \in \mathcal{U}_j^{s_j} : i_j(a^{s_j}) = m_{j1}\}$ denote the set containing indices corresponding to message m_{j1}.

The structure of the cloud center codebook plays an important role in this chapter and we formalize the same through the following definition.

Definition 4.6.3 *Recall that a coset code* $\lambda \subseteq \mathcal{F}_\pi^n$ *is a coset of a linear code* $\overline{\lambda} \subseteq \mathcal{F}_\pi^n$. *The coset code is completely specified by the generator matrix* $g \in \mathcal{F}_\pi^{k \times n}$ *and a bias vector* $b_j^n \in \mathcal{F}_\pi^n$. *Consider a partition of* λ *into* π^l *bins. Each codeword* $a^k g \oplus b^n$ *is assigned an index* $i(a^k) \in [\pi^l]$. *This coset code* λ *with it's partitions is referred to as partitioned coset code (PCC)* (n, k, l, g, b^n, i) *or succinctly as an* (n, k, l) *PCC. For each* $m \in [\pi^l]$, *let* $c(m) := \left\{a^k \in \mathcal{F}_\pi^k : i(a^k) = m\right\}$.

[11] Since the time sharing random variable Q is employed in a standard way, we choose to omit the same in this description.
[12] Recall $|\mathcal{U}_j| = \pi$.

User j's satellite codebook, built over $\mathcal{X}_j$, consists of $\exp\{nL_j\}$ bins, one for each private message $m_{jX} \in \mathcal{M}_{jX} := [\exp\{nL_j\}]$. Let $(x_j^n(m_{jX}, b_{jX}) \in \mathcal{X}_j^n : b_{jX} \in c_{jX} := [\exp\{nK_j\}])$ denote bin corresponding to message $m_{jX} \in \mathcal{M}_{jX}$. Having received message $M_j = (M_{j1}, M_{jX})$, the encoder identifies all pairs $(u_j^n(a^{s_j}), x_j^n(M_{jX}, b_{jX}))$ of jointly typical codewords with $(a^{s_j}, b_{jX}) \in c_{j1}(M_{j1}) \times c_{jX}$. If it finds one or more such pairs, one of them is chosen and the corresponding satellite codeword is fed as input on the channel. Otherwise, an error is declared.

We now describe the decoding rule. Predictably, the decoding rules of users 2 and 3 are identical and we describe this through the generic index $j \in \{2,3\}$. Except for a slight modification to handle the bins in the codebooks, decoder j's operation is identical in spirit to a point to point decoder described in section 4.4.2. Specifically, decoder j identifies all $(\hat{m}_{j1}, \hat{m}_{jX})$ for which there exists $(a^{s_j}, b_{jX}) \in c_{j1}(\hat{m}_{j1}) \times c_{jX}$ such that $(u_j^n(a^{s_j}), x_j^n(\hat{m}_{jX}, b_{jX}), Y_j^n)$ is jointly typical with respect to $p^n_{U_jX_j,Y_j}$. If there is exactly one such pair $(\hat{m}_{j1}, \hat{m}_{jX})$, this is declared as user j message. Otherwise an error is signaled.

Decoder 1 constructs the sum $\lambda_2 \oplus \lambda_3 := \left\{u_2^n \oplus u_3^n : u_j^n \in \lambda_j, j = 2,3\right\}$ of the cloud center codebooks. Having received Y_1^n, it looks for all potential message $\hat{m}_1$ for which there exists a $u^n_{2\oplus 3} \in \lambda_2 \oplus \lambda_3$ such that $(u^n_{2\oplus 3}, x_1^n(\hat{m}_1), Y_1^n)$ is jointly typical with respect to $p^n_{U_2\oplus U_3, X_1, Y_1}$. If it finds exactly one such message $\hat{m}_1$, it declares this as decoded message of user 1. Otherwise, it declares an error.

As is typical in information theory, we derive an upper bound on the probability of error by averaging over the ensemble of codebooks. Let $\mathcal{C}_1 = (X_1^n(m_1) : m_1 \in \mathcal{M}_1)$ denote random codebook of user 1. Let $\{j_1, j_2\} = \{2,3\}$, be such that $s_{j_1} \leq s_{j_2}$ and $\Lambda_{j_1}, \Lambda_{j_2}$ denote the random coset codes of users j_1 and j_2 respectively. Let matrices $G_{j_1}, G_{j_2}^t = \begin{bmatrix} G_{j_1}^t & G_{j_2/j_1}^t \end{bmatrix}$ and vectors $B_{j_1}^n, B_{j_2}^n$ denote the generator matrices and bias vectors of $\Lambda_{j_1}, \Lambda_{j_2}$ respectively. For each $j = 2,3, a^{s_j} \in \mathcal{U}_j^{s_j}$, let $I_j(a^{s_j}) \in [\pi^{t_j}]$ denote the random bin to which codeword $U_j^n(a^{s_j}) := a^{s_j}G_j \oplus B_j^n$ is assigned. For $j = 2,3$, let $\mathcal{C}_j = (X_j^n(m_{jX}, b_{jX}) : (m_{jX}, b_{jX}) \in \mathcal{M}_{jX} \times c_{jX})$, where $c_{jX} := [\exp\{nK_j\}]$ denote user j's random satellite codebook. We let $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3$ $G_{j_1}, G_{j_2/j_1}, B_2^n, B_3^n$ and indices $I_j(a^{s_j}) : a^{s_j} \in \mathcal{U}_j^{s_j}, j = 2,3$ be mutually independent. Moreover, for $j = 1,2,3$, we let (i) the codewords in $\mathcal{C}_j$ be mutually independent and identically distributed according to $\prod_{t=1}^n p_{X_j}$ and (ii) generator matrices $G_{j_1}, G_{j_2}^t = \begin{bmatrix} G_{j_1}^t & G_{j_2/j_1}^t \end{bmatrix}$ and bias vectors $B_{j_1}^n, B_{j_2}^n$ be uniformly distributed over their respective range spaces, and (iii) random indices $I_j(a^{s_j}) : a^{s_j} \in \mathcal{U}_j^{s_j}, j = 2,3$ be uniformly distributed over their respective range spaces.

In the following proof we derive upper bounds on the parameters of the code and thereby characterize $\alpha_f^{3\text{-}1}(\underline{\tau})$. Here, we only provide a sketch of the arguments and indicate the upper bounds. The codewords of Λ_j are uniformly distributed and pairwise independent (Lemma A.0.1). An informed reader can now recognize that if

$$\frac{s_j - t_j}{n} \log \pi \overset{(a)}{>} \log \pi - H(U_j)\ ,\ K_j > 0\ ,\ \frac{(s_j - t_j)\log \pi}{n} + K_j > \log \pi + H(X_j) - H(X_j, U_j) \text{ for } j = 2,3, \quad (4.8)$$

then encoders 2 and 3 will find at least one pair of typical codewords in the indexed pair of bins. Decoders 2 and 3

perform point-to-point decoding of their cloud center and satellite codebooks. If

$$K_j + L_j \overset{(a)}{<} I(X_j; Y_j, U_j) \ , \ \frac{s_j \log \pi}{n} + K_j + L_j \overset{(b)}{<} \log \pi + H(X_j) - H(U_j, X_j|Y_j),$$
$$\frac{s_j \log \pi}{n} < \log \pi - H(U_j|X_j, Y_j) \tag{4.9}$$

then probability of error at decoder j decays exponentially with block length n.

$\Lambda_2 \oplus \Lambda_3$ is a codebook of rate $\max\{\frac{s_j \log \pi}{n} : j = 2, 3\}$ with codewords being uniformly distributed and pairwise independent. An informed reader will be able to reason that if

$$R_1 \overset{(a)}{<} I(X_1; Y_1, U_2 \oplus U_3) \ \text{ and } \ R_1 + \frac{s_j \log \pi}{n} \overset{(b)}{<} \log \pi + H(X_1) - H(X_1, U_2 \oplus U_3|Y_1) \text{ for } j = 2, 3 \tag{4.10}$$

then probability of decoding error at receiver 1 can be driven arbitrarily small by choosing a large enough block length n.

Substituting $R_j = \frac{t_j \log \pi}{n} + L_j$ for $j = 2, 3$, incorporating non-negativity constraints for $R_1, t_j, L_j : j = 2, 3$ and eliminating $\frac{s_j \log \pi}{n}, K_j : j = 2, 3$ using the technique of Fourier Motzkin [26, Appendix D] yields the achievable rate region mentioned in theorem 4.6.2. We formalize the above arguments through the following proof.

Proof: Let $p_{QU_2U_3\underline{XY}} \in \mathbb{D}_f(\underline{\tau})$, $\underline{R} \in \alpha_f^{3\text{-}1}(p_{QU_2U_3\underline{XY}})$ and $\tilde{\eta} > 0$. Let us assume $\mathcal{U}_2 = \mathcal{U}_3 = \mathcal{F}_\pi$ is the finite field of size π. For each $n \in \mathbb{N}$ sufficiently large, we prove existence of a 3−IC code $(n, \underline{\mathscr{M}}, \underline{e}, \underline{d})$ for which $\frac{\log \mathscr{M}_k}{n} \geq R_k - \tilde{\eta}$, $\tau_k(e_k) \leq \tau_k + \tilde{\eta}$ for $k \in [3]$ and $\overline{\xi}(\underline{e}, \underline{d}) \leq \tilde{\eta}$.

Taking a cue from the above coding technique, we begin with an alternative characterization of $\alpha_f^{3\text{-}1}(p_{QU_2U_3\underline{XY}})$ in terms of the parameters of the code.

Definition 4.6.4 *Consider $p_{QU_2U_3\underline{XY}} \in \mathbb{D}_f(\underline{\tau})$ and let $\mathcal{F}_\pi := \mathcal{U}_2 = \mathcal{U}_3$. Let $\tilde{\alpha}_f^{3\text{-}1}(p_{QU_2U_3\underline{XY}})$ be defined as the set of rate triples $(R_1, R_2, R_3) \in [0, \infty)^3$ for which $\underset{\delta>0}{\cup} \tilde{\mathcal{S}}(\underline{R}, p_{QU_2U_3\underline{XY}}, \delta, \delta)$ is non-empty, where $\tilde{\mathcal{S}}(\underline{R}, p_{QU_2U_3\underline{XY}}, \delta_S, \delta_C)$ is defined as the collection of vectors $(S_2, T_2, K_2, L_2, S_3, T_3, K_3, L_3) \in [0, \infty)^8$ that satisfy*

$$R_j = T_j \log \pi + L_j, \ K_j > \delta_S, \ \ (S_j - T_j) \log \pi > \log \pi - H(U_j|Q) + \delta_S,$$
$$(S_j - T_j) \log \pi + K_j > \log \pi + H(X_j|Q) - H(U_j, X_j|Q) + \delta_S$$
$$T_j > \delta_C, \ L_j > \delta_C, \ K_j + L_j < I(X_j; Y_j, U_j|Q) - \delta_C, \ \ S_j \log \pi < \log \pi - H(U_j|X_j, Y_j, Q) - \delta_C,$$
$$S_j \log \pi + K_j + L_j < \log \pi + H(X_j|Q) - H(U_j, X_j|Y_j, Q) - \delta_C, \ \ R_1 < I(X_1; Y_1, U_2 \oplus U_3|Q) - \delta_C$$
$$R_1 + S_j \log \pi < \log \pi + H(X_1|Q) - H(X_1, U_2 \oplus U_3|Y_1, Q) - \delta_C$$

for $j = 2, 3$.

Lemma 4.6.5 $\tilde{\alpha}_f^{3\text{-}1}(p_{QU_2U_3\underline{XY}}) = \alpha_f^{3\text{-}1}(p_{QU_2U_3\underline{XY}})$. □

Proof: The proof follows by substituting $R_j = T_j \log \pi + L_j$ in the bounds characterizing $\tilde{\mathcal{S}}(\underline{R}, p_{QU_2U_3\underline{XY}}, 0, 0)$ and eliminating $S_j, T_j, K_j, L_j : j = 2, 3$ via the technique of Fourier Motzkin. The resulting characterization will be that of $\alpha_f^{3\text{-}1}(p_{QU_2U_3\underline{XY}})$. The presence of strict inequalities in the bounds characterizing $\alpha_f^{3\text{-}1}(p_{QU_2U_3\underline{XY}})$ and $\tilde{\mathcal{S}}(\underline{R}, p_{QU_2U_3\underline{XY}}, \delta, \delta)$ enables one prove $\underset{\delta>0}{\cup} \tilde{\mathcal{S}}(\underline{R}, p_{QU_2U_3\underline{XY}}, \delta, \delta)$ is non-empty for every $\underline{R} \in \alpha_f^{3\text{-}1}(p_{QU_2U_3\underline{XY}})$.[13] ■

Lemma 4.6.5 provides us with the parameters of the code whose existence we seek to prove. Let us now describe this code. For the rate triple $\underline{R}$ under consideration, Lemma 4.6.5 provides us with $\delta > 0$ and $(S_j, T_j, K_j, L_j, : j = 2, 3) \in \mathcal{S}(\underline{R}, p_{QU_2U_3\underline{X},\underline{Y}}, \delta)$. Define $\eta = \min\{\delta, \tilde{\eta}\}, \eta_1 = \eta_2 = \frac{\eta}{2^d}$, where $d \in \mathbb{N}$ will be specified in due course. Let $q^n \in T_{\eta_2}(Q)$ denote the time sharing sequence. User 1's code contains $\lceil \exp\{nR_1\} \rceil$ codewords $(x_1^n(m_1) \in \mathcal{X}_1^n : m_1 \in \mathcal{M}_1)$, where $\mathcal{M}_1 := [\lceil \exp\{nR_1\} \rceil]$. The structure of user 2 and 3's codebooks are identical and we describe it using the generic index $j \in \{2, 3\}$. User j's cloud center codebook λ_j is the partitioned coset code (definition 4.6.3) $(n, s_j, t_j, g_j, b_j^n, i_j)$ built over $\mathcal{U}_j^n = \mathcal{F}_\pi^n$ where $s_j := \lceil nS_j \rceil$ and $t_j := \lceil nT_j \rceil$. We let $u_j^n(a^{s_j}) := a^{s_j} g_j \oplus b_j^n$ denote a generic codeword in λ_j and $c_{j1}(m_{j1}) := \{a^{s_j} : i_j(a^{s_j}) = m_{j1}\}$ denote the indices of codewords in bin corresponding to message $m_{j1} \in \mathcal{M}_{j1}$. Moreover, the partitioned coset codes *overlap*, i.e., if $s_{j_1} \leq s_{j_2}$, then $g_{j_2}^T = [g_{j_1}^T \;\; g_{j_2/j_1}^T]$. Without loss of generality, we assume $s_2 \leq s_3$ and therefore $g_3^T = [g_2^T \;\; g_{3/2}^T]$ The satellite codebook $\mathcal{C}_j$, built over $\mathcal{X}_j$, contains $\lceil \exp\{nL_j\} \rceil$ bins, one for each message $m_{jX} \in \mathcal{M}_{jX} := [\lceil \exp\{nL_j\} \rceil]$. Each bin contains $\lfloor \exp\{nK_j\} \rfloor$ codewords. We let $c_{jX} := [\lfloor \exp\{nK_j\} \rfloor]$ denote the set of bin indices and thereby $(x_j^n(m_{jX}, b_{jX}) : b_{jX} \in c_{jX})$ denotes bin corresponding to message $m_{jX} \in \mathcal{M}_{jX}$. The following remarks on the parameters of the codebooks are in order. Define $\eta_3 = \frac{\eta}{2^d}$ and note that, for all $n \geq N_1(\eta_3) := \max\left\{\lceil \frac{\log 2}{\eta_3} \rceil, \lceil \frac{1}{\eta_3} \rceil\right\}$

$$nR_1 \leq \log |\mathcal{M}_1| \leq n(R_1 + \eta_3)$$

$$S_j \leq \tfrac{s_j}{n} \leq S_j + \eta_3, \quad nL_j \leq \log |\mathcal{M}_{jX}| \leq n(L_j + \eta_3) \text{ for } j = 2, 3 \tag{4.11}$$

$$T_j \leq \tfrac{t_j}{n} \leq T_j + \eta_3, \quad n(K_j - \eta_3) \leq \log |c_{jX}| \leq nK_j \text{ for } j = 2, 3 \tag{4.12}$$

$$\tag{4.13}$$

We now specify encoding rules. Encoder 1 feeds codeword $x_1^n(M_1)$ indexed by the message as input. For $j = 2, 3$, encoder j populates

$$\mathcal{L}_j(M_j) := \{(u_j^n(a^{s_j}), x_j^n(M_{jX}, b_{jX})) \in T_{2\eta_2}(U_j, X_j | q^n) : (a^{s_j}, b_{jX}) \in c_{j1}(M_{j1}) \times c_{jX}\}.$$

If $\mathcal{L}_j(M_j)$ is non-empty, one of these pairs is chosen. Otherwise, one pair from $\lambda_j \times \mathcal{C}_j$ is chosen. Let $(U_j^n(A^{s_j}), X_j^n(M_{jX}, B_{jX}))$ denote the chosen pair. $X_j^n(M_{jX}, B_{jX})$ is fed as input on the channel.

Decoder 1 constructs the sum $\lambda_2 \oplus \lambda_3 := \left\{u_2^n \oplus u_3^n : u_j^n \in \lambda_j, j = 2, 3\right\}$ of the cloud center codebooks. Let

[13] Indeed, substituting $S_j = \frac{s_j}{n}, T_j = \frac{t_j}{n} : j = 2, 3$, one can identify the bounds in the definition of $\tilde{\mathcal{S}}(\underline{R}, p_{QU_2U_3\underline{XY}}, 0, 0)$ with those in (4.8), (4.9) and (4.10). As indicated then, the proof follows from the technique of Fourier-Motzkin [26, Appendix D].

$u^n_{\oplus}(a^{s_3}) := a^{s_3}g_3 \oplus b^n_2 \oplus b^n_3$ denote a generic codeword in $\lambda_2 \oplus \lambda_3$. Note that $\lambda_2 \oplus \lambda_3 = \left\{u^n_{\oplus}(a^{s_3}) : a^{s_3} \in \mathcal{U}^{s_3}_3\right\}$.[14] Having received Y^n_1, it looks for all potential message $\hat{m}_1$ for which there exists a $u^n_{2\oplus 3} \in \lambda_2 \oplus \lambda_3$ such that $(q^n, u^n_{2\oplus 3}, x^n_1(\hat{m}_1), Y^n_1) \in T_{4\eta_4}(Q, U_2 \oplus U_3, X_1, Y_1)$[15]. If it finds exactly one such message $\hat{m}_1$, it declares this as decoded message of user 1. Otherwise, it declares an error.

For $j \in \{2,3\}$, decoder j identifies all $(\hat{m}_{j1}, \hat{m}_{jX})$ for which there exists $(a^{s_j}, b_{jX}) \in c_{j1}(\hat{m}_{j1}) \times c_{jX}$ such that $(q^n, u^n_j(a^{s_j}), x^n_j(\hat{m}_{jX}, b_{jX}), Y^n_j) \in T_{4\eta_4}(Q, U_j, X_j, Y_j)$, where Y^n_j is the received vector. If there is exactly one such pair $(\hat{m}_{j1}, \hat{m}_{jX})$, this is declared as user j message. Otherwise an error is signaled.

The above encoding and decoding rules map every quintuple of codes $(\mathcal{C}_1, \lambda_2, \lambda_3, \mathcal{C}_2, \mathcal{C}_3)$ into a corresponding 3−IC code $(n, \underline{\mathscr{M}}, \underline{e}, \underline{d})$ of rate $\frac{\log \mathscr{M}_1}{n} \geq R_1, \frac{\log \mathscr{M}_j}{n} \geq \frac{t_j}{n}\log\pi + L_j \geq T_j \log\pi + L_j = R_j : j \in \{2,3\}$, thus characterizing an ensemble of 3−IC codes, one for each $n \in \mathbb{N}$,. We average error probability over this ensemble of 3−IC codes by letting (i) the codewords of $\mathcal{C}_1 := (X^n_1(m_1) : m_1 \in \mathcal{M}_1)$, generator matrices $G_2, G_{3/2}$[16], bias vectors B^n_1, B^n_2, bin indices $(I_j(a^{s_j}) : a^{s_j} \in \mathcal{U}^{s_j}_j) : j = 2,3$ and codewords of $\mathcal{C}_j = (X^n_j(m_{jX}, b_{jX}) : (m_{jX}, b_{jX}) \in \mathcal{M}_{jX} \times c_{jX}) : j = 2,3$ be mutually independent, (ii) the codewords of $\mathcal{C}_j : j = 1,2,3$ are identically distributed according to $\prod^n_{t=1} p_{X_j|Q}(\cdot|q_t)$, (iii) generator matrices $G_{j_1}, G_{j_2/j_1}$, bias vectors B^n_1, B^n_2, bin indices $(I_j(a^{s_j}) : a^{s_j} \in \mathcal{U}^{s_j}_j) : j = 2,3$ be uniformly distributed over their respective range spaces. We denote the random partitioned coset code $(n, s_j, t_j, G_j, B^n_j, I_j)$ of user j as Λ_j and let (i) $U^n_j(a^{s_j}) := a^{s_j}G_j \oplus B^n_j$ denote a generic random codeword in Λ_j, (ii) $U^n_{\oplus}(a^{s_3}) := a^{s_3}G_3 \oplus B^n_2 \oplus B^n_3$ denote a generic codeword in $\Lambda_2 \oplus \Lambda_3$, and (iii) $C_{j1}(m_{j1}) = \{a^{s_j} \in \mathcal{U}^{s_j}_j : i_j(a^{s_j}) = m_{j1}\}$ denote the random collection of indices corresponding to message M_{j1}.

We now proceed towards deriving an upper bound on the probability of error. Towards that end, we begin with a characterization of error events. Let

$$
\begin{aligned}
\epsilon_{1j} &:= \bigcap_{\substack{(a^{s_j}, b_{jX}) \in \\ C_{j1}(M_{j1}) \times c_{jX}}} \left\{(q^n, U_j(a^{s_j}), X_j(M_{jX}, b_{jX})) \notin T_{2\eta_2}(Q, U_j, X_j)\right\}, \text{ for } j = 2,3 \\
\epsilon_{11} &:= \left\{(q^n, X^n_1(M_1)) \notin T_{2\eta_2}(Q, X_1)\right\} \quad , \quad \epsilon_{31} := \left\{(q^n, X^n_1(M_1), Y^n_1) \notin T_{2\eta_4}(Q, X_1, Y_1)\right\} \\
\epsilon_{3j} &:= \bigcap_{\substack{(a^{s_j}, b_{jX}) \in \\ C_{j1}(M_{j1}) \times c_{jX}}} \left\{(q^n, U_j(a^{s_j}), X_j(M_{jX}, b_{jX}), Y^n_j) \notin T_{2\eta_4}(Q, U_j, V_j, Y_j)\right\}, \text{ for } j = 2,3 \\
\epsilon_{41} &:= \bigcup_{\hat{m}_1 \neq M_1} \bigcup_{a^{s_3} \in \mathcal{U}^{s_3}_3} \left\{(q^n, U^n_{\oplus}(a^{s_3}), X^n_1(\hat{m}_1), Y^n_1) \in T_{4\eta_4}(Q, U_2 \oplus U_3, X_1, Y_1)\right\} \\
\epsilon_{4j} &:= \bigcup_{\hat{m}_j \neq M_j} \bigcup_{\substack{a^{s_j} \in \\ C_{j1}(\hat{m}_{j1})}} \bigcup_{b_{jX} \in c_{jX}} \left\{(q^n, U_j(a^{s_j}), X_j(\hat{m}_{jX}, b_{jX}), Y^n_j) \in T_{4\eta_4}(Q, U_j, V_j, Y_j)\right\} \text{ for } j = 2,3.
\end{aligned}
$$

[14]Here we have used the assumption $s_2 \leq s_3$. In general, if $s_{j_1} \leq s_{j_2}$, we have $\lambda_2 \oplus \lambda_3 = \left\{u^n_{\oplus}(a^{s_{j_2}}) : a^{s_{j_2}} \in \mathcal{U}^{s_{j_2}}_{j_2}\right\}$, where $u^n_{\oplus}(a^{s_{j_2}}) := a^{s_{j_2}}g_{j_2} \oplus b^n_2 \oplus b^n_3$ denotes a generic codeword.

[15]The value of η_4 is specified in due course.

[16]Recall, that we have assumed $s_2 \leq s_3$.

Note that $\epsilon := \bigcup_{j=1}^{3} (\epsilon_{1j} \cup \epsilon_{3j} \cup \epsilon_{4j})$ contains the error event. We derive an upper bound on the probability of this event by partitioning it appropriately. The following events will aid us identify such a partition. Define $\epsilon_l := \epsilon_{l_2} \cup \epsilon_{l_3}$, where

$$\epsilon_{l_j} := \{\phi_j(q^n, M_j) < \mathcal{L}_j(n)\}, \text{ and } \phi_j(q^n, M_j) := \sum_{\substack{(a^{s_j}, b_{jX}) \in \\ C_{j1}(M_{j1}) \times c_{jX}}} 1_{\left\{(q^n, U_j(a^{s_j}), X_j(M_{jX}, b_{jX})) \in T_{2\eta_2}(Q, U_j, X_j)\right\}}.$$

$\mathcal{L}_j(n)$ is half of the expected number of jointly typical pairs in the indexed pair of bins.[17] Let

$$\epsilon_1 := \bigcup_{j=2}^{3} \left\{(q^n, U_j^n(A^{s_j}), X_j^n(M_{jX}, B_{jX})) \notin T_{2\eta_2}(Q, U_j, X_j)\right\} \bigcup \left\{(q^n, X_1^n(M_1)) \notin T_{2\eta_2}(Q, X_1)\right\}, \quad (4.14)$$

$$\epsilon_2 := \{(q^n, U_2^n(A^{s_2}), U_3^n(A^{s_3}), X_1^n(M_1), X_2^n(M_{2X}, B_{2X}), X_3^n(M_{3X}, B_{3X})) \notin T_{\eta_4}(Q, U_2, U_3, \underline{X})\} \quad (4.15)$$

$$\epsilon_3 := \{(q^n, U_2^n(A^{s_2}), U_3^n(A^{s_3}), X_1^n(M_1), X_2^n(M_{2X}, B_{2X}), X_3^n(M_{3X}, B_{3X}), \underline{Y}^n) \notin T_{2\eta_4}(Q, X_1, U_2, U_3, \underline{X}, \underline{Y})\}. \quad (4.16)$$

For sufficiently large $n \in \mathbb{N}$, we prove $\mathcal{L}_j(n) > 2$. For such an n, $\epsilon_{1j} \subseteq \epsilon_{l_j} : j = 2, 3$. Since, we can choose n sufficiently large, we will henceforth assume $\epsilon_{1j} \subseteq \epsilon_{l_j} : j = 2, 3$. Therefore, the error event $\epsilon \subseteq \bigcup_{j=1}^{3} (\epsilon_{11} \cup \tilde{\epsilon}_1 \cup \epsilon_{3j} \cup \epsilon_{4j})$ where $\tilde{\epsilon}_1 := \epsilon_1 \cup \epsilon_l$.[18] From the encoding rule, we have $(\epsilon_{11} \cup \epsilon_l)^c \subseteq \epsilon_1^c$, and hence $P((\epsilon_{11} \cup \epsilon_l)^c \cap \epsilon_1) = 0$. Moreover, $(\tilde{\epsilon}_1 \cup \epsilon_2)^c \cap \bigcup_{j=1}^{3} \epsilon_{3j} \subseteq (\tilde{\epsilon}_1 \cup \epsilon_2)^c \cap \epsilon_3$. It therefore suffices to derive upper bounds on $P(\epsilon_{11}), P(\epsilon_{l_j}) : j = 2, 3$, $P(\tilde{\epsilon}_1^c \cap \epsilon_2)$, $P((\tilde{\epsilon}_1 \cup \epsilon_2)^c \cap \epsilon_3)$ and $P((\tilde{\epsilon}_1 \cup \epsilon_3)^c \cap \epsilon_{4j}) : j = 1, 2, 3$.

Upper bound on $P(\epsilon_{11})$:– By lemma 2.4.1, there exists $N_2(\eta_2) \in \mathbb{N}$, such that for all $n \geq N_2(\eta_2)$, $P(\epsilon_{11}) \leq \frac{\eta}{32}$.

Upper bound on $P(\epsilon_{l_j})$:– Using a second moment method similar to that employed in [47, Appendix A], we derive an upper bound on $P(\epsilon_{l_j})$ in appendix C. In particular, we prove existence of $N_5(\eta) \in \mathbb{N}$ such that for all $n \geq N_5(\eta)$

$$P(\epsilon_{1_j}) \leq 12 \exp\left\{-n\left(\delta - \frac{\eta\,[36 + \log \pi]}{2^d}\right)\right\}. \quad (4.17)$$

In deriving the above upper bound, we employed, among others, the bounds

$$K_j \geq \delta > 0, \quad (S_j - T_j) \log \pi - [\log \pi - H(U_j|Q)] \geq \delta > 0$$
$$(S_j - T_j) \log \pi + K_j - [\log \pi + H(X_j|Q) - H(U_j, X_j|Q)] \geq \delta > 0.$$

Upper bounds on $P(\tilde{\epsilon}_1^c \cap \epsilon_2)$, $P((\tilde{\epsilon}_1 \cup \epsilon_2)^c \cap \epsilon_3)$:– These events are related to the following two events. (i) The codewords chosen by the distributed encoders are *not* jointly typical, and (ii) the channel produces a triple of outputs that is *not* jointly typical with the chosen and input codewords. In deriving an upper bound on $P(\tilde{\epsilon}_1^c \cap \epsilon_2)$,

[17] Since the precise value of $\mathcal{L}_j(n)$ is necessary only in the derivation of the upper bound, it is provided in appendix C.
[18] The reader will note that we have included ϵ_1 on the right hand side.

$P((\tilde{\epsilon}_1 \cup \epsilon_2)^c \cap \epsilon_3)$, we employ (i) conditional mutual independence of the triplet $X_1, (U_j, X_j) : j = 2, 3$ given Q and (ii) the Markov chain $(U_j : j = 2, 3) - \underline{X} - \underline{Y}$. For a technique based on unstructured and independent code, the analysis of this event is quite standard. However, since our coding technique relies on codewords chosen from statistically correlated codebooks, we present the steps in deriving an upper bound in appendix D. In particular, we prove existence of $N_6(\eta_4), N_8(\eta_4) \in \mathbb{N}$, such that for $n \geq \max\{N_1(\eta_3), N_6(\eta_4), N_8(\eta_4)\}$,

$$P(\tilde{\epsilon}_1^c \cap \epsilon_2) + P((\tilde{\epsilon}_1 \cup \epsilon_2)^c \cap \epsilon_3) \leq 2\exp\{-n\left(n^2\mu\eta_4^2 - \frac{\eta}{2^{d-5}}\right)\} + \frac{\eta}{32} \tag{4.18}$$

Upper bound on $P((\tilde{\epsilon}_1 \cup \epsilon_2 \cup \epsilon_3)^c \cap \epsilon_{41})$:– In appendix E equation (E.9), we prove existence of $N_{10}(\eta_4) \in \mathbb{N}$ such that for all $n \geq \max\{N_1(\eta_3), N_9(\eta_4), N_{10}(\eta_4)\}$ we have

$$P((\tilde{\epsilon}_1 \cup \epsilon_2 \cup \epsilon_3)^c \cap \epsilon_{41}) \leq 4\exp\left\{-n\left[\delta - 28\eta_4 - \frac{\eta(13+\log\pi)}{2^d}\right]\right\}. \tag{4.19}$$

In deriving (4.19), we employed, among others, the bounds

$$\log\pi + H(X_1|Q) - H(X_1, U_2 \oplus U_3|Y_1, Q) - (R_1 + \max\{S_2, S_3\}\log\pi) \geq \delta > 0, \quad I(X_1; Y_1, U_2 \oplus U_3|Q) - R_1 \geq \delta > 0.$$

Upper bound on $P((\tilde{\epsilon}_1 \cup \epsilon_3)^c \cap \epsilon_{4j})$:– In appendix F equation (F.11), we prove existence of $N(\eta) \in \mathbb{N}$, such that for all $n \geq N(\eta)$

$$P((\tilde{\epsilon}_1 \cup \epsilon_2 \cup \epsilon_3)^c \cap \epsilon_{4j}) \leq 10\exp\left\{-n\left(\delta - \left(\frac{\eta(9+\log\pi)}{2^d} + 16\eta_4\right)\right)\right\}. \tag{4.20}$$

In deriving (4.20), we employed, among others, the bounds

$$(\log\pi - H(U_j|X_j, Y_j, Q)) - S_j\log\pi \geq \delta > 0, \quad (\log\pi + H(X_j|Q) - H(U_j, X_j|Y_j, Q)) - (S_j\log\pi + K_j) \geq \delta > 0,$$
$$I(X_1; Y_1, U_2 \oplus U_3|Q) - R_1 \geq \delta > 0, \quad (I(X_j; U_j, Y_j|Q)) - (K_j + L_j) \geq \delta > 0,$$
$$(\log\pi + H(X_j|Q) - H(X_j, U_j|Y_j, Q)) - (K_j + L_j + S_j\log\pi) \geq \delta > 0$$

We now collect the derived upper bounds. From (4.17), (4.18), (4.19) and (4.20), we have

$$\begin{aligned}
P(\bigcup_{j=1}^{3}(\epsilon_{1j} \cup \epsilon_{3j} \cup \epsilon_{4j})) \leq\ & \frac{\eta}{32} + 3\exp\left\{-n\left(\delta - \frac{\eta\,[36+\log\pi]}{2^d}\right)\right\} + 2\exp\left\{-n\left(n^2\mu\eta_4^2 - \frac{\eta(18+2\log\pi)}{2^d}\right)\right\} + \frac{\eta_4}{32} \\
& + 2\exp\left\{-n\left(\delta - 30\eta_4 - \frac{\eta(1+\log\pi)}{2^{d-1}}\right)\right\} + 5\exp\left\{-n\left(\delta - \left(\frac{\eta(13+2\log\pi)}{2^d} + 16\eta_4\right)\right)\right\} \\
\leq\ & 10\exp\left\{-n\left(\delta - \left(\frac{\eta(36+2\log\pi)}{2^d} + 30\eta_4\right)\right)\right\} + 2\exp\{-n\left(n^2\mu\eta_4^2 - \frac{\eta(18+2\log\pi)}{2^d}\right)\} \\
& + \frac{\eta+\eta_4}{32}
\end{aligned}$$

We also recall, that $\eta_4 \geq 4\eta_2$ for (D.11), (4.19) and (4.20) to hold true. If we are able to find η_4 that satisfies

$$\frac{1}{30}\left[\delta - \frac{\eta(36+2\log\pi)}{2^d}\right] > \eta_4 > \max\left\{4\eta_2 = \frac{\eta}{2^{d-2}}, \sqrt{\frac{\eta}{\mu 2^d}(36+2\log\pi)}\right\}, \tag{4.21}$$

then we can choose n sufficiently large enough to drive down $P(\overset{3}{\underset{j=1}{\cup}}(\epsilon_{1j} \cup \epsilon_{3j} \cup \epsilon_{4j}))$. Recall that $\delta = \min\{\delta_2, \delta_3\}$ is a function of the rate triple $\underline{R}$ and $\eta = \min\{\delta, \tilde{\eta}\}$ are not affected by the choice of d. Clearly, by choosing d large enough, the upper bound in (4.21) can be made sufficiently close to $\frac{\delta}{30}$ and the lower bound can be made sufficiently close 0 permitting a range of values for η_4. This completes derivation of an upper bound on the probability of error.

We only need to argue that the chosen input codewords satisfy the cost constraint. For sufficiently large n, we have proved $P(\epsilon_2) \leq \frac{\eta}{16}$. Since ϵ_2^c implies that chosen input codewords are jointly typical with respect to $p_{QU_2U_3\underline{XY}}$, a distribution that satisfies $\mathbb{E}\{\kappa_j(X_j)\} \leq \tau_j$. Using standard typicality arguments and finiteness of $\max\{\kappa_k(x_k) : x_k \in \mathcal{X}_k : k \in [3]\}$, it is straight forward to show that the average cost of the codeword input by encoder j is close to τ^j per symbol. ■

The coding technique proposed in the proof of theorem 4.6.2 is indeed a generalization of that proposed for example 4.5.1, and moreover capacity achieving for the same. We formalize this through the following corollary.

Corollary 4.6.6 *For the $3-to-1$ IC in example 4.5.1, if $\tau * \delta_1 < \min\{\delta_2, \delta_3\}$, then $\alpha_f^{3-1}(\tau, \frac{1}{2}, \frac{1}{2}) = \mathbb{C}(\tau)$. Moreover, if $\delta := \delta_2 = \delta_3$ and $h_b(\tau * \delta_1) \leq h_b(\delta) < \frac{1+h_b(\delta_1 * \tau)}{2}$, then $\alpha_u(\tau, \frac{1}{2}, \frac{1}{2}) \neq \mathbb{C}(\tau)$ and $\mathbb{C}(\tau) = \alpha_f^{3-1}(\tau, \frac{1}{2}, \frac{1}{2})$.*

It can be verified that $\beta(\tau, \frac{1}{2}, \frac{1}{2}, \underline{\delta}) = \alpha_f^{3-1}(p_{QU_2U_3\underline{XY}})$ where $P(U_j = X_j = 0) = P(U_j = X_j = 1) = \frac{1}{2}$, $P(X_1 = 1) = \tau$ and $\mathcal{Q} = \phi$, the empty set, where $\beta(\underline{\tau}, \underline{\delta})$ is given in (4.3).

In the sequel, we illustrate through an example the central claim of this thesis that utility of codes endowed with algebraic structure, and in particular coset codes, are not restricted to particular symmetric and additive problems. Furthermore, this example establishes the need (i) to achieve rates corresponding to non-uniform distributions which is accomplished via the technique of binning, (ii) to build coset codes over larger fields, and (iii) to analyze decoding of sums of transmitted codewords over arbitrary channels which hinges on typical set decoding.

Example 4.6.7 *Consider a binary $3-to-1$ IC illustrated in figure 4.3 with $\mathcal{X}_j = \mathcal{Y}_j = \{0,1\} : j \in [3]$ with channel transition probabilities $W_{\underline{Y}|\underline{X}}(\underline{y}|\underline{x}) = BSC_{\delta_1}(y_1|x_1 \oplus (x_2 \vee x_3))BSC_{\delta_2}(y_2|x_2)BSC_{\delta_3}(y_3|x_3)$, where $\vee$ denotes logical OR and $BSC_\eta(0|1) = BSC_\eta(1|0) = 1 - BSC_\eta(0|0) = 1 - BSC_\eta(1|1) = \eta$ denotes the transition probabilities of a binary symmetric channel (BSC) with cross over probability $\eta \in [0, \frac{1}{2}]$. Users' inputs are constrained with respect to a Hamming cost function, i.e., $\kappa_j(x) = x$ for $x \in \{0,1\}$. Assume user jth input is constrained to an average cost per symbol of $\tau_j \in (0, \frac{1}{2})$.*

We begin by stating the conditions for sub-optimality of $\mathscr{USB}-$technique.

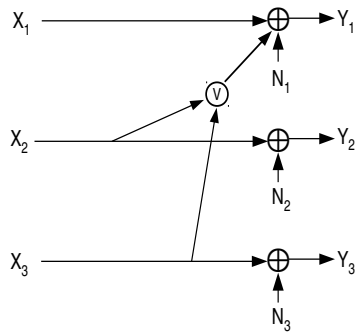

Figure 4.3: A binary 3−to−1 IC described in example 4.6.7.

Lemma 4.6.8 *Consider example 4.6.7 with* $\delta := \delta_2 = \delta_3 \in (0, \frac{1}{2})$ *and* $\tau := \tau_2 = \tau_3 \in (0, \frac{1}{2})$. *Let* $\beta := (1-\tau)^2\delta_1 + (2\tau - \tau^2)(1-\delta_1)$. *The rate triple* $(h_b(\tau_1 * \delta_1) - h_b(\delta_1), h_b(\tau * \delta) - h_b(\delta), h_b(\tau * \delta) - h_b(\delta)) \notin \alpha_u(\underline{\tau})$ *if*

$$h_b(\tau_1 * \delta_1) - h_b(\delta_1) + 2(h_b(\tau * \delta) - h_b(\delta)) > h_b(\tau_1(1-\beta) + (1-\tau_1)\beta) - h_b(\delta_1) \tag{4.22}$$

In particular, if (4.22) is true, $\alpha_u(\underline{\tau}) \subsetneq \beta(\underline{\tau}, \underline{\delta})$, *where* $\beta(\underline{\tau}, \underline{\delta})$ *is defined in (4.3).* □

Proof: We prove this by contradiction. Suppose $(h_b(\tau_1 * \delta_1) - h_b(\delta_1), h_b(\tau * \delta) - h_b(\delta), h_b(\tau * \delta) - h_b(\delta)) \in \alpha_f^{3\text{-}1}(p_{QU_2U_3\underline{XY}})$ for some $p_{QU_2U_3\underline{XY}} \in \mathbb{D}_{3-1}(\tau_1, \tau, \tau)$. Our first claim is that $p_{X_2|Q}(1|q) = p_{X_3|Q}(1|q) = \tau$ for all $q \in \mathcal{Q}$.

From (4.1) we have

$$R_j \leq I(U_jX_j; Y_j|Q) = H(Y_j|Q) - H(Y_j|X_jU_jQ) = H(Y_j|Q) - h_b(\delta) = \sum_{q\in\mathcal{Q}} p_Q(q)H(Y_j|Q=q) - h_b(\delta)$$
$$= \sum_{q\in\mathcal{Q}} p_Q(q)H(X_j \oplus N_j|Q=q) - h_b(\delta) \text{ for } j = 2, 3. \tag{4.23}$$

If $\tau_q := p_{X_j|Q}(1|q)$, then independence of the pair N_j and (X_j, Q) implies $p_{X_j\oplus N_j|Q}(1|q) = \tau_q(1-\delta) + (1-\tau_q)\delta = \tau_q(1-2\delta) + \delta$. Substituting the same in (4.23), we have

$$R_j \leq \sum_{q\in\mathcal{Q}} p_Q(q)h_b(\tau_q(1-2\delta)+\delta) - h_b(\delta) \leq h_b(\sum_{q\in\mathcal{Q}} p_Q(q)[\tau_q(1-2\delta)+\delta]) - h_b(\delta)$$
$$= h_b([p_{X_j}(1)(1-2\delta)+\delta]) - h_b(\delta)$$

from Jensen's inequality. Since $p_{X_j}(1) \leq \tau < \frac{1}{2}$, we have $p_{X_j}(1)(1-2\delta) + \delta \leq \tau(1-2\delta) + \delta < \frac{1}{2}(1-2\delta) + \delta = \frac{1}{2}$.[19] The term $h_b([p_{X_j}(1)(1-2\delta)+\delta])$ is therefore strictly increasing in $p_{X_j}(1)$ and is at most $h_b(\tau * \delta)$.[20] Moreover, the

[19] Here we have used the positivity of $(1-2\delta)$, or equivalently δ being in the range $(0, \frac{1}{2})$.
[20] This is consequence of $p_{X_j}(1) \leq \tau$.

condition for equality in Jensen's inequality implies $R_j = h_b(\tau * \delta) - h_b(\delta)$ if and only if $p_{X_j|Q}(1|q) = \tau$ for all $q \in \mathcal{Q}$ that satisfies $p_Q(q) > 0$. We have therefore proved our first claim.

Our second claim is an analogous statement for $p_{X_1|Q}(1|q)$. In particular, our second claim is that $p_{X_1|Q}(1|q) = \tau_1$ for each $q \in \mathcal{Q}$ of positive probability. We begin with the upper bound on R_1 in (4.1). As in proof of theorem 4.5.3, we let $\tilde{\mathcal{Q}} := \mathcal{Q} \times \mathcal{U}_2 \times \mathcal{U}_3$, $\tilde{q} = (q, u_2, u_3) \in \tilde{\mathcal{Q}}$ denote a generic element and $\tilde{Q} := (Q, U_2, U_3)$. The steps we employ in proving the second claim borrows steps from proof of theorem 4.5.3 and the proof of the first claim presented above. Note that

$$
\begin{aligned}
R_1 \leq I(X_1; Y_1|\tilde{Q}) &= H(Y_1|\tilde{Q}) - H(Y_1|\tilde{Q}X_1) \\
&= \sum_{\tilde{q}} p_{\tilde{Q}}(\tilde{q}) H(Y_1|\tilde{Q} = \tilde{q}) - \sum_{x_1, \tilde{q}} p_{\tilde{Q}X_1}(\tilde{q}, x_1) H(Y_1|X_1 = x_1, \tilde{Q} = \tilde{q}) \\
&= \sum_{\tilde{q}} p_{\tilde{Q}}(\tilde{q}) H(X_1 \oplus N_1 \oplus (X_2 \vee X_3)|\tilde{Q} = \tilde{q}) - \sum_{x_1, \tilde{q}} p_{X_1\tilde{Q}}(x_1, \tilde{q}) H(x_1 \oplus N_1 \oplus (X_2 \vee X_3)|X_1 = x_1, \tilde{Q} = \tilde{q}) \\
&= \sum_{\tilde{q}} p_{\tilde{Q}}(\tilde{q}) H(X_1 \oplus N_1 \oplus (X_2 \vee X_3)|\tilde{Q} = \tilde{q}) - \sum_{x_1, \tilde{q}} p_{X_1\tilde{Q}}(x_1, \tilde{q}) H(N_1 \oplus (X_2 \vee X_3)|\tilde{Q} = \tilde{q}) \qquad (4.24) \\
&= \sum_{\tilde{q}} p_{\tilde{Q}}(\tilde{q}) H(X_1 \oplus N_1 \oplus (X_2 \vee X_3)|\tilde{Q} = \tilde{q}) - \sum_{\tilde{q}} p_{\tilde{Q}}(\tilde{q}) H(N_1 \oplus (X_2 \vee X_3)|\tilde{Q} = \tilde{q}) \\
&\leq \sum_{\tilde{q}} p_{\tilde{Q}}(\tilde{q}) H(X_1 \oplus N_1|\tilde{Q} = \tilde{q}) - \sum_{\tilde{q}} p_{\tilde{Q}}(\tilde{q}) H(N_1|\tilde{Q} = \tilde{q}) = \sum_{q} p_{\tilde{Q}}(\tilde{q}) H(X_1 \oplus N_1|\tilde{Q} = \tilde{q}) - h_b(\delta_1) \qquad (4.25) \\
&= \sum_{\tilde{q}} p_{\tilde{Q}}(\tilde{q}) h_b(\tau_{1\tilde{q}} * \delta_1) - h_b(\delta_1) \leq h_b(\mathbb{E}_{\tilde{Q}} \{\tau_{1\tilde{q}} * \delta_1\}) - h_b(\delta_1) = h_b(p_{X_1}(1) * \delta_1) - h_b(\delta_1), \qquad (4.26)
\end{aligned}
$$

where (i) (4.24) follows from independence of (N_1, X_2, X_3) and X_1 conditioned on realization of $\tilde{Q}$, (ii) (4.25) follows from substituting $p_{X_1 \oplus N_1|\tilde{Q}}(\cdot|\tilde{q})$ for p_{Z_1}, $p_{N_1|\tilde{Q}}(\cdot|\tilde{q})$ for p_{Z_2} and $p_{X_2 \vee X_3|\tilde{Q}}(\cdot|\tilde{q})$ for p_{Z_3} in lemma 4.5.4, (iii) the first inequality in (4.26) follows from Jensen's inequality. Since $p_{X_1}(1) \leq \tau_1 < \frac{1}{2}$, we have $p_{X_1}(1) * \delta_1 = p_{X_1}(1-\delta_1) + (1 - p_{X_1}(1))\delta_1 = p_{X_1}(1)(1-2\delta_1) + \delta_1 \leq \tau_1(1-2\delta_1) + \delta_1 \leq \frac{1}{2}(1-2\delta_1) + \delta_1 = \frac{1}{2}$. Therefore $h_b(p_{X_1}(1) * \delta_1)$ is increasing[21] in $p_{X_1}(1)$ and is bounded above by $h_b(\tau_1 * \delta_1)$.[22] Moreover, the condition for equality in Jensen's inequality implies $R_1 = h_b(\tau_1 * \delta_1) - h_b(\delta_1)$ if and only if $p_{X_1|\tilde{Q}}(1|\tilde{q}) = \tau_1$ for all $\tilde{q} \in \tilde{\mathcal{Q}}$. We have therefore proved our second claim.[23]

Our third claim is that either $H(X_2|Q, U_2) > 0$ or $H(X_3|Q, U_3) > 0$. Suppose not, i.e., $H(X_2|Q, U_2) =$

[21] This also employs the positivity of $1 - 2\delta_1$, or equivalently δ_1 being in the range $(0, \frac{1}{2})$.

[22] This is consequence of $p_{X_1}(1) \leq \tau_1$.

[23] We have only proved $p_{X_1|QU_2U_3}(1|q, u_2, u_3 = \tau_1)$ for all $(q, u_2, u_3) \in \mathcal{Q} \times \mathcal{U}_2 \times \mathcal{U}_3$ of positive probability. The claim now follows from conditional independence of X_1 and U_2, U_3 given Q.

$H(X_3|Q,U_3)=0$. In this case, the upper bound on $R_1+R_2+R_3$ in (4.2) is

$$\begin{aligned} R_1+R_2+R_3 &\leq I(X_2,X_3,X_1;Y_1|Q)=H(Y_1|Q)-H(Y_1|Q,X_1,X_2,X_3) \\ &= H(X_1\oplus(X_2\vee X_3)\oplus N_1|Q)-H(X_1\oplus(X_2\vee X_3)\oplus N_1|Q,X_1,X_2,X_3) \\ &= h_b(\tau_1(1-\beta)+(1-\tau_1)\beta)-h_b(\delta_1), \end{aligned}$$

where the last equality follows from substituting $p_{X_j|Q}: j=1,2,3$ derived in the earlier two claims.[24] The hypothesis (4.22) therefore precludes $(h_b(\tau_1*\delta_1)-h_b(\delta_1),h_b(\tau*\delta)-h_b(\delta),h_b(\tau*\delta)-h_b(\delta))\in\alpha_f^{3\text{-}1}(p_{QU_2U_3\underline{XY}})$ if $H(X_2|Q,U_2)=H(X_3|Q,U_3)=0$. This proves our third claim.

Our fourth claim is $H(X_2\vee X_3|Q,U_2,U_3)>0$. The proof of this claim rests on each of the earlier three claims. Note that we have either $H(X_2|Q,U_2)>0$ or $H(X_3|Q,U_3)>0$. Without loss of generality, we assume $H(X_2|Q,U_2)>0$. Note that

$$H(X_2|QU_2)=\sum_{q\in\mathcal{Q}}p_Q(q)\sum_{u_2\in\mathcal{U}_2}p_{U_2|Q}(u_2|q)H(X_2|U_2=u_2,Q=q)>0.$$

There exists $q^*\in\mathcal{Q}$ such that $p_Q(q^*)>0$ and $H(X_2|U_2,Q=q^*)=\sum_{u_2\in\mathcal{U}_2}p_{U_2|Q}(u_2|q^*)H(X_2|U_2=u_2,Q=q^*)>0$. We therefore have a $u_2^*\in\mathcal{U}_2$ such that $p_{U_2|Q}(u_2^*|q^*)>0$ and $H(X_2|U_2=u_2^*,Q=q^*)>0$. This implies $p_{X_2|U_2Q}(x_2|u_2^*,q^*)\notin\{0,1\}$ for each $x_2\in\{0,1\}$.

Since $p_Q(q^*)>0$, from the first claim we have

$$0<1-\tau=p_{X_3|Q}(0|q^*)=\sum_{u_3\in\mathcal{U}_3}p_{X_3U_3|Q}(0,u_3|q^*).$$

This guarantees existence of $u_3^*\in\mathcal{U}_3$ such that $p_{X_3U_3|Q}(0,u_3^*|q^*)>0$. We therefore have $p_{U_3|Q}(u_3^*|q^*)>0$ and $1\geq p_{X_3|U_3Q}(0|u_3^*,q^*)>0$.

We have therefore identified $(q^*,u_2^*,u_3^*)\in\mathcal{Q}\times\mathcal{U}_2\times\mathcal{U}_3$ such that $p_Q(q^*)>0$, $p_{U_2|Q}(u_2^*|q^*)>0$, $p_{U_3|Q}(u_3^*|q^*)>0$, $p_{X_2|U_2Q}(x_2|u_2^*,q^*)\notin\{0,1\}$ for each $x_2\in\{0,1\}$ and $1\geq p_{X_3|U_3Q}(0|u_3^*,q^*)>0$. By conditional independence of the pairs (X_2,U_2) and (X_3,U_3) given Q, we also have $p_{X_2|U_2U_3Q}(x_2|u_2^*,u_3^*,q^*)\notin\{0,1\}$ for each $x_2\in\{0,1\}$ and $1\geq p_{X_3|U_2U_3Q}(0|u_2^*,u_3^*,q^*)>0$. The reader may now verify $p_{X_2\vee X_3|U_2U_3Q}(x|u_2^*,u_3^*,q^*)\notin\{0,1\}$ for each $x\in\{0,1\}$. Since $p_{QU_2U_3}(q^*,u_2^*,u_3^*)=p_Q(q^*)p_{U_2|Q}(u_2^*|q^*)p_{U_3|Q}(u_3^*|q^*)>0$, we have proved the fourth claim.

Our fifth and final claim is $R_1<h_b(\tau_1*\delta_1)-h_b(\delta_1)$. This follows from a sequence of steps employed in proof of the second claim or in the proof of theorem. Denoting $\tilde{Q}:=(Q,U_2,U_3)$ and a generic element $\tilde{q}:=(q,u_2,u_3)\in\tilde{\mathcal{Q}}$:

[24] $\beta:=(1-\tau)^2\delta_1+(2\tau-\tau^2)(1-\delta_1)$ is as defined in the statement of the lemma.

$= \mathcal{Q} \times \mathcal{U}_2 \times \mathcal{U}_3$, we observe that

$$\begin{aligned}
R_1 \leq I(X_1;Y_1|\tilde{Q}) &= \sum_{\tilde{q}} p_{\tilde{Q}}(\tilde{q}) H(Y_1|\tilde{Q}=\tilde{q}) - \sum_{x_1,\tilde{q}} p_{\tilde{Q}X_1}(\tilde{q},x_1) H(Y_1|X_1=x_1,\tilde{Q}=\tilde{q}) \\
&= \sum_{\tilde{q}} p_{\tilde{Q}}(\tilde{q}) H(X_1 \oplus N_1 \oplus (X_2 \vee X_3)|\tilde{Q}=\tilde{q}) - \sum_{x_1,\tilde{q}} p_{X_1\tilde{Q}}(x_1,\tilde{q}) H(x_1 \oplus N_1 \oplus (X_2 \vee X_3)|X_1=x_1,\tilde{Q}=\tilde{q}) \\
&= \sum_{\tilde{q}} p_{\tilde{Q}}(\tilde{q}) H(X_1 \oplus N_1 \oplus (X_2 \vee X_3)|\tilde{Q}=\tilde{q}) - \sum_{x_1,\tilde{q}} p_{X_1\tilde{Q}}(x_1,\tilde{q}) H(N_1 \oplus (X_2 \vee X_3)|\tilde{Q}=\tilde{q}) \qquad (4.27) \\
&= \sum_{\tilde{q}} p_{\tilde{Q}}(\tilde{q}) H(X_1 \oplus N_1 \oplus (X_2 \vee X_3)|\tilde{Q}=\tilde{q}) - \sum_{\tilde{q}} p_{\tilde{Q}}(\tilde{q}) H(N_1 \oplus (X_2 \vee X_3)|\tilde{Q}=\tilde{q}) \\
&< \sum_{\tilde{q}} p_{\tilde{Q}}(\tilde{q}) H(X_1 \oplus N_1|\tilde{Q}=\tilde{q}) - \sum_{\tilde{q}} p_{\tilde{Q}}(\tilde{q}) H(N_1|\tilde{Q}=\tilde{q}) = \sum_{q} p_{\tilde{Q}}(\tilde{q}) H(X_1 \oplus N_1|\tilde{Q}=\tilde{q}) - h_b(\delta_1) \qquad (4.28) \\
&= \sum_{\tilde{q}} p_{\tilde{Q}}(\tilde{q}) h_b(\tau_{1\tilde{q}} * \delta_1) - h_b(\delta_1) \leq h_b(\mathbb{E}_{\tilde{Q}}\{\tau_{1\tilde{q}} * \delta_1\}) - h_b(\delta_1) = h_b(p_{X_1}(1) * \delta_1) - h_b(\delta_1), \qquad (4.29)
\end{aligned}$$

where (i) (4.27) follows from independence of (N_1, X_2, X_3) and X_1 conditioned on realization of $\tilde{Q}$, (ii) (4.28) follows from existence of a $\tilde{q}^* \in \tilde{\mathcal{Q}}$ for which $H(X_2 \vee X_3|\tilde{Q}=\tilde{q}^*) > 0$ and substituting $p_{X_1 \oplus N_1|\tilde{Q}}(\cdot|\tilde{q}^*)$ for p_{Z_1}, $p_{N_1|\tilde{Q}}(\cdot|\tilde{q}^*)$ for p_{Z_2} and $p_{X_2 \vee X_3|\tilde{Q}}(\cdot|\tilde{q}^*)$ for p_{Z_3} in lemma 4.5.4, (iii) the first inequality in (4.29) follows from Jensen's inequality. Since $p_{X_1}(1) * \delta_1 = p_{X_1}(1-\delta_1) + (1-p_{X_1}(1))\delta_1 = p_{X_1}(1)(1-2\delta_1) + \delta_1 \leq \tau_1(1-2\delta_1) + \delta_1 \leq \frac{1}{2}(1-2\delta_1) + \delta_1 = \frac{1}{2}$. Therefore $h_b(p_{X_1}(1) * \delta_1)$ is increasing[25] in $p_{X_1}(1)$ and is bounded above by $h_b(\tau_1 * \delta_1)$. We therefore have $R_1 < h_b(\tau_1 * \delta_1) - h_b(\delta_1)$. ■

We now derive conditions under which $\alpha_f^{3-1}(\tau_1,\tau,\tau) = \mathbb{C}(\tau_1,\tau,\tau)$. Clearly, $\mathbb{C}(\tau_1,\tau,\tau) \subseteq \beta(\underline{\tau},\underline{\delta})$ where $\underline{\tau} = (\tau_1,\tau,\tau)$ and $\underline{\delta} = (\delta_1,\delta,\delta)$ and $\beta(\underline{\tau},\underline{\delta})$ is as given in (4.3). It therefore suffices to derive conditions under which $(h_b(\tau_1 * \delta_1) - h_b(\delta_1), h_b(\tau * \delta) - h_b(\delta), h_b(\tau * \delta) - h_b(\delta)) \in \alpha_f^{3-1}(\tau_1,\tau,\tau)$.

Lemma 4.6.9 *Consider example 4.6.7 with $\delta := \delta_2 = \delta_3 \in (0, \frac{1}{2})$ and $\tau := \tau_2 = \tau_3 \in (0, \frac{1}{2})$. Let $\beta := (1-\tau)^2\delta_1 + (2\tau - \tau^2)(1-\delta_1)$. The rate triple $(h_b(\tau_1 * \delta_1) - h_b(\delta_1), h_b(\tau * \delta) - h_b(\delta), h_b(\tau * \delta) - h_b(\delta)) \in \alpha_f^{3-1}(\tau_1,\tau,\tau)$ i.e., achievable using coset codes, if,*

$$h_b(\tau * \delta) - h_b(\delta) \leq \theta, \qquad (4.30)$$

where $\theta = h_b(\tau) - h_b((1-\tau)^2) - (2\tau - \tau^2)h_b(\frac{\tau^2}{2\tau-\tau^2}) - h_b(\tau_1 * \delta_1) + h_b(\tau_1(1-\beta) + (1-\tau_1)\beta)$. We therefore have $\alpha_f^{3-1}(\tau_1,\tau,\tau) = \mathbb{C}(\tau_1,\tau,\tau)$ if (4.30) holds. □

Proof: The proof only involves identifying the appropriate test channel $p_{QU_2U_3\underline{XY}} \in \mathbb{D}_f^{3-1}(\tau_1,\tau,\tau)$. Let $\mathcal{Q} = \phi$ be empty, $\mathcal{U}_2 = \mathcal{U}_3 = \{0,1,2\}$. Let $p_{X_1}(1) = 1 - p_{X_1}(0) = \tau_1$. Let $p_{U_jX_j}(0,0) = 1 - p_{U_jX_j}(1,1) = 1 - \tau$ and therefore $P(U_j = 2) = P(X_j \neq U_j) = 0$ for $j = 2,3$. It is easily verified that $p_{QU_2U_3\underline{XY}} \in \mathbb{D}_f^{3-1}(\tau_1,\tau,\tau)$, i.e, in particular respects the cost constraints.

[25] This also employs the positivity of $1-2\delta_1$, or equivalently δ_1 being in the range $(0, \frac{1}{2})$.

The choice of this test channel, particularly the ternary fields, is motivated by $H(X_2 \vee X_3 | U_2 \oplus_3 U_3) = 0$. The decoder 1 can reconstruct the interfering pattern after having decoded the ternary sum of the codewords. It maybe verified that for this test channel $p_{QU_2U_3\underline{XY}}$, $\alpha_f^{3\text{-}1}(\tau_1, \tau, \tau)$ is defined as the set of rate triples $(R_1, R_2, R_3) \in [0, \infty)^3$ that satisfy

$$\begin{aligned} R_1 < \min\{0, \theta\} + h_b(\tau_1 * \delta_1) - h_b(\delta_1), \quad R_j < h_b(\tau * \delta) - h_b(\delta) : j = 2, 3 \\ R_1 + R_j < h_b(\tau_1 * \delta_1) - h_b(\delta_1) + \theta, \end{aligned} \tag{4.31}$$

where $\theta = h_b(\tau) - h_b((1-\tau)^2) - (2\tau - \tau^2)h_b(\frac{\tau^2}{2\tau - \tau^2}) - h_b(\tau_1 * \delta_1) + h_b(\tau_1(1-\beta) + (1-\tau_1)\beta)$ is as defined in the statement of the lemma. Clearly, $(h_b(\tau_1 * \delta_1) - h_b(\delta_1), h_b(\tau * \delta) - h_b(\delta), h_b(\tau * \delta) - h_b(\delta)) \in \alpha_f^{3\text{-}1}(p_{U_2U_3\underline{XY}})$ if (4.30) is satisfied. ■

Conditions (4.22) and (4.30) are *not* mutually exclusive. It maybe verified that the choice $\tau_1 = \frac{1}{90}$, $\tau = 0.15$, $\delta_1 = 0.01$ and $\delta = 0.067$ satisfies both conditions thereby establishing the utility of structured codes for examples well beyond particular additive ones.

4.6.2 Step II: The PCC rate region for a general discrete 3−IC

In this section, we employ PCC to manage interference seen by each receiver. In the sequel, we describe the coding technique and provide a characterization of the corresponding achievable rate region. In the interest of brevity, we omit the proof of achievability. All the non-trivial and new elements of such a proof have been detailed in the proof of theorem 4.6.2.

User j splits it's message M_j of rate $R_j = L_j + T_{ji} + T_{jk}$ into three parts $(M_{ji}^U, M_{jk}^U, M_j^V)$, where i, j, k are distinct indices in $\{1, 2, 3\}$. Let $\mathcal{U}_{ji} = \mathcal{F}_{\pi_i}, \mathcal{U}_{jk} = \mathcal{F}_{\pi_k}$ be finite fields and $\mathcal{V}_j$ be an arbitrary finite set. Let $\lambda_{ji} \subseteq \mathcal{U}_{ji}^n$ denote an $(n, s_{ji} + t_{ji}, t_{ji})$ PCC and $\lambda_{jk} \subseteq \mathcal{U}_{jk}^n$ denote an $(n, s_{jk} + t_{jk}, t_{jk})$ PCC. If we let $S_{ji} := \frac{s_{ji}}{n} \log \pi_i, T_{ji} := \frac{t_{ji}}{n} \log \pi_i$ and $S_{jk} := \frac{s_{jk}}{n} \log \pi_k, T_{jk} := \frac{t_{jk}}{n} \log \pi_k$ then recall that recall that $\lambda_{ji}, \lambda_{jk}$ are coset codes of rates $S_{ji} + T_{ji}, S_{jk} + T_{jk}$ partitioned into $\exp\{nT_{ji}\}, \exp\{nT_{jk}\}$ bins respectively.[26] Observe that cosets λ_{ji} and λ_{ki} are built over the same finite field $\mathcal{F}_{\pi_i}$. To enable contain the range the sum of these cosets, the larger of λ_{ji}, λ_{ki} contains the other. A codebook $\mathcal{C}_j$ of rate $K_j + L_j$ is built over $\mathcal{V}_j$. Codewords of $\mathcal{C}_j$ are partitioned into $\exp\{nL_j\}$ bins.

M_{ji}^U,M_{jk}^U and M_j^V index bins in λ_{ji}, λ_{jk} and $\mathcal{C}_j$ respectively. Encoder looks for a triplet of codewords from the indexed bins that are jointly typical with respect to a pmf $p_{U_{ji}U_{jk}V_j}$ defined on $\mathcal{U}_{ji} \times \mathcal{U}_{jk} \times \mathcal{V}_j$. Having chosen one such jointly typical collection, say $(U_{ji}^n, U_{jk}^n, V_j^n)$, the encoder generates a vector X_j^n according to $\prod_{t=1}^{n} p_{X_j|U_{ji}U_{jk}V_j}(\cdot|U_{jit}, U_{jkt}, V_{jt})$ and inputs the same on the channel.

[26] The reader will note a change in our notation. In section 4.6.1, we let $S_j \log \pi$ denote rate of user j's cloud center codebook. This was partitioned into $\exp\{nT_j \log \pi\}$ bins. In this section, we let the cloud center coset codes to be of rate $S_{ji} + T_{ji}$ and $S_{jk} + T_{jk}$ partitioned into $\exp\{nT_{ji}\}$ and $\exp\{nT_{jk}\}$ bins respectively.

Decoder j receives Y_j^n and looks for all triples $(u_{ji}^n, u_{jk}^n, v_j^n)$ of codewords in $\lambda_{ji} \times \lambda_{jk} \times \mathcal{C}_j$ for which there exists a $u_{ij\oplus kj}^n \in (\lambda_{ij} \oplus \lambda_{kj})$ such that $(u_{ij\oplus kj}^n, u_{ji}^n, u_{jk}^n, v_j^n, Y_j^n)$ are jointly typical with respect to $p_{U_{ij}\oplus U_{kj}, U_{ji}, U_{jk}, V_j, Y_j}$. If it finds all such triples in a unique triple of bins, the corresponding triple of bin indices is declared as decoded message of user j. Otherwise, an error is declared.

In order to characterize an achievable rate region, we average the performance of the above coding technique via random coding. The distribution induced on the ensemble of codebooks is a simple generalization of that employed in proof of theorem 4.6.2. In particular, the codewords of $\mathcal{C}_j$ are chosen independently according to $\prod_{t=1}^{n} p_{V_j|Q}(\cdot|q^t)$, where q^n is an appropriately chosen time sharing sequence. The three pairs $(\Lambda_{12}, \Lambda_{32}), (\Lambda_{21}, \Lambda_{31}), (\Lambda_{13}, \Lambda_{23})$ of random PCC are mutually independent. Within each such pair, (i) the generator matrix of the smaller PCC is obtained by choosing each of it's rows uniformly and independently, and (ii) the generator matrix of the larger is obtained by appending the generator matrix of the smaller with an appropriately chosen number mutually independent and uniformly distributed rows. All the vectors specifying the coset shifts are chosen independently and uniformly. Moreover, partitioning of all codes into their bins is effected uniformly and independently.[27] Deriving an upper bound on the average probability of error of this random collection of codebooks coupled with the above coding technique yields the following rate region.

Definition 4.6.10 *Let $\mathbb{D}_f(W_{\underline{Y}|\underline{X}}, \kappa, \tau)$ denote the collection of probability mass functions $(p_{Q\underline{U}\underline{V}\underline{X}\underline{Y}})$ defined on $\mathcal{Q} \times \underline{\mathcal{U}} \times \underline{\mathcal{V}} \times \underline{\mathcal{X}} \times \underline{\mathcal{Y}}$, where*

(i) $\mathcal{Q}, \mathcal{V}_1, \mathcal{V}_2, \mathcal{V}_3$ are arbitrary finite sets, $\underline{\mathcal{V}} := \mathcal{V}_1 \times \mathcal{V}_2 \times \mathcal{V}_3$,

(ii) $\mathcal{U}_{ij} = \mathcal{F}_{\pi_j}$[28] for each $1 \leq i, j \leq 3$, and $\underline{\mathcal{U}} := \mathcal{U}_{12} \times \mathcal{U}_{13} \times \mathcal{U}_{21} \times \mathcal{U}_{23} \times \mathcal{U}_{31} \times \mathcal{U}_{32}$,

(iii) $\underline{V} := (V_1, V_2, V_3)$ and $\underline{U} := (U_{12}, U_{13}, U_{21}, U_{23}, U_{31}, U_{32})$,

such that (i) the three quadruples $(U_{12}, U_{13}, V_1, X_1)$, $(U_{23}, U_{21}, V_2, X_2)$ and $(U_{31}, U_{32}, V_3, X_3)$ are conditionally mutually independent given Q, (ii) $p_{\underline{Y}|\underline{X}\underline{V}\underline{U}} = p_{\underline{Y}|\underline{X}} = W_{\underline{Y}|\underline{X}}$, (iii) $\mathbb{E}\{\kappa_j(X_j)\} \leq \tau_j$ for $j = 1, 2, 3$.

For $p_{\underline{U}\underline{V}\underline{X}\underline{Y}} \in \mathbb{D}_f(W_{\underline{Y}|\underline{X}}, \kappa, \tau)$, let $\beta_f(p_{\underline{U}\underline{V}\underline{X}\underline{Y}})$ be defined as the set of rate triples $(R_1, R_2, R_3) \in [0, \infty]^3$ for which there exists nonnegative numbers $S_{ij} : ij \in \{12, 13, 21, 23, 31, 32\}, T_{jk} : jk \in \{12, 13, 21, 23, 31, 32\}, K_j : j \in \{1, 2, 3\}, L_j : j \in \{1, 2, 3\}$ that satisfy $R_1 = T_{12} + T_{13} + L_1, R_2 = T_{21} + T_{23} + L_2, R_3 = T_{31} + T_{32} + L_3$ and

$$S_{A_j} + K_j > \sum_{a_j \in A_j} \log |\mathcal{U}_{a_j}| + H(V_j|Q) - H(U_{A_j}, V_j|Q), \tag{4.32}$$

$$S_{A_j} > \sum_{a_j \in A_j} \log |\mathcal{U}_{a_j}| - H(U_{A_j}|Q), \tag{4.33}$$

[27] The reader is encouraged to confirm that the distribution induced herein is a simple generalization of that employed in proof of theorem 4.6.2.

[28] *Recall $\mathcal{F}_{\pi_j}$ is the finite field of cardinality π_j.*

$$
\begin{aligned}
S_{A_j} + T_{A_j} &< \sum_{a\in A_j} \log|\mathcal{U}_a| - H(U_{A_j}|Q, U_{A_j^c}, U_{ij}\oplus U_{kj}, V_j, Y_j)\\
S_{A_j} + T_{A_j} + S_{ij} + T_{ij} &< \sum_{a\in A_j} \log|\mathcal{U}_a| + \log\pi_j - H(U_{A_j}, U_{ij}\oplus U_{kj}|Q, U_{A_j^c}, V_j, Y_j)\\
S_{A_j} + T_{A_j} + S_{kj} + T_{kj} &< \sum_{a\in A_j} \log|\mathcal{U}_a| + \log\pi_j - H(U_{A_j}, U_{ij}\oplus U_{kj}|Q, U_{A_j^c}, V_j, Y_j)\\
S_{A_j} + T_{A_j} + K_j + L_j &< \sum_{a\in A_j} \log|\mathcal{U}_a| + H(V_j) - H(U_{A_j}, V_j|Q, U_{A_j^c}, U_{ij}\oplus U_{kj}, Y_j)\\
S_{A_j} + T_{A_j} + K_j + L_j + S_{ij} + T_{ij} &< \sum_{a\in A_j} \log|\mathcal{U}_a| + \log\pi_j + H(V_j) - H(U_{A_j}, V_j, U_{ij}\oplus U_{kj}|Q, U_{A_j^c}, Y_j)\\
S_{A_j} + T_{A_j} + K_j + L_j + S_{kj} + T_{kj} &< \sum_{a\in A_j} \log|\mathcal{U}_a| + \log\pi_j + H(V_j) - H(U_{A_j}, V_j, U_{ij}\oplus U_{kj}|Q, U_{A_j^c}, Y_j),
\end{aligned}
\tag{4.34}
$$

for every $A_j \subseteq \{ji, jk\}$ *with distinct indices* i, j, k *in* $\{1, 2, 3\}$*, where* $S_{A_j} := \sum_{a_j\in A_j} S_{a_j}, U_{A_j} = (U_{a_j} : a_j \in A_j)$*. Let*

$$
\beta_f(W_{\underline{Y}|\underline{X}}, \kappa, \tau) = \mathit{cocl}\left(\bigcup_{\substack{p_{\underline{U}\underline{V}\underline{X}\underline{Y}}\in\\ \mathbb{D}_f(W_{\underline{Y}|\underline{X}},\kappa,\tau)}} \beta_f(p_{\underline{U}\underline{V}\underline{X}\underline{Y}})\right).
$$

Theorem 4.6.11 *For 3-IC* $(\underline{\mathcal{X}}, \underline{\mathcal{Y}}, W_{\underline{Y}|\underline{X}}, \kappa)$*,* $\beta_f(W_{\underline{Y}|\underline{X}}, \kappa, \tau)$ *is achievable, i.e.,* $\beta_f(W_{\underline{Y}|\underline{X}}, \kappa, \tau) \subseteq \mathbb{C}(W_{\underline{Y}|\underline{X}}, \kappa, \tau)$*.* □

Since all the non-trivial elements of this proof are captured in the proof of theorem 4.6.2, and is only more involved in notation, we omit the same.

4.6.3 Enlarging the PCC rate region using unstructured codes

Let us describe a coding technique that incorporates both unstructured and partitioned coset codes. We follow the approach of Ahlswede and Han [48]. Refer to figure 4.4 for an illustration of the random variables involved. Each user splits it's message into 5 parts. The $W-$random variable is decoded by all users. In addition, each user decodes a univariate component of the other user messages. This is represented by the random variable T. Furthermore, it decodes a bivariate interference component denoted using U. Lastly, each decoder decodes all it's parts. As indicated by Han and Kobayashi [13], this achievable rate region could potentially be enlarged through the use of a time sharing random variable. A description of the corresponding achievable rate region being sufficiently involved, we omit the details. The reader may refer to section 7.3 unstructured and coset codes are glued together to derive an achievable rate region for the computation over MAC problem.

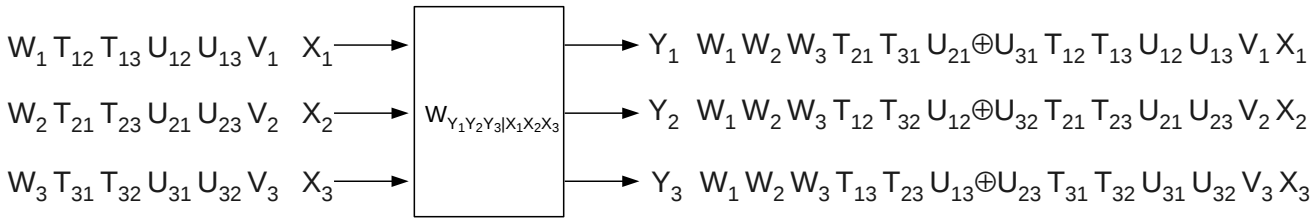

Figure 4.4: Collection of random variables associated with coding technique that incorporates unstructured and partitioned coset codes

Chapter 5

Three user broadcast channel

We begin with a brief description of a three user discrete broadcast channel (3−DBC) and the problem statement. A 3−DBC, as depicted in figure 5.1, consists of a single transmitter (Tx) and three receivers (Rx). The transmitter wishes to communicate specific information streams, that are assumed statistically independent, to each of the three receivers. The transmitter is provided with a finite input alphabet set $\mathcal{X}$ and receiver j observes symbols in a finite output alphabet set $\mathcal{Y}_j$. Let $W_{Y_1Y_2Y_3|X}$ denote the channel transition probabilities. As always, we assume the channel is memoryless, time invariant and used without feedback. The problem of interest is to characterize it's capacity region. Please refer to section 5.4 for a precise statement of this problem. In the following, we provide a discussion of current coding techniques and our findings.

The problem of designing efficient coding techniques for communicating over a broadcast channel (BC) was initiated [20] in the context of a BC with two receivers (2−BC). Over a 2−BC, the transmitter maps two information bearing signals, meant for the two receivers, into one signal that can be input on the channel. The channel produces an output signal at each receiver based on the input signal. From the perspective of each receiver, it's signal undergoes a transformation, in accordance with the other receiver's signal. This transformation is akin to the effect

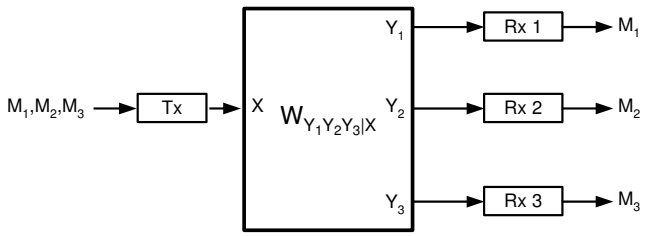

Figure 5.1: Three user broadcast channel (3−BC)

of interference, and is therefore undesirable[1]. Since this transformation is inevitable, one technique to minimize it's effect is to make available to each receiver the information bearing signal of the other receiver, henceforth referred to as interfering signal. Each receiver can therefore reconstruct, as best as possible, this transformation, or in other words, peel the effect of the interfering signal, and thereby enhance it's ability to decode the desired signal.

If one employs this technique, observe that each receiver's rate is constrained, not just by it's ability, but also by the other receiver's ability to decode it's signal. This technique is therefore prohibitive, unless one of the receivers is stronger, i.e., capable of decoding everything that the other receiver can.[2] In this case, one can fix a rate of communication to the weaker receiver that it can support. Since the stronger receiver can decode it's interfering signal at this rate, and thereby peel it's effect off, it can decode it's signal at a rate that is limited only by the channel. Not surprisingly, this technique, which has come to be known as superposition coding, came to light [20] [21] in the context of a degraded 2−BC, which precisely models the above scenario.

For communicating over a general 2−BC, it is natural to consider a generalization of the superposition technique to enable each receiver decode a *part* of the interfering signal. By choosing the parts carefully, one might be able to better trade off the benefit of decoding the interfering signal and the constraint it imposes. Hajek and Pursley [49] proved that this was indeed the case. This led to the technique of splitting the transmission into three parts. The signal meant for each receiver is split into two parts - public and private. The two public parts are combined together to form the base layer signal.[3] Each private part is identified with a signal in the satellite layer. The three signals are mapped into a signal input on the channel. Each receiver decodes the base layer signal and the signal corresponding to it's private part.

By decoding the base layer signal, each receiver decodes the public part of the interfering signal. This enables each receiver peel off the effect of the public part of it's interfering signal. How does one accommodate the presence of private parts? Following Gel'fand's ingenious coding technique [8], devised for a particular two user discrete BC (2−DBC), Marton proposed the technique of precoding via binning [9]. Instead of choosing the signals for the private parts independently, precoding via binning enables the encoder jointly choose a compatible pair of signals. By jointly choosing the pair of signals, the effect of each private part on the other receiver is minimized. In other words, the transformation effected by the private part of the interfering signal is lent more benign by jointly choosing the pair of signals. Precoding via binning turns out to be an efficient technique for multiplexing information bearing signals meant for different receivers.

Splitting the transmission of each user into public and private parts, precoding the private parts via binning and superposing the latter over a base layer comprising of public parts are the current known coding techniques for

[1]This transformation is the combined effect of (i) the map employed by the transmitter, and (ii) the channel. While the latter is inevitable, the former is a necessary evil.

[2]This technique is also not prohibitive if each receiver can decode what the other can.

[3]The informed reader may associate this with the codeword chosen from the cloud center codebook.

communicating over a BC with any number of receivers.[4] We will henceforth refer to this combination of the three coding techniques in the context of a 2−BC as Marton's coding technique [9]. Before we move onto a 3−DBC, let us make note of one important finding that determines allocation between public and private parts. In general, the knowledge of any part of the interfering signal can be better exploited at the receiver than via precoding. In other words, precoding, in general, results in a *rate loss*, i.e., precoding for any component results in a lower rate when compared to a scenario wherein the decoder is made available that component. It is therefore desirable to decode as large a part of the interfering signal as possible, without constraining the rate of the other receiver, and precode for the rest.[5]

Over a 2−BC, each receiver is plagued by the presence of a *single* interfering signal. The technique of superposition coding using unstructured codes provides an efficient technique for decoding a suitably large part of the single interfering signal. In this chapter, we study the problem of communicating over a 3−DBC. Over a three user BC (3−BC), each receiver is plagued by two interfering signals. What are the current known coding techniques for communicating over a 3−DBC and how they deal with two interfering signals?

The current known coding techniques for communicating over a 3−DBC are based on Marton's coding technique. We henceforth refer to this as $\mathscr{UM}-$technique. Each information bearing signal is split into four parts - one public, two semi-private and one private part. The public part of every receiver is decoded by all receivers. In addition, each of it's semi-private part is decoded by exactly one other receiver.[6] The technique of superposition coding and precoding via binning are appropriately combined to multiplex the twelve parts of the three signals.[7] Without going into the details of this technique, we highlight one element that will play a key role herein. The $\mathscr{UM}-$technique enables each receiver decode individual parts[8] of the interfering signals. The contributions of this chapter are based on the following three questions. Firstly, does it suffice to decode individual parts of the interfering signal? If not, what parts of the two interfering signals must a receiver decode? How does one enable a receiver decode these parts *efficiently*?

In this chapter, we prove that in addition to individual parts of the interfering signal, it benefits for receivers over a 3−DBC to decode *bivariate* parts of the same. Since the $\mathscr{UM}-$technique is based on unstructured codes, it is suited for decoding individual parts of the interfering signals. It is therefore constrained to decode the arguments of the bivariate function. If the bivariate function is sufficiently compressive, i.e., entropy of the function is significantly lower than the joint entropy of the arguments, then decoding the arguments is an inefficient technique. These ideas

[4]These coding techniques achieve capacity for several interesting classes [20, 50, 21, 51, 52, 53, 54, 55, 56, 9, 57, 58, 59, 60, 49].

[5]Precoding for the Gaussian channel, referred to as dirty paper coding (DPC) is a popular instance of no rate loss. Indeed, over a vector Gaussian BC, the transmitter can precode for all of the interfering signal, lending a trivial public part. This is closely linked to the optimality of dirty paper coding for the Gaussian MIMO BC [61]. As the reader will later note, this is the reason why lattices are superfluous for communicating over vector Gaussian broadcast channels with any number of receivers.

[6]Clearly, every receiver decodes all it's parts too.

[7]In section 5.5.2 we provide an exposition of this coding technique.

[8]If V_2 and V_3 denote the interfering signals for receiver 1, we refer to univariate functions of these signals, say $f_2(V_2)$ and $f_3(V_3)$, as individual parts of V_2 and V_3. In contrast, we refer to a bivariate function of the same, say $g(V_2, V_3)$ as a bivariate part or a bivariate interference component.

lead us to identify a vector additive 3−DBC for which one of the receivers, say receiver 1, benefits by decoding the sum - a bivariate part - of the interfering signals. If the pair of interfering signals have a high entropy, the $\mathscr{UM}-$technique cannot enable receiver 1 decode this pair. It is then forced to decode strictly smaller public parts of the same and precode for the non-trivial private parts. In contrast to the Gaussian setting, this precoding results in a rate loss, i.e., receiver 1 pays for it's inability to reconstruct the sum of the interfering signals.

For this vector additive 3−DBC, we propose a linear coding technique wherein the interfering signals are encoded using cosets of the same linear code. We exploit the algebraic closure property of these codes to enable receiver 1 *efficiently* reconstruct the sum of interfering signals *without* decoding the pair. This technique is therefore not constrained by the high entropy of the pair of interfering signals. It is only constrained by the entropy of the function. The function in this case, being a sum, is compressive and the linear coding technique can decode the sum while the $\mathscr{UM}-$technique is unable to decode the arguments. We therefore prove that the proposed linear coding technique strictly outperforms $\mathscr{UM}-$technique.[9]

5.1 Our contributions

Our findings in the context of a 3−IC (chapter 4) illustrated a similar phenomenon. Therein, we observed coset codes aid efficient communication even over non-additive scenarios, thus motivating the need to generalize the linear coding technique for communicating over a general 3−DBC. This leads us to develop an analogous *framework* based on partitioned coset codes (PCC) (definition 3.4.2) to communicate over an arbitrary 3−DBC. The following are the central elements of this framework. Firstly, the PCC are carefully chosen with mutual relationship that aids decoding the sum of chosen codewords. Secondly, in order to exploit algebraic closure property of PCC, we propose new decoding rules. Thirdly, we resort joint typicality encoding and decoding that enables us communicate over arbitrary 3−DBC and achieve rates corresponding to arbitrary single-letter distributions. This framework enables us derive a new achievable rate region for a general 3−DBC. Since it generalizes the linear coding technique for the vector 3−DBC, which in turn strictly outperforms $\mathscr{UM}-$technique, the derived achievable rate region is strictly larger than $\mathscr{UM}-$region for the vector additive 3−DBC.

The natural question to ask is whether the derived achievable rate region subsumes $\mathscr{UM}-$region for a general 3−DBC. As we have mentioned, (i) coset codes enable efficient decoding of bivariate parts of the two interfering signals and (ii) superposition coding using unstructured codes enable efficient decoding of individual parts of the interfering signal. Over a general 3−DBC, it maybe necessary to decode all of these parts. We therefore conclude

[9]The current known coding techniques being optimal for vector Gaussian BC [61], an observant reader might wonder why the same phenomenon cannot be exploited therein using lattice codes. The answer lies in the absence of a rate loss for the Gaussian case. Indeed, all of the interfering signal is precoded for, resulting in no part of the same needing to be decoded. This also emphasizes (i) what governs the choice of public and private parts and (ii) presence of rate loss in discrete channels leading to the phenomenon identified herein. Rate loss being a general phenomenon and the absence of the same being particular to the Gaussian setting, the theory developed herein is widely applicable.

that the proposed framework based on PCCs in conjunction with the current known coding techniques based on unstructured codes strictly enhances the latter. In other words, by incorporating the framework based on PCCs the current known largest achievable rate region can be strictly enlarged. We indicate a technique for the same in this chapter.

5.2 Significance of our contributions

The coding technique based on coset codes developed herein strictly outperforms the best known coding technique for communicating over a 3−DBC. In fact, even within the larger class of BC's with arbitrary input and output alphabets, any number of receivers and any number of antennae, we have been unaware, for over three decades since the findings of [9] came to light, of a BC for which the coding techniques of superposition coding and precoding via binning can be strictly improved upon. This chapter presents the first such example which elegantly ties together the two ideas of decoding bivariate components and rate loss in a novel setting.

Going beyond proposing a coding technique for a particular example, we develop a framework for communicating over an arbitrary 3−DBC. Bringing together techniques studied in disparate contexts - joint typicality coding techniques and codes endowed with algebraic closure properties - we derive a new achievable rate region for the general 3−DBC that includes rate regions corresponding to non-uniform distributions.[10] The derived rate region strictly enlarges upon the current known largest, that has remained so for over three decades now.

5.3 Content and organization

We begin with preliminaries in section 5.4. In section 5.5, we present the current known largest achievable rate regions for 2−DBC and 3−DBC. In particular, we describe the current known coding techniques for communicating over a BC, in the context of 2−DBC (section 5.5.1) and derive the corresponding achievable rate region. This is henceforth referred to as Marton's rate region in recognition of Marton, who derived the same in [9]. In section 5.5.2, appropriately stitch together all (relevant) current known coding techniques and derive an achievable rate region for 3−DBC. These are henceforth referred to as $\mathscr{UM}-$technique and $\mathscr{UM}-$region respectively. Section 5.6 contains our first main finding - identification of a vector additive 3−DBC for which the $\mathscr{UM}-$technique is proved to be strictly sub-optimal. The proof of strict sub-optimality is provided in section 5.10. Section 5.6 contains our second main finding - a framework based on partitioned coset codes (PCC) to communicate over an arbitrary 3−DBC.[11] To aid the reader, we present the same in three pedagogical steps. In section 5.8, we indicate how to glue

[10] It maybe noted that current works employing linear code based techniques such as [18], [15], [16], [17] restrict attention to additive scenarios and the coding techniques proposed therein do not generalize and moreover only achieve rates corresponding to uniform distributions.

[11] This is a generalization of the linear coding technique proposed for the vector additive 3−DBC studied in section 5.6.

together $\mathscr{UM}-$technique and the framework based on PCC and communicate efficiently over an arbitrary 3−DBC. An analysis of such a technique yields a new achievable rate region for an arbitrary 3−DBC that strictly enlarges upon $\mathscr{UM}-$region. A characterization of this region is quite involved and the reader is spared of the same.

5.4 Preliminaries: Notation, definitions and problem statement

We begin with remarks on notation in section 5.4.1. In section 5.4.2, we state relevant definitions - code, achievability and capacity - with respect to a 3−DBC and provide a precise statement of the problem of interest.

5.4.1 Notation

The empty sum has value 0, i.e, $\sum_{a\in\phi} = 0$. If p_{UV} is a distribution on $\mathcal{U}\times\mathcal{V}$, we let $p_{UV}^n := \prod_{t=1}^{n} p_{UV}$ unless otherwise specified. Similarly, if $q^n \in \mathcal{Q}^n$ and $p_{UV|Q}(\cdot,\cdot|q) : q \in \mathcal{Q}$ is a collection of conditional probabilities, we let $p_{UV|Q}^n(\cdot,\cdot|q^n) := \prod_{t=1}^{n} p_{UV|Q}(\cdot,\cdot|q_t)$. In this chapter, we will need to define pairs and triples of objects of the same type. In order to reduce clutter, we use an underline to denote aggregates of objects of similar type. For example, (i) if $\mathcal{Y}_1,\mathcal{Y}_2,\mathcal{Y}_3$ denote (finite) sets, we let $\underline{\mathcal{Y}}$ either denote the Cartesian product $\mathcal{Y}_1\times\mathcal{Y}_2\times\mathcal{Y}_3$ or abbreviate the collection $(\mathcal{Y}_1,\mathcal{Y}_2,\mathcal{Y}_3)$ of sets, the particular reference being clear from context, (ii) if $y_k \in \mathcal{Y}_k : k = 1,2,3$, we let $\underline{y} \in \underline{\mathcal{Y}}$ abbreviate $(y_1,y_2,y_3) \in \mathcal{Y}$, (iii) if $d_k : \mathcal{Y}_k^n \to \mathcal{M}_k : k = 1,2,3$ denote (decoding) maps, then we let $\underline{d}(\underline{y}^n)$ denote $(d_1(y_1^n), d_2(y_2^n), d_3(y_3^n))$, (iv) if U_2, U_3 are random variables taking values in $\mathcal{U}_2,\mathcal{U}_3$ respectively, we let $\underline{U} := U_2, U_3$ and similarly $\underline{u} := (u_2,u_3) \in \underline{\mathcal{U}} := \mathcal{U}_2\times\mathcal{U}_3$ denote a generic element.

5.4.2 Definitions: Broadcast channel, code, achievability and capacity

A 3−DBC consists of a finite input alphabet set $\mathcal{X}$ and three finite output alphabet sets $\mathcal{Y}_1,\mathcal{Y}_2,\mathcal{Y}_3$. The discrete time channel is (i) time invariant, i.e., the pmf of $\underline{Y}_t = (Y_{1t}, Y_{2t}, Y_{3t})$, the output at time t, conditioned on X_t, the input at time t, is invariant with t, (ii) memoryless, i.e., conditioned on present input X_t, the present output $\underline{Y}_t$ is independent of past inputs $X_1,\cdots,X_{t-1}$, past outputs $\underline{Y}_1,\cdots,\underline{Y}_{t-1}$ and (iii) used without feedback, i.e., the encoder has no information of the symbols received by the decoder. Let $W_{\underline{Y}|X}(\underline{y}|x) = W_{Y_1Y_2Y_3|X}(y_1,y_2,y_3|x)$ denote probability of observing $\underline{y} \in \underline{\mathcal{Y}}$ at the respective outputs conditioned on $x \in \mathcal{X}$ being input. Input is constrained with respect to a cost function $\kappa : \mathcal{X} \to [0,\infty)$. The cost function is assumed additive, i.e., cost of transmitting the vector $x^n \in \mathcal{X}^n$ is $\sum_{t=1}^n \kappa(x_t)$. For each $n \in \mathbb{N}$, let $\bar{\kappa}^n(x^n) := \frac{1}{n}\sum_{t=1}^n \kappa(x_t)$ denote the average cost of transmitting x^n, per symbol. We refer to this 3−DBC as $(\mathcal{X}, \underline{\mathcal{Y}}, W_{\underline{Y}|X}, \kappa)$.

In general, a 3−DBC can be employed to communicate seven messages - one to each non-empty subset of receivers (users). Throughout this chapter, we assume that none of the messages are to be shared among two or more receivers. In other words, the transmitter has one distinct message to be communicated to each receiver. The focus of this

chapter therefore is the (private message) capacity region of a 3−DBC, and in particular corresponding achievable rate regions. The following definitions make the relevant notions precise.

Definition 5.4.1 *A 3−DBC code $(n, \underline{\mathcal{M}}, e, \underline{d})$ consist of (i) finite index sets $\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3$ of messages, (ii) encoder map $e : \underline{\mathcal{M}} \to \mathcal{X}^n$, and (iii) three decoder maps $d_k : \mathcal{Y}_k^n \to \mathcal{M}_k : k = 1, 2, 3$.*

Definition 5.4.2 *The error probability of a 3−DBC code $(n, \underline{\mathcal{M}}, e, \underline{d})$ conditioned on message triple $(m_1, m_2, m_3) \in \underline{\mathcal{M}}$ is*

$$\xi(e, \underline{d}|\underline{m}) := 1 - \sum_{\underline{y}^n : \underline{d}(\underline{y}^n) = \underline{m}} W_{\underline{Y}|X}(\underline{y}^n | e(\underline{m})).$$

The average error probability of a 3−DBC code $(n, \underline{\mathcal{M}}, e, \underline{d})$ is $\bar{\xi}(e, \underline{d}) := \sum_{\underline{m} \in \underline{\mathcal{M}}} \frac{1}{|\mathcal{M}_1||\mathcal{M}_2||\mathcal{M}_3|} \xi(e, \underline{d}|\underline{m})$. Cost of transmitting message $\underline{m} \in \underline{\mathcal{M}}$ per symbol is $\tau(e|\underline{m}) := \bar{\kappa}^n(e(\underline{m}))$ and average cost per symbol of 3−DBC code $(n, \underline{\mathcal{M}}, e, \underline{d})$ is $\tau(e) := \frac{1}{|\mathcal{M}_1||\mathcal{M}_2||\mathcal{M}_3|} \sum_{\underline{m} \in \underline{\mathcal{M}}} \tau(e|\underline{m})$.

Definition 5.4.3 *A rate-cost quadruple $(R_1, R_2, R_3, \tau) \in [0, \infty)^4$ is achievable if for every $\eta > 0$, there exists $N(\eta) \in \mathbb{N}$ such that for all $n > N(\eta)$, there exists a 3−DBC code $(n, \underline{\mathcal{M}}^{(n)}, e^{(n)}, \underline{d}^{(n)})$ such that (i) $\frac{\log |\mathcal{M}_k^{(n)}|}{n} \geq R_k - \eta : k = 1, 2, 3$, (ii) $\bar{\xi}(e^{(n)}, \underline{d}^{(n)}) \leq \eta$, and (iii) average cost $\tau(e^{(n)}) \leq \tau + \eta$. The capacity region is $\mathbb{C}(W_{\underline{Y}|X}, \kappa, \tau) := cl\big\{\underline{R} \in \mathbb{R}^3 : (\underline{R}, \tau)$ is achievable$\big\}$.*

In this chapter, our objective is to characterize an inner bound to $\mathbb{C}(W_{\underline{Y}|X}, \kappa, \tau)$, i.e., an achievable rate region for a general 3−DBC. In the following section, we provide a characterization of the currently known largest achievable rate region for the same.

5.5 Current known largest achievable rate region a DBC

The currently known largest achievable rate region for 3−DBC is obtained via superposition and binning of unstructured codes. We henceforth refer to this $\mathscr{UM}$−technique and the corresponding achievable rate region as $\mathscr{UM}$−region. We begin with a brief review of Marton's rate region for the 2−DBC in section 5.5.1 and characterize $\mathscr{UM}$−region in section 5.5.2.

5.5.1 Marton's rate region

Marton's coding incorporates two fundamental coding techniques - superposition and precoding. Superposition involves each user decode a part of the signal carrying the other user's information and thereby enhance it's ability to decode the intended signals. The technique of jointly choosing each user's message bearing signal to contain mutual interference is precoding. Superposition coding is accomplished using a two layer coding scheme. First layer, which is public, contains a codebook over $\mathcal{W}$. Second layer is private and contains two codebooks one each on $\mathcal{V}_1$ and $\mathcal{V}_2$.

Precoding is accomplished by setting aside a *bin* of codewords for each private message, thus enabling the encoder to choose a compatible pair of codewords in the indexed bins. User jth message is split into two parts - public and private. The public parts together index a codeword in $\mathcal{W}-$codebook and the private part of user jth message index a codeword in $\mathcal{V}_j-$codebook. Both users decode from the public codebook and their respective private codebooks.

Definition 5.5.1 and theorem 5.5.2 provide a characterization of rate pairs achievable using Marton's coding technique. We omit restating the definitions analogous to definitions 5.4.1, 5.4.2, 5.4.3 for a 2-BC.

Definition 5.5.1 *Let $\mathbb{D}_M(W_{\underline{Y}|X}, \kappa, \tau)$ denote the collection of distributions $p_{QWV_1V_2XY_1Y_2}$ defined on $\mathcal{Q}\times\mathcal{W}\times\mathcal{V}_1\times\mathcal{V}_2\times\mathcal{X}\times\mathcal{Y}_1\times\mathcal{Y}_2$, where (i) $\mathcal{Q}$, $\mathcal{W}$, $\mathcal{V}_1$ and $\mathcal{V}_2$ are finite sets of cardinality at most $|\mathcal{X}|+4$, $|\mathcal{X}|+4$, $|\mathcal{X}|+1$ and $|\mathcal{X}|+1$ respectively, (ii) $p_{\underline{Y}|X\underline{Y}WQ} = p_{\underline{Y}|X} = W_{\underline{Y}|X}$, (iii) $\mathbb{E}\{\kappa(X)\} \leq \tau$. For $p_{QW\underline{V}X\underline{Y}} \in \mathbb{D}_M(W_{\underline{Y}|X}, \kappa, \tau)$, let $\alpha_M(p_{QW\underline{V}X\underline{Y}})$ denote the set of $(R_1, R_2) \in [0,\infty)^2$ that satisfy*

$$\begin{aligned} R_k &< I(WV_k; Y_k|Q) : k = 1, 2, \\ R_1 + R_2 &< \min\{I(W;Y_1|Q), I(W;Y_2|Q)\} + I(V_1;Y_1|QW) + I(V_2;Y_2|W,Q) - I(V_1;V_2|W,Q) \end{aligned}$$

and

$$\alpha_M(W_{\underline{Y}|X}, \kappa, \tau) = cocl\left(\bigcup_{\substack{p_{QW\underline{V}X\underline{Y}} \\ \in \mathbb{D}_M(W_{\underline{Y}|X}, \kappa, \tau)}} \alpha_M(p_{QW\underline{V}X\underline{Y}})\right)$$

Theorem 5.5.2 *For $2-DBC$ $(\mathcal{X}, \underline{\mathcal{Y}}, W_{\underline{Y}|X}, \kappa)$, $\alpha(W_{\underline{Y}|X}, \kappa, \tau)$ is achievable, i.e., $\alpha(W_{\underline{Y}|X}, \kappa, \tau) \subseteq \mathbb{C}(W_{\underline{Y}|X}, \kappa, \tau)$.* □

Remark 5.5.3 *The bounds on cardinality of $\mathcal{W}, \mathcal{V}_1$ and $\mathcal{V}_2$ were derived by Gohari and Anantharam in [62].*

We refer the reader to [9] for a proof of achievability. El Gamal and Meulen [59] provide a simplified proof using the method of second moment.

5.5.2 $\mathscr{UM}-$region : Current known largest achievable rate region for $3-$DBC

The $\mathscr{UM}-$technique is a 3 layer coding technique. For simplicity, we describe the coding technique without referring to the time sharing random variable and employ the same in characterizing $\mathscr{UM}-$region. User jth message M_j is split into four parts - two semi-private parts, and one, private and public parts each. We let message (i) $M_j^W \in \mathcal{M}_j^W$ of rate K_j denote it's public part (ii) $M_{i\not{j}}^U \in \mathcal{M}_{i\not{j}}^U, M_{\not{j}k}^U \in \mathcal{M}_{\not{j}k}^U$ of rates L_{ij}, K_{jk} respectively, denote it's semi-private parts, where (i,j,k) is an appropriate triple in $\{(1,2,3),(2,3,1),(3,1,2)\}$, and (iii) $M_j^V \in \mathcal{M}_j^V$ of rate T_j denote it's private part.

The first layer is public with a single codebook $(w^n(\underline{m}^W) : \underline{m}^W \in \underline{\mathcal{M}}^W)$ of rate $K_1+K_2+K_3$ over $\mathcal{W}$. $\underline{M}^W := (M_1^W, M_2^W, M_3^W)$ indexes a codeword in $\mathcal{W}-$codebook and each user decodes from $\mathcal{W}-$codebook.

Each codeword in $\mathcal{W}-$codebook is linked to a triple of codebooks - one each on $\mathcal{U}_{ij} : (i,j) \in \{(1,2),(2,3),(3,1)\}$- in the second layer. The second layer is semi-private. Each of the three semi-private codebooks is composed of *bins*, wherein each bin comprises a collection of codewords. For each pair $(i,j) \in \{(1,2),(2,3),(3,1)\}$ the following hold. $M^U_{i\not j}$ and $M^U_{\not i j}$ together index a bin in $\mathcal{U}_{ij}-$codebook. Each bin in $\mathcal{U}_{ij}-$codebook is of rate S_{ij}. Let $(u^n_{ij}(\underline{m}^W, m^U_{\not i j}, m^U_{i\not j}, s_{ij}) : s_{ij} \in [\exp\{nS_{ij}\}])$ denote bin corresponding to semi-private messages $\underline{m^U_{ij}} := (m^U_{\not i j}, m^U_{i\not j})$ in the $\mathcal{U}_{ij}-$codebook linked to public message $\underline{m}^W$. Users i,j decode from $\mathcal{U}_{ij}-$codebook and it maybe verified that $\mathcal{U}_{ij}-$codebook is of rate $K_{ij}+L_{ij}+S_{ij}$.

Let (i,j) and (j,k) be distinct pairs in $\{(1,2),(2,3),(3,1)\}$. Every pair of codewords in $\mathcal{U}_{ij}-$ and $\mathcal{U}_{jk}-$codebooks is linked to a codebook on $\mathcal{V}_j$. The codebooks over $\mathcal{V}_j : j = 1,2,3$ comprise the third layer which is private. M^V_j indexes a bin in $\mathcal{V}_j-$codebook, each of which is of rate S_j, and thus $\mathcal{V}_j-$codebook is of rate $T_j + S_j$. Let $(v^n_j(\underline{m}^W, \underline{m^U_{ij}}, s_{ij}, \underline{m^U_{jk}}, s_{jk}, m^V_j, s_j) : s_j \in [\exp\{nS_j\}])$ denote bin corresponding to private message m^V_j in the $\mathcal{V}_j-$codebook linked to codeword pair $(u^n_{ij}(\underline{m}^W, \underline{m^U_{ij}}, s_{ij}), u^n_{jk}(\underline{m}^W, \underline{m^U_{jk}}, s_{jk}))$. User j decodes from the private codebook over $\mathcal{V}_j$.

How does the encoder map messages to a codeword? Let $p_{W\underline{UV}X}$ be a distribution on $\mathcal{W}\times\underline{\mathcal{U}}\times\underline{\mathcal{V}}\times\mathcal{X}$ such that $\mathbb{E}\{\kappa(X)\} \le \tau$. The encoder looks for $(s_{12}, s_{23}, s_{31}, s_1, s_2, s_3)$ such that the septuple

$$\begin{pmatrix} w^n(\underline{M}^W), u^n_{ij}(\underline{M}^W, \underline{M^U_{ij}}, s_{ij}) : (i,j) = (1,2),(2,3),(3,1), \\ v^n_j(\underline{M}^W, \underline{M^U_{ij}}, s_{ij}, \underline{M^U_{jk}}, s_{jk}, M^V_j, s_j) : (i,j,k) = (1,2,3),(2,3,1),(3,1,2) \end{pmatrix}$$

of codewords is jointly typical with respect to $p_{W\underline{UV}}$. If such a septuple is found, this is mapped to a codeword on $\mathcal{X}^n$ which is input to the channel. If it does not find any such septuple, an error is declared.

Decoder j looks for all quadruples $(\hat{\underline{m}}^W, \underline{\hat{m}_{ij}}^U, \underline{\hat{m}_{jk}}^U, \hat{m}^V_j)$ such that

$$\left(w^n(\hat{\underline{m}}^W), u^n_{ij}(\hat{\underline{m}}^W, \underline{\hat{m}^U_{ij}}, s_{ij}), u^n_{jk}(\hat{\underline{m}}^W, \underline{\hat{m}^U_{jk}}, s_{jk}), v^n_j(\underline{m}^W, \underline{m^U_{ij}}, s_{ij}, \underline{m^U_{jk}}, s_{jk}, m^V_j, s_j), Y^n_j\right)$$

is jointly typical with respect to $p_{W\underline{UV}X\underline{Y}} := p_{W\underline{UV}X}W_{\underline{Y}|X}$, where (i) (i,j,k) is the appropriate triple in $\{(1,2,3),(2,3,1),(3,1,2)\}$ and (ii) Y^n_j is the received vector. If there is a unique such quadruple, it declares $\hat{m}_j := (\hat{m}^W_j, \hat{m}^U_{i\not j}, \hat{m}^U_{\not j k}, \hat{m}^V_j)$ as user jth message. Otherwise, i.e., none or more than one such quadruple is found, it declares an error.

As is typical in information theory, we average error probability over the entire ensemble of codebooks and upper bound the same. Moreover, we incorporate the time sharing random variable in the above coding technique using the standard approach. Let Q, taking values over the finite alphabet $\mathcal{Q}$, denote the time sharing random variable. Let p_Q be a pmf on $\mathcal{Q}$ and $q^n \in \mathcal{Q}^n$ denote a sequence picked according to p^n_Q. q^n is revealed to the encoder and all decoders. The codewords in $\mathcal{W}-$codebook are identically and independently distributed according to $p^n_{W|Q}(\cdot|q^n)$. Conditioned on entire public codebook $(W^n(\underline{m}^W) = w^n(\underline{m}^W) : \underline{m}^W \in \underline{\mathcal{M}}^W)$ and the time sharing sequence q^n, each of the codewords $U^n_{ij}(\underline{m}^W, \underline{m^U_{ij}}, s_{ij}) : (\underline{m^U_{ij}}, s_{ij}) \in \underline{\mathcal{M}_{ij}}^U \times [\exp\{nS_{ij}\}]$ are independent and identically distributed according

to $p^n_{U_{ij}|WQ}(\cdot|w^n(\underline{m}^W), q^n)$. Conditioned on a realization of the entire collection of public and semi-private codebooks, the private codewords $(V_j^n(\underline{m}^W, \underline{m}^U_{ij}, s_{ij}, \underline{m}^U_{jk}, s_{jk}, m^V_j, s_j) : s_j \in [\exp\{nS_j\}])$ are independent and identically distributed according to

$$p^n_{V_j|U_{ij}U_{jk}WQ}\left(\cdot|w^n(\underline{m}^W), u^n_{ij}(\underline{m}^W, \underline{m}^U_{ij}, s_{ij}), u^n_{jk}(\underline{m}^W, \underline{m}^U_{jk}, s_{jk}), q^n\right).$$

We now average error probability over the ensemble of codebooks. An upper bound on the error event at the encoder is derived using the method of second moment [59]. The probability of the error event at the encoder decays exponentially with n if for each triple $(i,j,k) \in \{(1,2,3),(2,3,1),(3,1,2)\}$

$$S_i > 0 \tag{5.1}$$

$$S_{ij} + S_{jk} > I(U_{ij}; U_{jk}|WQ) \tag{5.2}$$

$$S_{ij} + S_{jk} + S_{ki} > I(U_{ij}; U_{jk}; U_{ki}|WQ)^{12} \tag{5.3}$$

$$S_i + S_{ij} + S_{jk} + S_{ki} > I(U_{ij}; U_{jk}; U_{ki}|WQ) + I(V_i; U_{jk}|U_{ij}, U_{ki}, WQ) \tag{5.4}$$

$$\begin{aligned} S_i + S_j + S_{ij} + S_{jk} + S_{ki} > {} & I(V_i; U_{jk}|U_{ij}, U_{ki}, WQ) + I(V_j; U_{ki}|U_{ij}, U_{jk}, WQ) \\ & + I(U_{ij}; U_{jk}; U_{ki}|WQ) + I(V_i; V_j|U_{jk}, U_{ij}, U_{ki}, WQ) \end{aligned} \tag{5.5}$$

$$\begin{aligned} S_1 + S_2 + S_3 + S_{12} + S_{23} + S_{31} > {} & I(V_1; U_{23}|U_{12}, U_{31}, WQ) + I(V_2; U_{31}|U_{12}, U_{23}, WQ) + I(V_1; V_2; V_3|QW\underline{U}) \\ & + I(U_{12}; U_{23}; U_{31}|WQ) + I(V_3; U_{12}|U_{23}, U_{31}, WQ). \end{aligned} \tag{5.6}$$

The probability of decoder error event decays exponentially if for each triple $(i,j,k) \in \{(1,2,3),(2,3,1),(3,1,2)\}$

$$I(V_i; Y_i|QWU_{ij}U_{ki}) > T_i + S_i \tag{5.7}$$

$$I(U_{ij}V_i; Y_i|QWU_{ki}) + I(U_{ij}; U_{ki}|QW) > K_{ij} + L_{ij} + S_{ij} + T_i + S_i \tag{5.8}$$

$$I(U_{ki}V_i; Y_i|QWU_{ij}) + I(U_{ij}; U_{ki}|QW) > K_{ki} + L_{ki} + S_{ki} + T_i + S_i \tag{5.9}$$

$$I(U_{ij}U_{ki}V_i; Y_i|QW) + I(U_{ij}; U_{ki}|QW) > K_{ij} + L_{ij} + S_{ij} + K_{ki} + L_{ki} + S_{ki} + T_i + S_i \tag{5.10}$$

$$I(WU_{ij}U_{ki}V_i; Y_i|Q) + I(U_{ij}; U_{ki}|QW) > K_i + K_j + K_k + K_{ij} + L_{ij} + S_{ij} + K_{ki} + L_{ki} + S_{ki} + T_i + S_i \tag{5.11}$$

For each pmf $p_{QW\underline{UV}X}W_{\underline{Y}|X}$ defined on $\mathcal{Q} \times \mathcal{W} \times \underline{\mathcal{U}} \times \underline{\mathcal{V}} \times \mathcal{X} \times \underline{\mathcal{Y}}$, let $\alpha_{\mathscr{U}}(p_{QW\underline{UV}X\underline{Y}})$ denote the set of all triples $(R_1, R_2, R_3) \in [0,\infty)^4$ such that (i) there exists non-negative real numbers $K_{ij}, L_{ij}, S_{ij}, K_j, T_j, S_j$ that satisfies (5.1)-(5.11) for each pair $(i,j) \in \{(1,2),(2,3),(3,1)\}$ and (ii) $R_j = T_j + K_{jk} + L_{ij} + K_j$ for each triple $(i,j,k) \in$

[12] For three random variables, $A, B, C, I(A;B;C) = I(A;B) + I(AB;C)$.

$\{(1,2,3),(2,3,1),(3,1,2)\}$. The $\mathscr{UM}-$region is

$$\alpha_{\mathscr{U}}(W_{\underline{Y}|X},\kappa,\tau)=\mathrm{cocl}\left(\bigcup_{\substack{p_{QW\underline{UV}X\underline{Y}}\\ \in\mathbb{D}_{\mathscr{U}}(W_{\underline{Y}|X},\kappa,\tau)}}\alpha_{\mathscr{U}}(p_{QW\underline{UV}X\underline{Y}})\right),$$

where $\mathbb{D}_{\mathscr{U}}(W_{\underline{Y}|X},\kappa,\tau)$ denote the collection of distributions $p_{QW\underline{UV}X\underline{Y}}$ defined on $\mathcal{Q}\times\mathcal{W}\times\underline{\mathcal{U}}\times\underline{\mathcal{V}}\times\mathcal{X}\times\underline{\mathcal{Y}}$, where (i) $\mathcal{Q},\mathcal{W},\underline{\mathcal{U}},\underline{\mathcal{V}}$ are finite sets, (ii) $p_{\underline{Y}|X\underline{VU}WQ}=p_{\underline{Y}|X}=W_{\underline{Y}|X}$, (iii) $\mathbb{E}\{\kappa(X)\}\leq\tau$.

5.6 A vector additive $3-$DBC and a linear coding technique

In this section, we lay the groundwork for our first main finding - strict sub-optimality of $\mathscr{UM}-$technique. In particular, we identify a vector additive $3-$DBC (example Ex:3-BCExample) and propose a linear coding technique for the same. In section 5.10, we prove strict sub-optimality of $\mathscr{UM}-$technique for this vector additive $3-$DBC. We remark that even within the larger class of BC's that include continuous valued alphabets, any number of receivers and multiple antennae, we have been unaware, for over three decades, of any BC for which the coding techniques of superposition coding and precoding with binning can be strictly improved upon. The vector additive $3-$DBC presented herein is indeed a significant finding.

Example 5.6.1 *Consider the $3-$DBC depicted in figure 5.2. Let the input alphabet $\mathcal{X}=\mathcal{X}_1\times\mathcal{X}_2\times\mathcal{X}_3$ be a triple Cartesian product of the binary field $\mathcal{X}_1=\mathcal{X}_2=\mathcal{X}_3=\mathbb{F}_2$ and the output alphabets $\mathcal{Y}_1=\mathcal{Y}_2=\mathcal{Y}_3=\mathbb{F}_2$ be binary fields. If $X=X_1X_2X_3$ denote the three binary digits input to the channel, then the outputs are $Y_1=X_1\oplus X_2\oplus X_3\oplus N_1$, $Y_2=X_2\oplus N_2$ and $Y_3=X_3\oplus N_3$, where (i) N_1,N_2,N_3 are independent binary random variables with $P(N_j=1)=\delta_j\in(0,\frac{1}{2})$ and (ii) (N_1,N_2,N_3) is independent of the input X. The binary digit X_1 is constrained to an average Hamming weight of $\tau\in(0,\frac{1}{2})$. In other words, $\kappa(x_1x_2x_3)=1_{\{x_1=1\}}$ and the average cost of input is constrained to $\tau\in(0,\frac{1}{2})$. For the sake of clarity, we provide a formal description of this channel in terms of section 5.4.2. This $3-$DBC maybe referred to as $(\mathcal{X},\underline{\mathcal{Y}},W_{\underline{Y}|X},\kappa)$ where $\mathcal{X}:=\{0,1\}\times\{0,1\}\times\{0,1\},\mathcal{Y}_1=\mathcal{Y}_2=\mathcal{Y}_3=\{0,1\},W_{\underline{Y}|X}(y_1,y_2,y_3|x_1x_2x_3)=BSC_{\delta_1}(y_1|x_1\oplus x_2\oplus x_3)BSC_{\delta_2}(y_2|x_2)BSC_{\delta_3}(y_3|x_3)$, where $\delta_j\in(0,\frac{1}{2}):j=1,2,3$, $BSC_\eta(1|0)=BSC_\eta(0|1)=1-BSC_\eta(0|0)=1-BSC_\eta(1|1)=\eta$ for any $\eta\in(0,\frac{1}{2})$ and the cost function $\kappa(x_1x_2x_3)=1_{\{x_1=1\}}$.*

We begin with some observations for the above channel. Users 2 and 3 see *interference free point to point* links from the input. It is therefore possible to communicate to them simultaneously at their point to point capacities using any point to point channel codes achieving their respective capacities. For the purpose of this discussion, let us assume $\delta:=\delta_2=\delta_3$. This enables us employ the same capacity achieving code of rate $1-h_b(\delta)$ for both users 2 and 3. What about user 1? Three observations are in order. Firstly, if users 2 and 3 are being fed at their respective point

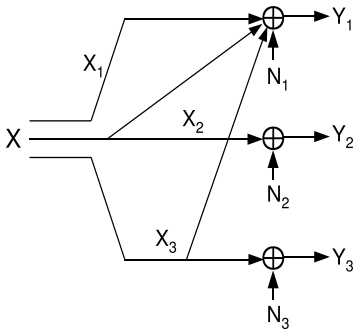

Figure 5.2: A 3−DBC with octonary input and binary outputs described in example 5.6.1.

to point capacities, then information can be pumped to user 1 only through the first binary digit, henceforth referred to as X_1. In this case, we recognize that the sum of user 2 and 3's transmissions interferes at receiver 1. Thirdly, the first binary digit X_1 is costed, and therefore cannot cancel the interference caused by users 2 and 3.

Since average Hamming weight of X_1 is restricted to τ, $X_1 \oplus N_1$ is restricted to an average Hamming weight of $\tau * \delta_1$. If the rates of users 2 and 3 are sufficiently small, receiver 1 can attempt to decode codewords transmitted to users 2 and 3, cancel the interference and decode the desired codeword. This will require $2 - 2h_b(\delta) \leq 1 - h_b(\delta_1 * \tau)$ or equivalently $\frac{1+h_b(\delta_1*\tau)}{2} \leq h_b(\delta)$. What if this were not the case?

In the case $\frac{1+h_b(\delta_1*\tau)}{2} > h_b(\delta)$, we are left with two choices. The first choice is to enable decoder 1 decode as large a part of the interference as possible and precode for the rest of the uncertainty.[13] The second choice is to attempt decoding the sum of user 2 and 3's codewords, instead of the pair. In the sequel, we pursue the second choice using linear codes. In section 5.10, we prove $\mathscr{UM}-$technique is forced to take the first choice which results in it's sub-optimality.

Since linear codes achieve capacity of binary symmetric channels, there exists a single linear code, or a coset thereof, of rate $1 - h_b(\delta)$ that achieves capacity of both user 2 and 3 channels. Let us employ this linear code for communicating to users 2 and 3. The code being linear or affine, the collection of sums of all possible pairs of codewords is restricted to a coset of rate $1 - h_b(\delta)$. This suggests that decoder 1 decode the sum of user 2 and 3 codewords. Indeed, if $1-h_b(\delta) \leq 1-h_b(\tau*\delta_1)$, or equivalently $\tau*\delta_1 \leq \delta$, then user 1 can first decode the interference, peel it off, and then go on to decode the desired signal. Under this case, a rate $h_b(\tau * \delta_1) - h_b(\delta_1)$ is achievable for user 1 even while communicating independent information at rate $1-h_b(\delta)$ for both users 2 and 3. We have therefore proposed a coding technique based on linear codes that achieves the rate triple $(h_b(\tau*\delta_1)-h_b(\delta_1), 1-h_b(\delta), 1-h_b(\delta))$ if $\tau * \delta_1 \leq \delta = \delta_2 = \delta_3$.

[13]Since X_1 is costed, precoding results in a rate loss, i.e., in terms of rate achieved, the technique of precoding is in general inferior to the technique of decoding interference. This motivates a preference for decoding the interference as against to precoding. However, for the Gaussian case, precoding suffers *no* rate loss. This is the precise reason for dirty paper coding being optimal for vector Gaussian BCs [61].

Let us now consider the general case with respect to δ_2, δ_3. Without loss of generality we may assume $\delta_2 \leq \delta_3$. We employ a capacity achieving linear code to communicate to user 2. This code is sub sampled (uniformly and randomly) to yield a capacity achieving code for user 3. This construction ensures the sum of all pairs of user 2 and 3 codewords to lie within user 2's linear code, or a coset thereof, of rate $1 - h_b(\delta_2)$. If $1 - h_b(\delta_2) \leq 1 - h_b(\tau * \delta_1)$, or equivalently $\tau * \delta_1 \leq \delta_2$, then decoder 1 can decode the sum of user 2 and 3's codewords, i.e., the interfering signal, peel it off and decode the desired message at rate $h_b(\tau * \delta_1) - h_b(\delta_1)$. If $\delta_3 \leq \delta_2$, then user 2's code is obtained by sub-sampling a capacity achieving linear code provided to user 3. In this case, user 1 can be fed at rate of $h_b(\tau * \delta_1) - h_b(\delta_1)$ if $1 - h_b(\delta_3) \leq 1 - h_b(\tau * \delta_1)$, or equivalently $\tau * \delta_1 \leq \delta_3$ The above arguments are summarized in the following lemma.

Lemma 5.6.2 *Consider the* $3-$*DBC in example 5.6.1. If* $\tau * \delta_1 \leq \min\{\delta_2, \delta_3\}$*, then* $(h_b(\tau * \delta_1) - h_b(\delta_1), 1 - h_b(\delta_2), 1 - h_b(\delta_3)) \in \mathbb{C}(\tau)$. □

In section 5.10, we prove that if $1 + h_b(\delta_1 * \tau) > h_b(\delta_2) + h_b(\delta_3)$, then $(h_b(\tau * \delta_1) - h_b(\delta_1), 1 - h_b(\delta_2), 1 - h_b(\delta_3)) \notin \alpha_{\mathscr{U}}(\tau)$. We therefore conclude in corollary 5.10.5 that if $\tau, \delta_1, \delta_2, \delta_3$ are such that $1 + h_b(\delta_1 * \tau) > h_b(\delta_2) + h_b(\delta_3)$ and $\min\{\delta_2, \delta_3\} \geq \delta_1 * \tau$, then $\mathscr{UM}-$technique is strictly suboptimal for the $3-$DBC presented in example 5.6.1. In particular, if $\tau, \delta_1, \delta = \delta_2 = \delta_3$ are such that $\frac{1+h_b(\delta_1 * \tau)}{2} > h_b(\delta) \geq h_b(\delta_1 * \tau)$, then $\mathscr{UM}-$technique is strictly suboptimal for the $3-$DBC presented in example 5.6.1. While the proof of this statement is long, the curious reader may sample our conclusion in theorem 5.10.4 and corollary 5.10.5.

5.7 Achievable rate regions for $3-$DBC using partitioned coset codes

In this section we present our second main finding - a new framework based on partitioned coset codes (PCC) for communicating over an arbitrary $3-$DBC - that enables us derive a new achievable rate region for the same. We present our framework in three pedagogical steps. Step I, presented in section 5.7.1, describes all the new elements of our framework in a simple setting. In particular, we employ PCC to manage interference seen by one receiver, and derive a corresponding achievable rate region. For this step, we also provide a complete and elaborate proof of achievability. Step II (section 5.7.2) builds on step I by incorporating private codebooks. Finally in step III (section 5.7.3), we employ PCC to manage interference seen by all receivers.

5.7.1 Step I: Using PCC to manage interference seen by a single receiver

Since this section describes the key elements of our findings in a simplified setting, the reader is strongly encouraged to study through the same carefully. We begin with a simple description of the coding technique. Subsequently, we formalize the same through a proof of achievability.

Description of the coding technique

The coding technique proposed herein is a very simple generalization of the linear coding technique proposed for example 5.6.1. The reader may find it useful to review the same. The two elements put forth in this step are (i) binning the linear codes into PCC to enable achieve rates corresponding to non-uniform distributions,[14] (ii) decoding the sum of user 2 and 3 codewords via the technique of joint typicality to enable analyze the performance of this decoding technique over arbitrary 3−DBC. We now state the coding technique.

Consider auxiliary alphabet sets $\mathcal{V}_1,\mathcal{U}_2,\mathcal{U}_3$ where $\mathcal{U}_2 = \mathcal{U}_3 = \mathcal{F}_\pi$ is the finite field of cardinality π and let $p_{V_1U_2U_3X\underline{Y}}$ be a pmf on $\mathcal{V}_1 \times \mathcal{U}_2 \times \mathcal{U}_3 \times \mathcal{X} \times \underline{\mathcal{Y}}$. Consider a random codebook $\mathcal{C}_1 \subseteq \mathcal{V}_1^n$ of rate $K_1 + R_1$ whose codewords are independently chosen according to $p_{V_1}^n$. Codewords of $\mathcal{C}_1$ are independently and uniformly partitioned into $\exp\{nR_1\}$ bins. For $j = 2, 3$, consider random partitioned coset codes (PCC) $(n, nS_j, nT_j, G_j, B_j^n, I_j)$ (definition 3.4.2) denoted Λ_j. The corresponding linear codes are nested, i.e., if $S_{j_1} \leq S_{j_2}$, then $G_{j_2}^t = \begin{bmatrix} G_{j_1}^t & G_{j_2/j_1}^t \end{bmatrix}$ where $G_{j_2/j_1} \in \mathcal{F}_\pi^{n(S_{j_2}-S_{j_1})\times n}$. $G_{j_1}, G_{j_2/j_1}, B_1^n, B_2^n, (I_j(a_j^{nS_j}) : a_j^{nS_j} \in \mathcal{F}_\pi^{nS_j}) : j = 2, 3$ are mutually independent and uniformly distributed over their respective range spaces. Moreover, random codebook $\mathcal{C}_1$ is independent of the pair Λ_2, Λ_3. We have thus specified the distribution of the triplet $\mathcal{C}_1, \Lambda_2, \Lambda_3$ of random codebooks. Messages of users $1, 2, 3$ at rates $R_1, T_2 \log \pi, T_3 \log \pi$ are used to index bins, one each in $\mathcal{C}_1, \Lambda_2, \Lambda_3$ respectively.[15] The encoder looks for a jointly typical triple, with respect to $p_{V_1U_2U_3}$, of codewords in the indexed triple of bins. Following a second moment method similar to that employed in [63], it can be proved that the encoder finds at least one jointly typical triple if for $j = 2, 3$

$$K_1 > 0, \quad (S_j - T_j)\log\pi > \log\pi - H(U_j), \quad (S_j - T_j)\log\pi + K_1 > \log\pi - H(U_j) + I(U_j; V_1), \tag{5.12}$$

$$\sum_{j=2}^{3}(S_j - T_j)\log\pi > 2\log\pi - H(U_2) - H(U_3) + I(U_2; U_3) \tag{5.13}$$

$$K_1 + \max\{S_2, S_3\}\log\pi > \log\pi - H(U_2 \oplus U_3) + I(V_1; U_2 \oplus U_3), \quad \max\{S_2, S_3\}\log\pi \geq \log\pi - H(U_2 \oplus U_3) \tag{5.14}$$

$$\sum_{j=2}^{3}(S_j - T_j)\log\pi + K_1 > 2\log\pi - \sum_{j=2}^{3}H(U_j) + I(U_2; U_3; V_1). \tag{5.15}$$

Having chosen one such jointly typical triple, say V_1^n, U_2^n, U_3^n, it generates a vector X^n according to

$$p_{X|V_1U_2U_3}^n(\cdot|V_1^n, U_2^n, U_3^n) = \prod_{t=1}^{n} p_{X|V_1U_2U_3}(\cdot|V_{1t}, U_{2t}, U_{3t})$$

and feeds the same as input, on the channel.

Decoders 2 and 3 perform a standard PTP decoding. For example, decoder 2 receives Y_2^n and looks for all codewords in Λ_2 that are jointly typical with Y_2^n. If it finds all such codewords in a unique bin it declares the

[14] This is akin to binning for channels with state information, wherein $\exp\{nI(U;S)\}$ codewords, each picked according to $\prod_{t=1}^{n} p_U$, are chosen for each message in order to find a codeword in $T_\delta(U|s^n)$ jointly typical with state sequence s^n.

[15] For $j = 2, 3$, user j's codebook of block length n must provide $\exp\{nT_j \log\pi\} = \pi^{nT_j}$ bins. Indeed Λ_j contains π^{nT_j} bins.

corresponding bin index as the decoded message. It can be proved by following the technique similar to [63, Proof of Theorem 1] that if

$$S_j < \log \pi - H(U_j|Y_j) \text{ for } j = 2, 3 \tag{5.16}$$

then probability of decoding error at decoders 2 and 3 can be made arbitrarily small for sufficiently large n.

Having received Y_1^n, decoder 1 looks for all codewords $v_1^n \in \mathcal{C}_1$ for which there exists a codeword $u_{2\oplus 3}^n \in \Lambda_2 \oplus \Lambda_3$ such that $(v_1^n, u_{2\oplus 3}^n, Y_1^n)$ are jointly typical with respect to $p_{V_1, U_2 \oplus U_3, Y_1}$. Here

$$\Lambda_2 \oplus \Lambda_3 := \left\{ U_2^n \oplus U_3^n : U_j^n \in \Lambda_j^n : j = 2, 3 \right\}.$$

If all such codewords in $\mathcal{C}_1$ belong to a unique bin, the corresponding bin index is declared as the decoded message. Again following the technique similar to [63, Proof of Theorem 1], it can be proved, that if, for $j = 2, 3$

$$K_1 + R_1 < H(V_1) - H(V_1|U_2 \oplus U_3, Y_1), \;\; K_1 + R_1 + (S_j + T_j) \log \pi < \log \pi + H(V_1) - H(V_1, U_2 \oplus U_3|Y_1), \tag{5.17}$$

then probability of decoding error at decoder 1 falls exponentially with n. In the sequel, we provide a formal proof of achievability. We begin with a characterization of the rate region proved achievable herein. For completeness, we include a random variable for time sharing in it's description.

Proof of achievability

Definition 5.7.1 *Let $\mathbb{D}_1^f(W_{\underline{Y}|X}, \kappa, \tau)$ denote the collection of pmfs $p_{QV_1U_2U_3X\underline{Y}}$ defined on $\mathcal{Q} \times \mathcal{V}_1 \times \mathcal{U}_2 \times \mathcal{U}_3 \times \mathcal{X} \times \underline{\mathcal{Y}}$, where (i) $\mathcal{Q}, \mathcal{V}_1$ are finite sets, $\mathcal{U}_2 = \mathcal{U}_3$ is a finite field, (ii) $p_{\underline{Y}|XV_1\underline{U}} = p_{\underline{Y}|X} = W_{\underline{Y}|X}$, [16] and (iii) $\mathbb{E}\{\kappa(X)\} \leq \tau$.*

Definition 5.7.2 *Consider $p_{QV_1\underline{U}X\underline{Y}} \in \mathbb{D}_1^f(W_{\underline{Y}|X}, \kappa, \tau)$ and let $\pi := |\mathcal{U}_2| = |\mathcal{U}_3|$. Let $\beta_1(p_{QV_1\underline{U}X\underline{Y}})$ be defined as the set of rate triples $\underline{R} := (R_1, R_2, R_3) \in [0, \infty)^3$ for which $\mathcal{S}(\underline{R}, p_{QV_1\underline{U}X\underline{Y}}, 0)$ is non-empty, where, for any $\delta > 0$, $\mathcal{S}(\underline{R}, p_{QV_1\underline{U}X\underline{Y}}, \delta)$ is defined as the set of vectors $(K_1, R_1, S_2, T_2, S_3, T_3) \in [0, \infty)^6$ that satisfy $R_j = T_j \log \pi$,*

$$K_1 > \delta, \quad (S_j - T_j) \log \pi > \log \pi - H(U_j|Q) + \delta, \tag{5.18}$$

$$K_1 + (S_j - T_j) \log \pi > \log \pi - H(U_j|Q, V_1) + \delta, \quad \sum_{l=2}^{3} (S_l - T_l) \log \pi > 2 \log \pi - H(\underline{U}|Q) + \delta, \tag{5.19}$$

$$K_1 + \sum_{l=2}^{3} (S_l - T_l) \log \pi > 2 \log \pi - H(\underline{U}|Q, V_1) + \delta, \quad \max\{S_2, S_3\} \log \pi > \log \pi - H(U_2 \oplus U_3|Q) + \delta, \tag{5.20}$$

$$K_1 + \max\{S_2, S_3\} \log \pi \overset{(a)}{>} \log \pi - H(U_2 \oplus U_3|Q, V_1) + \delta, \quad K_1 + R_1 < I(V_1; Y_1, U_2 \oplus U_3|Q) - \delta, \tag{5.21}$$

$$K_1 + R_1 + \max\{S_2, S_3\} \log \pi < \log \pi + H(V_1|Q) - H(V_1, U_2 \oplus U_3|Q, Y_1) - \delta$$

$$S_j \log \pi < \log \pi - H(U_j|Q, Y_j) - \delta,$$

[16] *In this subsection, $\underline{U}$ denotes the pair U_2, U_3. Similarly, the other objects such as $\underline{\mathcal{U}}, \underline{u}$ denote corresponding pairs.*

for $j = 2, 3$. Furthermore, let

$$\beta_1(W_{\underline{Y}|X}, \kappa, \tau) := cocl\left(\bigcup_{\substack{p_{QV_1\underline{U}X\underline{Y}} \in \\ \mathbb{D}_1^f(W_{\underline{Y}|X}, \kappa, \tau)}} \beta_1(p_{QV_1\underline{U}X\underline{Y}})\right).$$

Theorem 5.7.3 *For a 3−DBC $(\mathcal{X}, \underline{\mathcal{Y}}, W_{\underline{Y}|X}, \kappa)$, $\beta_1(W_{\underline{Y}|X}, \kappa, \tau)$ is achievable, i.e., $\beta_1(W_{\underline{Y}|X}, \kappa, \tau) \subseteq \mathbb{C}(W_{\underline{Y}|X}, \kappa, \tau)$.* □

Proof: Given $p_{QV_1\underline{U}X\underline{Y}} \in \mathbb{D}_1^f(W_{\underline{Y}|X}, \kappa, \tau)$, $\underline{R} \in \beta_1(p_{QV_1\underline{U}X\underline{Y}}), \tilde{\eta} > 0$, our task is to identify a 3−DBC code $(n, \underline{\mathcal{M}}, e, \underline{d})$ of rate $\frac{\log \mathcal{M}_j}{n} \geq R_j - \tilde{\eta} : j = 1, 2, 3$, average error probability $\overline{\xi}(e, \underline{d}) \leq \tilde{\eta}$, and average cost $\tau(e) \leq \tau + \tilde{\eta}$.

For the given rate triple $\underline{R} \in \beta_1(p_{QV_1\underline{U}X\underline{Y}})$, we have $\delta_1 > 0$ and $(K_1, R_1, S_2, T_2, S_3, T_3) \in \mathcal{S}(\underline{R}, p_{QV_1\underline{U}X\underline{Y}}, \delta_1)$. Set $\eta := \min\{\tilde{\eta}, \delta_1\}$. Consider a codebook $\mathcal{C}_1 = (v_1^n(m_1, b_1) : m_1 \in \mathcal{M}_1, b_1 \in \mathcal{B}_1)$ built over $\mathcal{V}_1$ consisting of $|\mathcal{M}_1|$ bins, each consisting of $|\mathcal{B}_1|$ codewords. We let $\mathcal{M}_1 = [\lfloor \exp\{n(R_1 - \frac{\eta}{2})\} \rfloor]$ and $\mathcal{B}_1 = [\lceil \exp\{n(K_1 + \frac{\eta}{8})\} \rceil]$. $\mathcal{C}_1$ is employed to encode user 1's message. Codebooks employed to encode user 2 and 3's messages are partitioned coset codes (definitions 3.4.2) which are described in the sequel. Henceforth, we let $\pi := |\mathcal{U}_2| = |\mathcal{U}_3|$ and therefore $\mathcal{F}_\pi = \mathcal{U}_2 = \mathcal{U}_3$. Consider a linear code $\overline{\lambda} \subseteq \mathcal{F}_\pi^n$ with generator matrix $g \in \mathcal{F}_\pi^{s \times n}$ and let $\lambda \subseteq \mathcal{F}_\pi^n$ denote the coset of $\overline{\lambda}$ with respect to shift $b^n \in \mathcal{F}_\pi^n$. Clearly, the codewords of λ are given by $u(a^s) := a^s g \oplus b^n : a^s \in \mathcal{F}_\pi^s$. Consider a partition of λ into π^t bins. Each codeword $u(a^s)$ is assigned a bin index $i(a^s) \in \mathcal{F}_\pi^t$. For every $m^t \in \mathcal{F}_\pi^t$, $c(m^t) := \{a^s : i(a^s) = m^t\}$ denotes the set of indices whose codewords are assigned to bin m^t. The coset code λ with it's partitions is called a *partitioned coset code* and denoted (n, s, t, g, b^n, i).[17]

For $j = 2, 3$, user j is provided the partitioned coset code $(n, s_j, t_j, g_j, b_j^n, i_j)$, where $s_j = \lfloor nS_j \rfloor, t_j := \lceil n(T_j - \frac{\eta}{4 \log \pi}) \rceil$. Let $u_j^n(a_j^{s_j}) := a_j^{s_j} g_j \oplus b_j^n$ denote a generic codeword in λ_j and $c_j(m_j^{t_j}) := \left\{a_j^{s_j} : i_j(a_j^{s_j}) = m_j^{t_j}\right\}$ denote the indices of codewords in bin corresponding to message $m_j^{t_j}$. These codes are such that if $s_{j_1} \leq s_{j_2}$, then $g_{j_2}^t = \begin{bmatrix} g_{j_1}^t & g_{j_2/j_1}^t \end{bmatrix}$. In other words, the linear code corresponding to the larger coset code contains the linear code corresponding to the smaller coset code. Without loss of generality, we henceforth assume $s_2 \leq s_3$ and therefore $g_3^t = \begin{bmatrix} g_2^t & g_{3/2}^t \end{bmatrix}$. It is now appropriate to derive some relationships between the code parameters that would be of use at a later time. There exists $N_1(\eta) \in \mathbb{N}$ such that for all $n \geq N_1(\eta)$

$$nS_j - 1 \leq s_j \leq nS_j \text{ and therefore } S_j - \tfrac{\eta}{8 \log \pi} \leq S_j - \tfrac{1}{n} \leq \tfrac{s_j}{n} \leq S_j, \tag{5.22}$$

$$n\left(T_j - \tfrac{\eta}{4 \log \pi}\right) \leq t_j \leq n\left(T_j - \tfrac{\eta}{4 \log \pi}\right) + 1 \text{ and therefore } T_j - \tfrac{\eta}{4 \log \pi} \leq \tfrac{t_j}{n} \leq T_j - \tfrac{\eta}{8 \log \pi}, \tag{5.23}$$

$$R_1 - \eta \leq \tfrac{\log |\mathcal{M}_1|}{n} \leq R_1 - \tfrac{\eta}{2} \text{ and } K_1 + \tfrac{\eta}{8} \leq \tfrac{\log |\mathcal{B}_1|}{n} \leq K_1 + \tfrac{\eta}{4}. \tag{5.24}$$

We now describe the encoding and decoding rules. A vector $q^n \in T_{\eta_2}(Q)$ is chosen to be the time sharing vector, where

[17] A careful and diligent reader who has studied through definitions 3.4.2 and 4.6.3 will note a minor difference between those and the one stated here. In definitions 3.4.2 and 4.6.3, the set indexing the partitions was chosen to be $[\pi^t]$. Here the corresponding set is $\mathcal{F}_\pi^t$.

η_2 will be specified in due course. Without loss of generality, we assume the message sets are $\mathcal{M}_j := \mathcal{F}_\pi^{t_j}$ for $j = 2, 3$ and as stated before $\mathcal{M}_1 := [\lfloor \exp\left\{n(R_1 - \frac{\eta}{2})\right\} \rfloor]$. Let $(M_1, M_2^{t_2}, M_3^{t_3}) \in \underline{\mathcal{M}}$ denote the uniformly distributed triple of message random variables to be communicated to the respective users. Having received $(M_1, M_2^{t_2}, M_3^{t_3})$, the encoder looks for a triple of codewords in the indexed bin of codewords that are jointly typical. Formally, the encoder looks for a triplet $(b_1, a_2^{s_2}, a_3^{s_3}) \in \mathcal{B}_1 \times c_2(M_2^{t_2}) \times c_3(M_3^{t_3})$ such that $(v_1^n(M_1, b_1), u_2^n(a_2^{s_2}), u_3^n(a_3^{s_3})) \in T_{2\eta_2}(V_1, U_2, U_3|q^n)$.[18] If it finds at least one such triple, one of them is chosen according to a predefined rule. Otherwise, i.e, if it finds no triple of codewords in the indexed triple of bins that is jointly typical, it chooses a fixed triple of codewords in $\mathcal{C}_1 \times \lambda_2 \times \lambda_3$. In either case, let $(v_1^n(M_1, B_1), u_2^n(A_2^{s_2}), u_3(A_3^{s_3}))$ denote the chosen triple of codewords. In the former case, the encoder maps the triple to a vector in $T_{4\eta_2}(X|v_1^n(M_1, B_1), u_2^n(A_2^{s_2}), u_3(A_3^{s_3}))$ and feeds the same as input on the channel. In the latter case, it picks a fixed vector in $\mathcal{X}^n$ and feeds the same as input on the channel. In either case, let $x^n(M_1, M_2^{t_2}, M_3^{t_3})$ denote the vector input on the channel.

The operations of decoders 2 and 3 are identical and we describe the same through the generic index j. Having received vector Y_j^n, it looks for all messages $\hat{m}_j^{t_j} \in \mathcal{M}_j$ such that for some $a_j^{s_j} \in c_j(\hat{m}_j^{t_j})$, $u_j(a_j^{s_j}) \in T_{8\eta_2}(U_j|q^n, Y_j^n)$. If it finds exactly one such message, this is declared as the decoded message. Otherwise, an error is declared.

Decoder 1 is provided with the codebook $\lambda_2 \oplus \lambda_3 := \left\{u_2^n(a_2^{s_2}) \oplus u_3^n(a_3^{s_3}) : a_j^{s_j} \in \mathcal{F}_\pi^{s_j} : j = 2, 3\right\}$. Note that $\lambda_2 \oplus \lambda_3 = \{u_\oplus(a_3^{s_3}) := a_3^{s_3} g_3 \oplus b_2^n \oplus b_3^n : a_3^{s_3} \in \mathcal{F}_\pi^{s_3}\}$. Having received Y_1^n, decoder 1 looks for all messages $\hat{m}_1 \in \mathcal{M}_1$ such that $(v_1^n(\hat{m}_1, b_1), u_\oplus(a_3^{s_3})) \in T_{8\eta_2}(V_1, U_2 \oplus U_3|q^n, Y_1^n)$ for some $(b_1, a_3^{s_3}) \in \mathcal{B}_1 \times \mathcal{F}_\pi^{s_3}$. If it finds exactly one such $\hat{m}_1 \in \mathcal{M}_1$, this is declared as the decoded message. Otherwise, an error is declared.

The above encoding and decoding rules map a triplet $\mathcal{C}_1, \lambda_2, \lambda_3$ of codebooks into a 3−DBC code[19]. Moreover, (5.23) and (5.24) imply that the rates of the corresponding 3−DBC code satisfy $\frac{\log \mathcal{M}_1}{n} \geq R_1 - \eta$, $\frac{t_j \log \pi}{n} \geq R_j - \frac{\tilde{\eta}}{4}$ for $j = 2, 3$. Since every triple $\mathcal{C}_1, \lambda_2, \lambda_3$ of codebooks, and a choice for the predefined rules map to a corresponding 3−DBC code, we have characterized an ensemble of 3−DBC codes, one for each $n \in \mathbb{N}$. We now induce a distribution over this ensemble of 3−DBC codes.

Consider a random triple $\mathcal{C}_1, \Lambda_2, \Lambda_3$ of codebooks, where $\mathcal{C}_1 = (V_1^n(m_1, b_1) : (m_1, b_1) \in \mathcal{M}_1 \times \mathcal{B}_1)$ and Λ_j is the random partitioned coset code $(n, s_j, t_j, G_j, B_j^n, I_j)$. Note that the joint distribution of $V_1^n(m_1, b_1) : (m_1, b_1) \in \mathcal{M}_1 \times \mathcal{B}_1, G_2, G_{3/2}, B_2^n, B_3^n, I_2(a_2^{s_2}) : a_2^{s_2} \in \mathcal{F}_\pi^{s_2}, I_3(a_3^{s_3}) : a_3^{s_3} \in \mathcal{F}_\pi^{s_3}$ uniquely characterizes the distribution of $\mathcal{C}_1, \Lambda_2, \Lambda_3$. We let $V_1^n(m_1, b_1) : (m_1, b_1) \in \mathcal{M}_1 \times \mathcal{B}_1, G_2, G_{3/2}, B_2^n, B_3^n, I_2(a_2^{s_2}) : a_2^{s_2} \in \mathcal{F}_\pi^{s_2}, I_3(a_3^{s_3}) : a_3^{s_3} \in \mathcal{F}_\pi^{s_3}$ be mutually independent. For every $(m_1, b_1) \in \mathcal{M}_1 \times \mathcal{B}_1$, $v_1^n \in \mathcal{V}_1^n$, let $P(V_1^n(m_1) = v_1^n) = \prod_{t=1}^n p_{V_1|Q}(v_{1t}|q_t)$. The rest of the random objects $G_2, G_{3/2}, B_2^n, B_3^n, I_2(a_2^{s_2}) : a_2^{s_2} \in \mathcal{F}_\pi^{s_2}, I_3(a_3^{s_3}) : a_3^{s_3} \in \mathcal{F}_\pi^{s_3}$ are uniformly distributed over their respective range spaces. We have therefore specified the distribution of the random triple $\mathcal{C}_1, \Lambda_2, \Lambda_3$ of codebooks. For $j = 2, 3$, we let $U_j^n(a_j^{s_j}) = a_j^{s_j} G_j \oplus B_j^n$ denote a generic random codeword in the random codebook Λ_j. Likewise,

[18] Here, the typicality is with respect to $p_{Q V_1 \underline{U} X \underline{Y}}$.

[19] This map also relies on a 'predefined' rule to choose among many jointly typical triples within an indexed pair of bins and furthermore, a rule to decide among many input sequences that is conditionally typical with this chosen triple of codewords.

we let $U^n_{\oplus}(a_3^{s_3}) = a_3^{s_3}G_3 \oplus B_2^n \oplus B_3^n$ denote a generic codeword in $\Lambda_2 \oplus \Lambda_3$. Let $(V_1^n(M_1, B_1), U_2^n(A_2^{s_2}), U_3^n(A_3^{s_3}))$ denote the triple of codewords chosen by the encoder and $X^n(M_1, M_2^{t_2}, M_3^{t_3})$ denote the vector input on the channel.

While the above specifies the distribution of the random triple of $\mathcal{C}_1, \Lambda_2, \Lambda_3$ of codebooks, the predefined rules that map it to a 3−DBC code is yet unspecified. In other words, the distribution of $(V_1^n(M_1, B_1), U_2^n(A_2^{s_2}), U_3^n(A_3^{s_3}))$ and $X^n(M_1, M_2^{t_2}, M_3^{t_3})$ needs to be specified. All the 3−DBC codes that a particular triplet of codebooks $\mathcal{C}_1, \lambda_2, \lambda_3$ map to, are uniformly distributed. Alternatively, the encoder picks a triple in

$$\left\{(V_1^n(M_1, b_1), U_2(a_2^{s_2}), U_3(a_3^{s_3})) \in T_{2\eta_2}(V_1, \underline{U}|q^n) : (b_1, a_2^{s_2}, a_3^{s_3}) \in \mathcal{B}_1 \times C_2(M_2^{t_2}) \times C_3(M_3^{t_3})\right\}$$

uniformly at random and independent of other choices. Denoting this random triple as $(V_1^n(M_1, B_1), U_2^n(A_2^{s_2}), U_3^n(A_3^{s_3}))$, the encoder picks an input sequence in $T_{2\eta_2}(X|(V_1^n(M_1, B_1), U_2^n(A_2^{s_2}), U_3^n(A_3^{s_3})))$ uniformly at random and independent of other choices. We have therefore specified the distribution induced on the corresponding ensemble of 3−DBC codes. In the sequel, we characterize error events associated with this random 3−DBC code.

If

$$\begin{aligned}
\epsilon_1 &:= \bigcap_{\substack{(b_1, a_2^{s_2}, a_3^{s_3}) \\ \mathcal{B}_1 \times C_2(M_2^{t_2}) \times C_3(M_3^{t_3})}} \{(V_1(M_1, b_1), U_2(a_2^{s_2}), U_3(a_3^{s_3})) \notin T_{2\eta_2}(V_1, U_2, U_3|q^n)\} \\
\epsilon_{31} &:= \bigcap_{\substack{(b_1, a_3^{s_3}) \\ \in \mathcal{B}_1 \times \mathcal{F}_\pi^{s_3}}} \left\{(V_1(M_1, b_1), U^n_{\oplus}(a_3^{s_3}), Y_1^n) \notin T_{8\eta_2}(V_1, U_2 \oplus U_3, Y_1|q^n)\right\}, \\
\epsilon_{3j} &:= \bigcap_{a_j^{s_j} \in C_j(M_j^{t_j})} \left\{(U_j(a_j^{s_j}), Y_j^n) \notin T_{8\eta_2}(U_j, Y_j|q^n)\right\} \\
\epsilon_{41} &:= \bigcup_{\substack{(b_1, a_3^{s_3}) \\ \in \mathcal{B}_1 \times \mathcal{F}_\pi^{s_3}}} \bigcup_{\hat{m}_1 \neq M_1} \left\{(V_1(\hat{m}_1, b_1), U^n_{\oplus}(a_3^{s_3}), Y_1^n) \in T_{8\eta_2}(V_1, Y_1|q^n)\right\}, \\
\epsilon_{4j} &:= \bigcup_{\substack{a_j^{s_j} \in C_j(\hat{m}_j^{t_j}) \\ \hat{m}_j^{t_j} \neq M_j^{t_j}}} \left\{(U_j(a_j^{s_j}), Y_j^n) \in T_{8\eta_2}(U_j, Y_j|q^n)\right\},
\end{aligned}$$

then $\epsilon := \bigcup_{j=1}^{3} (\epsilon_1 \cup \epsilon_{3j} \cup \epsilon_{4j})$ contains the error event. Our next task is to derive an upper bound on $P(\epsilon)$.

Let

$$\begin{aligned}
\phi(m_1, m_2^{t_2}, m_3^{t_3}) &:= \sum_{\substack{(b_1, a^{s_2}, a^{s_3}) \in \\ \mathcal{B}_1 \times \mathcal{F}_\pi^{s_2} \times \mathcal{F}_\pi^{s_3}}} 1_{\left\{(V_1^n(m_1, b_1), U_2(a^{s_2}), U_3(a^{s_3})) \in T_{2\eta_2}(V_1, U_2, U_3|q^n), I(a^{s_j}) = m_j^{s_j} : j=2,3\right\}}, \\
\epsilon_l &:= \left\{\phi(M_1, M_2^{t_2}, M_3^{t_3}) < \mathcal{L}(n)\right\}, \text{ where } \mathcal{L}(n) := \frac{1}{2}\mathbb{E}\left\{\phi(M_1, M_2^{t_2}, M_3^{t_3})\right\}.
\end{aligned}$$

Clearly $P(\epsilon) \leq P(\epsilon_l) + P(\epsilon_l^c \cap \epsilon)$, and it therefore suffices to derive upper bounds on each of these terms.

Upper bound on $P(\epsilon_l)$:- Substituting for $\mathcal{L}(n)$, we have

$$P(\epsilon_l) \leq P(|\phi(M_1, M_2^{t_2}, M_3^{t_3}) - \mathbb{E}\left\{\phi(M_1, M_2^{t_2}, M_3^{t_3})\right\}| \geq \frac{\mathbb{E}\left\{\phi(M_1, M_2^{t_2}, M_3^{t_3})\right\}}{2}). \leq \frac{4\text{Var}\left\{\phi(M_1, M_2^{t_2}, M_3^{t_3})\right\}}{\left(\mathbb{E}\left\{\phi(M_1, M_2^{t_2}, M_3^{t_3})\right\}\right)^2}$$

from the Cheybyshev inequality. In appendix J, we evaluate the variance and expectation of $\phi(M_1, M_2^{t_2}, M_3^{t_3})$ and derive an upper bound on $P(\epsilon_l)$. In particular, we prove for $n \geq \max\{N_1(\eta), N_2(\eta_2)\}$,

$$P(\epsilon_1) \leq (28 + 8\pi) \exp\left\{-n\left(\delta_1 - \frac{\eta}{8} - 48\eta_2\right)\right\}. \tag{5.25}$$

In deriving the above, we have employed lower bounds (5.18), (5.19), (5.20) and (5.21)(a).

Now consider $\epsilon_l^c \cap \epsilon_1$. Note that $P(\epsilon_1) = P(\phi(M_1, M_2^{t_2}, M_3^{t_3}) = 0)$, and hence $\epsilon_l^c \cap \epsilon_1 = \phi$, the empty set, if $\mathcal{L}(n) > 1$. At the end of appendix J, we prove $\mathcal{L}(n) > 1$ for sufficiently large n. We are left to derive an upper bound on $P(\epsilon_l^c \cap \bigcup_{j=1}^{3} (\epsilon_{3j} \cup \epsilon_{4j}))$.

Since $\mathcal{L}(n) > 1$, $\epsilon_l^c \subseteq \epsilon_1^c$, it suffices to derive an upper bound on the terms $P(\epsilon_1^c \cap (\epsilon_{31} \cup \epsilon_{32} \cup \epsilon_{33}))$, $P(\epsilon_l^c \cap (\epsilon_{31} \cup \epsilon_{32} \cup \epsilon_{33})^c \cap \epsilon_{4j}) : j = 1, 2, 3$.

Upper bound on $P(\epsilon_1^c \cap (\epsilon_{31} \cup \epsilon_{32} \cup \epsilon_{33}))$:- Consider $P(\epsilon_1^c \cap \epsilon_2)$, where

$$\epsilon_2 := \left\{(V_1(M_1, B_1), U_2(A_2^{s_2}), U_3(A_3^{s_3}), X^n) \notin T_{4\eta_2}(V_1, \underline{U}, X|q^n)\right\}.$$

By the encoding rule $P(\epsilon_1^c \cap \epsilon_2) = 0$. Since the encoding rule also ensures $\epsilon_1^c \cap (\epsilon_{31} \cup \epsilon_{32} \cup \epsilon_{33}) \subseteq \epsilon_1^c \cap \epsilon_3$, where

$$\epsilon_3 := \left\{(V_1^n(M_1, B_1), U_2^n(A_2^{s_2}), U_3^n(A_3^{s_3}), X^n(M_1, M_2^{t_2}, M_3^{t_3}), \underline{Y}^n) \notin T_{8\eta_2}(V_1, \underline{U}, X, \underline{Y})\right\},$$

it suffices to derive an upper bound on $P((\epsilon_1 \cup \epsilon_2)^c \cap \epsilon_3)$. This follows from conditional frequency typicality (lemma 2.4.1) and $p_{\underline{Y}|XV_1\underline{U}Q} = p_{\underline{Y}|X} = W_{\underline{Y}|X}$ (statement (ii) of definition 5.7.1). We conclude the existence of $N_3(\eta_2)$ such that for all $n \geq N_4(\eta_2)$, $P((\epsilon_1 \cup \epsilon_2)^c \cap \epsilon_3) \leq \frac{\eta}{32}$.

Upper bound on $P((\epsilon_l \cup \epsilon_2 \cup \epsilon_3)^c \cap \epsilon_{41})$: We refer the reader to appendix K for the derivation of an upper bound on $P((\epsilon_1 \cup \epsilon_2 \cup \epsilon_3)^c \cap \epsilon_{41})$. Therein, we prove existence of $N_4(\eta_2) \in \mathbb{N}$ such that for all $n \geq \max\{N_1(\eta), N_4(\eta_2)\}$, we have

$$P((\epsilon_l \cup \epsilon_2 \cup \epsilon_3)^c \cap \epsilon_{41}) \leq 4 \exp\left\{-n\left(\delta_1 + \frac{\eta}{4} - 56\eta_2\right)\right\}. \tag{5.26}$$

Upper bound on $P((\epsilon_l \cup \epsilon_2 \cup \epsilon_3)^c \cap \epsilon_{4j})$: For $j = 2, 3$, decoder j performs a simple point-to-point decoding and therefore the reader might expect the analysis here to be quite standard. The partitioned coset code structure of user j's codebook that involves correlated codewords and bins lends some technical complexities. We flesh out the details

in appendix L. In particular, we prove (L.5) existence of $N_5(\eta_2) \in \mathbb{N}$ such that for all $n \geq \max\{N_1(\eta), N_5(\eta_2)\}$

$$P((\epsilon_1 \cup \epsilon_2 \cup \epsilon_3)^c \cap \epsilon_{4j}) \leq 2\exp\left\{-n\left(\delta_1 - 32\eta_2\right)\right\}. \tag{5.27}$$

Let us now compile the upper bounds derived in (5.25), (5.26) and (5.27). For $n \geq \max\{N_1(\eta), N_2(\eta_2)N_3(\eta_2), N_4(\eta_2), N_5(\eta_2)\}$, we have

$$P(\epsilon_1 \cup \epsilon_2 \cup \epsilon_3 \cup \epsilon_{41} \cup \epsilon_{42}) \leq \frac{\eta}{32} + (34 + 8\pi)\exp\left\{-n\left(\delta_1 - \frac{\eta}{8} - 56\eta_2\right)\right\}. \tag{5.28}$$

Recall that η is chosen to be $\min\{\tilde{\eta}, \delta_1\}$. By choosing $\eta_2 = \frac{\eta}{56\times 8}$, we have $\delta_1 - \frac{\eta}{8} - \frac{\eta}{8} > \frac{3\eta}{4}$ and we can drive the probability of error below $\tilde{\eta}$ by choosing n sufficiently large.

The only element left to argue is the random code satisfies the cost constraint. Since $P(\epsilon_1 \cup \epsilon_2)$ is lesser than $\frac{\tilde{\eta}}{2}$ for sufficiently large n, the encoder inputs a vector on the channel that is typical with respect p_X with probability $1 - \frac{\tilde{\eta}}{2}$. Since $\mathbb{E}\{\kappa(X)\} \leq \tau$, a standard argument proves that the expected cost of the input vector can be made arbitrarily close to τ by choosing n sufficiently large and η_2 sufficiently small. We leave the details to the reader. ■

The coding technique that yields achievability of $\beta_1(W_{\underline{Y}|X}, \kappa, \tau)$ is a simple generalization of the linear coding technique proposed for example 5.6.1. Therefore, it can be verified that, if $\tau * \delta_1 \leq \min\{\delta_2, \delta_3\}$, then $(h_b(\tau * \delta_1) - h_b(\delta_1), 1 - h_b(\delta_2), 1 - h_b(\delta_3)) \in \tilde{\beta}_1(W_{\underline{Y}|X}, \kappa, \tau)$. We leave it to the reader to verify that if $\tau * \delta_1 \leq \min\{\delta_2, \delta_3\}$, then $(h_b(\tau * \delta_1) - h_b(\delta_1), 1 - h_b(\delta_2), 1 - h_b(\delta_3)) \in \beta_1(p_{\underline{U}V_1X\underline{Y}})$, where $p_{\underline{U}V_1X} = p_{V_1}p_{U_2}p_{U_3}1_{\{X_1=V_1\}}1_{\{X_2=U_2\}}1_{\{X_3=U_3\}}$, $p_{U_2}(1) = p_{U_3}(1) = \frac{1}{2}$ and $p_{V_1}(1) = \tau$.

5.7.2 Step II: Incorporating private codebooks

We revisit the coding technique proposed in section 5.7.1. Observe that (i) user 1 decodes a sum of the entire codewords/signals transmitted to users 2 and 3 and (ii) users 2 and 3 decode only their respective codewords. This technique may be enhanced in the following way. User 1 can decode the sum of *one component* of user 2 and 3 signals each. In other words, we may include private codebooks for users 2 and 3.

We begin with a description of the coding technique. In addition to the codebooks $\mathcal{C}_1, \Lambda_2, \Lambda_3$ described in section 5.7.1, we incorporate private layer codebooks for users 2 and 3. Specifically, in addition to auxiliary alphabet sets $\mathcal{V}_1, \mathcal{U}_2, \mathcal{U}_3$ introduced in section 5.7.1, let $\mathcal{V}_2, \mathcal{V}_3$ denote arbitrary finite sets and $p_{U_2U_3V_1V_2V_3}$ denote a pmf on $\mathcal{U}_2 \times \mathcal{U}_3 \times \mathcal{V}_1 \times \mathcal{V}_2 \times \mathcal{V}_3$. For $j = 2, 3$, consider a random codebook $\mathcal{C}_j \subseteq \mathcal{V}_j^n$ of rate $K_j + L_j$ whose codewords are independently chosen according to $p_{V_j}^n$. Codewords of $\mathcal{C}_j$ are independently and uniformly partitioned into $\exp\{nL_j\}$ bins. The distribution induced on $\mathcal{C}_1, \Lambda_2, \Lambda_3$ is identical to that in section 5.7.1. Moreover, the triplet $\mathcal{C}_2, \mathcal{C}_3, (\mathcal{C}_1, \Lambda_2, \Lambda_3)$ are mutually independent.[20] Having specified the distribution of codewords of $\mathcal{C}_j : j = 2, 3$, we

[20] Here $(\mathcal{C}_1, \Lambda_2, \Lambda_3)$ is treated as a single random object.

have thus specified the distribution of quintuple of random codebooks. Messages of users' 2 and 3 are split into two parts each. One part of user 2's (3's) message, of rate $T_2 \log \pi$ ($T_3 \log \pi$), index a bin in Λ_2 (Λ_3), and the other part, of rate L_2 (L_3), index a bin in $\mathcal{C}_2$ ($\mathcal{C}_3$). The sole part of user 1's message indexes a bin in $\mathcal{C}_1$. The encoder looks for a quintuple of jointly typical codewords with respect to $p_{\underline{UV}}$, in the quintuple of indexed bins. Following a second moment method similar to that employed in [63], it can be proved that the encoder finds at least one jointly typical triple if

$$\begin{aligned} S_A \log \pi + K_B &> |A| \log_2 \pi + \sum_{b \in B} H(V_b) - H(U_A, V_B)^{21} && (5.29) \\ \max\{S_2 + T_2, S_3 + T_3\} \log \pi + K_B &> \log \pi + \sum_{b \in B} H(V_b) - \min_{\theta \in \mathcal{F}_\pi \setminus \{0\}} H(U_2 \oplus \theta U_3, V_B) && (5.30) \end{aligned}$$

for all $A \subseteq \{2,3\}, B \subseteq \{1,2,3\}$, where $S_A = \sum_{j \in A} S_j$, $K_B = \sum_{b \in B} K_b$, $U_A = (U_j : j \in A)$ and $V_B = (V_b : b \in B)$.[22] Having chosen one such jointly typical quintuple, say $(U_2^n, U_3^n, \underline{V}^n)$, the encoder generates a vector X^n according to $p^n_{X|\underline{V}U_2U_3}(\cdot|\underline{V}^n, U_2^n, U_3^n)$ and inputs the same on the channel.

The operations of decoders 2 and 3 are identical and we describe one of them. Decoder 3 receives Y_3^n and looks for all pairs of codewords in the Cartesian product $\Lambda_3 \times \mathcal{C}_3$ that are jointly typical with Y_3^n with respect to $p_{U_3V_3Y_3}$. If all such pairs belong to a unique pair of bins, the corresponding pair of bin indices is declared as the decoded message of user 3. Else an error is declared. It can be proved that if

$$\begin{aligned} (S_j + T_j) \log \pi < \log_2 \pi - H(U_j|V_j, Y_j), \qquad K_j + L_j < H(V_j) - H(V_j|Y_j, U_j) && (5.31) \\ (S_j + T_j) \log \pi + K_j + L_j &< \log_2 \pi + H(V_j) - H(V_j, U_j|Y_j) && (5.32) \end{aligned}$$

for $j = 2, 3$, then probability of users 2 or 3 decoding into an incorrect message falls exponentially with n.

Operation of decoder 1 is identical to that described in section 5.7.1. If (5.17) holds, then probability of error at decoder 1 falls exponentially with n. Substituting $R_1 = K_1, R_2 = T_2 \log \pi + L_2, R_3 = T_3 \log \pi + L_3$ and eliminating $S_2 \log \pi, S_3 \log \pi, K_1, K_2, K_3$ in (5.17)-(5.32) yields an achievable rate region. We provide a mathematical characterization of this achievable rate region.

Definition 5.7.4 *Let $\mathbb{D}_2^f(W_{\underline{Y}|X}, \kappa, \tau)$ denote the collection of pmfs $p_{QU_2U_3V_1V_2V_3X\underline{Y}}$ defined on $\mathcal{Q} \times \mathcal{U}_2 \times \mathcal{U}_3 \times \mathcal{V}_1 \times \mathcal{V}_2 \times \mathcal{V}_3 \times \mathcal{X} \times \underline{\mathcal{Y}}$, where (i) $\mathcal{U}_2 = \mathcal{U}_3 = \mathcal{F}_\pi$ is the finite field of cardinality π, $\mathcal{Q}, \mathcal{V}_1, \mathcal{V}_2, \mathcal{V}_3$ are finite sets, (ii) $p_{\underline{Y}|X\underline{V}\underline{U}Q} = p_{\underline{Y}|X} = W_{\underline{Y}|X}$, and (iii) $\mathbb{E}\{\kappa(X)\} \leq \tau$. For $p_{Q\underline{UV}X\underline{Y}} \in \mathbb{D}_2^f(W_{\underline{Y}|X}, \kappa, \tau)$, let $\beta_2^f(p_{Q\underline{UV}X\underline{Y}})$ be defined as the set of triples $(R_1, R_2, R_3) \in [0, \infty)^3$ for which there exists nonnegative numbers $S_2, T_2, S_3, T_3, K_j, L_j : j = 1, 2, 3$*

[21] We remind the reader that the empty sum has value 0, i.e, $\sum_{a \in \phi} = 0$

[22] Recall that $\mathcal{F}_\pi = \mathcal{U}_2 = \mathcal{U}_3$.

such that $R_1 = K_1, R_2 = T_2 \log \pi + L_2, R_3 = T_3 \log \pi + L_3$,

$$S_A \log \pi + K_B > |A| \log_2 \pi + \sum_{b \in B} H(V_b|Q) - H(U_A, V_B|Q),^{23}$$
$$\max\{S_2 + T_2, S_3 + T_3\} \log \pi + K_B > \log \pi + \sum_{b \in B} H(V_b|Q) - \min_{\theta \in \mathcal{F}_\pi \setminus \{0\}} H(U_2 \oplus \theta U_3, V_B|Q),$$
$$K_1 + R_1 < I(V_1; U_2 \oplus U_3, Y_1|Q), \quad K_1 + R_1 + (S_j + T_j) \log \pi < \log \pi + H(V_1|Q) - H(V_1, U_2 \oplus U_3|Q, Y_1) : j = 2, 3,$$
$$(S_j + T_j) \log \pi < \log_2 \pi - H(U_j|Q, V_j, Y_j) : j = 2, 3, \quad K_j + L_j < H(V_j|Q) - H(V_j|Q, Y_j, U_j) : j = 2, 3$$
$$(S_j + T_j) \log \pi + K_j + L_j < \log_2 \pi + H(V_j|Q) - H(V_j, U_j|Q, Y_j) : j = 2, 3$$

for all $A \subseteq \{2, 3\}, B \subseteq \{1, 2, 3\}$, *where* $S_A = \sum_{j \in A} S_j$, $K_B = \sum_{b \in B} K_b$, $U_A = (U_j : j \in A)$ *and* $V_B = (V_b : b \in B)$. *Let*

$$\beta_2^f(W_{\underline{Y}|X}, \kappa, \tau) = cocl \left(\bigcup_{\substack{p_{Q\underline{U}\underline{V}X\underline{Y}} \\ \in \mathbb{D}_2^f(W_{\underline{Y}|X}, \kappa, \tau)}} \beta_2^f(p_{Q\underline{U}\underline{V}X\underline{Y}}) \right).$$

Theorem 5.7.5 *For a* $3-DBC$ $(\mathcal{X}, \underline{\mathcal{Y}}, W_{\underline{Y}|X}, \kappa)$, $\beta_2^f(W_{\underline{Y}|X}, \kappa, \tau)$ *is achievable, i.e.,* $\beta_2^f(W_{\underline{Y}|X}, \kappa, \tau) \subseteq \mathbb{C}(W_{\underline{Y}|X}, \kappa, \tau)$. □

The proof is similar to that of theorem 5.7.3. The only differences being (i) the encoder looks for a quintuple of codewords instead of a triple, and (ii) decoders 2 and 3 decode from a pair of codebooks. From theorem 3.5.1, the informed reader can see why the second difference can be easily handled. Indeed, in theorem 3.5.1, we prove nested coset codes, and therefore partitioned coset codes, achieve capacity of arbitrary point-to-point channels. This indicates that for $j = 2, 3$, $\mathcal{U}_j-$codebook can be used to communicate at rate $I(U_j; Y_j)$ and the private layer $\mathcal{V}_j-$codebook can be used to communicate at rate $I(U_j : Y_j|U_j)$, thereby satisfying user j's rate. This leaves us to argue only the first difference pointed above. Using a second moment method similar to that employed in appendix J,[24] it can be shown that probability of encoder not finding a jointly typical quintuple decays exponentially if (5.29) holds.

5.7.3 Step III: Using PCC to manage interference over a $3-$DBC

Here we employ PCC to manage/decode interference seen by each receiver. In the sequel, we propose a simple extension of the technique presented in section 5.7.2 to enable each user decode a bivariate interference component.

Throughout the following discussion i, j, k denote distinct indices in $\{1, 2, 3\}$. Let $\mathcal{U}_{ji} = \mathcal{F}_{\pi_i}, \mathcal{U}_{jk} = \mathcal{F}_{\pi_k}$ be finite fields and $\mathcal{V}_j$ be an arbitrary finite set. User j splits it's message M_j into three parts $(M_{ji}^U, M_{jk}^U, M_j^V)$ of rates

[23] *We remind the reader that the empty sum has value* 0, *i.e,* $\sum_{a \in \phi} = 0$

[24] A diligent reader would have noted the same second moment method has been employed in appendices A and C.

$T_{ji}\log\pi_i, T_{jk}\log\pi_k, L_j$ respectively. User j's message indexes three codebooks - $\mathcal{C}_j, \Lambda_{ji}, \Lambda_{jk}$ - whose structure is described in the following. Consider a random codebook $\mathcal{C}_j \subseteq \mathcal{V}_j^n$ of rate $K_j + L_j$ whose codewords are independently chosen according to $p_{V_j}^n$. Codewords of $\mathcal{C}_j$ are independently and uniformly partitioned into $\exp\{nL_j\}$ bins. Consider random partitioned coset codes (PCC) $(n, nS_{ji}, nT_{ji}, G_{ji}, B_{ji}^n, I_{ji})$ and $(n, nS_{jk}, nT_{jk}, G_{jk}, B_{jk}^n, I_{jk})$ (definition 3.4.2) denoted Λ_{ji} and Λ_{jk} respectively. Observe that PCC Λ_{ji} and Λ_{ki} are built over the same finite field $\mathcal{F}_{\pi_i}$. The corresponding linear codes are nested, i.e., if $S_{ji} \leq S_{ki}$, then $G_{ki}^t = \begin{bmatrix} G_{ji}^t & G_{ki/ji}^t \end{bmatrix}$ where $G_{ki/ji} \in \mathcal{F}_\pi^{n(S_{ji}-S_{ki})\times n}$, and vice versa. We have thus specified the structure of 9 random codebooks. We now specify the distribution of these random codebooks.

The random PCCs are independent of $\mathcal{C}_j : j = 1, 2, 3$. $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3$ are mutually independent. We now specify the distribution of the PCCs. The triplet $(\Lambda_{12}, \Lambda_{32}), (\Lambda_{21}, \Lambda_{31}), (\Lambda_{23}, \Lambda_{13})$ are mutually independent. All of the bias vectors are mutually independent and uniformly distributed. The collection of generator matrices is independent of the collection of bias vectors. We only need to specify the distribution of the generator matrices. The rows of the larger of the two generator matrices G_{ji} and G_{ki} are uniformly and independently distributed. This specifies the distribution of the 9 random codebooks.

M_{ji}^U,M_{jk}^U and M_j^V index bins in Λ_{ji}, Λ_{jk} and $\mathcal{C}_j$ respectively. The encoder looks for a collection of 9 codewords from the indexed bins that are jointly typical with respect to a pmf $p_{\underline{UV}}$ defined on $\underline{\mathcal{U}} \times \underline{\mathcal{V}}$.[25] We now state the bounds that ensure the probability of encoder not finding a jointly typical collection of codewords from the indexed bins. We introduce some notation to aid reduce clutter. Throughout the following, in every instance i, j, k will denote distinct indices in $\{1, 2, 3\}$. For every $A \subseteq \{12, 13, 21, 23, 31, 32\}, B \subseteq \{1, 2, 3\}, C \subseteq \{1, 2, 3\}$, let $S_A = \sum_{jk\in A} S_{jk}, M_B := \sum_{j\in B} \max\{S_{ij} + T_{ij}, S_{kj} + T_{kj}\}, K_C = \sum_{c\in C} K_c$. For every $B \subseteq \{1, 2, 3\}$, let $A(B) = \cup_{j\in B}\{ji, jk\}$. Following a second moment method similar to that employed in appendix J, it can be proved that the encoder finds at least one jointly typical collection if (5.33) is satisfied for all $A \subseteq \{12, 13, 21, 23, 31, 32\}, B \subseteq \{1, 2, 3\}, C \subseteq \{1, 2, 3\}$, that satisfy $A \cap A(B) = \phi$, where $U_A = (U_{jk} : jk \in A)$ and $V_C = (V_c : c \in C)$. Having chosen one such jointly typical collection, say $(\underline{U}^n, \underline{V}^n)$, the encoder generates a vector X^n according to $p_{X|\underline{UV}}^n(\cdot|\underline{U}^n, \underline{V}^n)$ and feeds the same as input on the channel.

Decoder j receives Y_j^n and looks for all triples $(u_{ji}^n, u_{jk}^n, v_j^n)$ of codewords in $\lambda_{ji} \times \lambda_{jk} \times \mathcal{C}_j$ such that there exists a $u_{ij\oplus kj}^n \in (\lambda_{ij} \oplus \lambda_{kj})$ such that $(u_{ij\oplus kj}^n, u_{ji}^n, u_{jk}^n, v_j^n, Y_j^n)$ are jointly typical with respect to $p_{U_{ij}\oplus U_{kj}, U_{ji}, U_{jk}, V_j, Y_j}$. If it finds all such triples in a unique triple of bins, the corresponding triple of bin indices is declared as decoded message of user j. Else an error is declared. The probability of error at decoder j can be made arbitrarily small for sufficiently large block length if (5.34) holds for every $\mathcal{A}_j \subseteq \{ji, jk\}$ with distinct indices i, j, k in $\{1, 2, 3\}$, where $S_{\mathcal{A}_j} := \sum_{a\in\mathcal{A}_j} S_a, T_{\mathcal{A}_j} := \sum_{a\in\mathcal{A}_j} T_a, U_{\mathcal{A}_j} = (U_a : a \in \mathcal{A}_j)$. . Recognize that user j's rate $R_j = T_{ji}\log\pi_i + T_{jk}\log\pi_k + L_j$. We are now equipped to state an achievable rate region for a general 3−DBC using partitioned coset codes.

[25] $\underline{U}$ abbreviates $U_{12}U_{13}U_{21}U_{23}U_{31}U_{32}$.

Definition 5.7.6 *Let $\mathbb{D}^f(W_{\underline{Y}|X}, \kappa, \tau)$ denote the collection of probability mass functions $p_{Q\underline{UV}X\underline{Y}}$ defined on $\mathcal{Q}\times\underline{\mathcal{U}}\times\underline{\mathcal{V}}\times\mathcal{X}\times\underline{\mathcal{Y}}$, where (i) $\mathcal{Q}, \mathcal{V}_1, \mathcal{V}_2, \mathcal{V}_3$ are arbitrary finite sets, $\underline{\mathcal{V}} := \mathcal{V}_1\times\mathcal{V}_2\times\mathcal{V}_3$, (ii) $\mathcal{U}_{ij} = \mathcal{F}_{\pi_j}$[26] for each $1 \le i, j \le 3$, and $\underline{\mathcal{U}} := \mathcal{U}_{12}\times\mathcal{U}_{13}\times\mathcal{U}_{21}\times\mathcal{U}_{23}\times\mathcal{U}_{31}\times\mathcal{U}_{32}$, (iii) $\underline{V} := (V_1, V_2, V_3)$ and $\underline{U} := (U_{12}, U_{13}, U_{21}, U_{23}, U_{31}, U_{32})$, such that (i) $p_{\underline{Y}|X\underline{VU}} = p_{\underline{Y}|X} = W_{\underline{Y}|X}$, (ii) $\mathbb{E}\{\kappa(X)\} \le \tau$.*

For $p_{\underline{UV}X\underline{Y}} \in \mathbb{D}^f(W_{\underline{Y}|X}, \kappa, \tau)$, let $\beta^f(p_{\underline{UV}X\underline{Y}})$ be defined as the set of rate triples $(R_1, R_2, R_3) \in [0,\infty)^3$ for which there exists nonnegative numbers $S_{ij}, T_{ij} : ij \in \{12, 13, 21, 23, 31, 32\}, K_j, L_j : j \in \{1, 2, 3\}$ such that $R_1 = T_{12}\log\pi_2 + T_{13}\log\pi_3 + L_1, R_2 = T_{21}\log\pi_1 + T_{23}\log\pi_3 + L_2, R_3 = T_{31}\log\pi_1 + T_{32}\log\pi_2 + L_3$ and

$$\begin{aligned} &S_A + M_B + K_C > \Theta(A,B,C) \text{ where,} \\ &\Theta(A,B,C) := \max_{(\theta_j : j\in B)\in \prod_{j\in B}\mathcal{F}_{\pi_j}} \{\sum_{a\in A}\log|\mathcal{U}_a| + \sum_{j\in B}\log\pi_j + \sum_{c\in C} H(V_c|Q) - H(U_A, U_{ji}\oplus\theta_j U_{jk} : j\in B, V_C|Q)\} \end{aligned} \tag{5.33}$$

for all $A \subseteq \{12, 13, 21, 23, 31, 32\}, B \subseteq \{1, 2, 3\}, C \subseteq \{1, 2, 3\}$, that satisfy $A \cap A(B) = \phi$, where $A(B) = \cup_{j\in B}\{ji, jk\}$, $U_A = (U_{jk} : jk \in A)$, $V_C = (V_c : c \in C)$, $S_A = \sum_{jk\in A} S_{jk}, M_B := \sum_{j\in B}\max\{S_{ij} + T_{ij}, S_{kj} + T_{kj}\}, K_C = \sum_{c\in C} K_c$, and

$$\begin{aligned} &S_{\mathcal{A}_j} + T_{\mathcal{A}_j} \le \sum_{a\in\mathcal{A}_j}\log|\mathcal{U}_a| - H(U_{\mathcal{A}_j}|Q, U_{\mathcal{A}_j^c}, U_{ij}\oplus U_{kj}, V_j, Y_j) \\ &S_{\mathcal{A}_j} + T_{\mathcal{A}_j} + S_{ij} + T_{ij} \sum_{a\in\mathcal{A}_j}\log|\mathcal{U}_a| + \log\pi_j - H(U_{\mathcal{A}_j}, U_{ij}\oplus U_{kj}|Q, U_{\mathcal{A}_j^c}, V_j, Y_j) \\ &S_{\mathcal{A}_j} + T_{\mathcal{A}_j} + S_{kj} + T_{kj} \le \sum_{a\in\mathcal{A}_j}\log|\mathcal{U}_a| + \log\pi_j - H(U_{\mathcal{A}_j}, U_{ij}\oplus U_{kj}|Q, U_{\mathcal{A}_j^c}, V_j, Y_j) \\ &S_{\mathcal{A}_j} + T_{\mathcal{A}_j} + K_j + L_j \le \sum_{a\in\mathcal{A}_j}\log|\mathcal{U}_a| + H(V_j) - H(U_{\mathcal{A}_j}, V_j|Q, U_{\mathcal{A}_j^c}, U_{ij}\oplus U_{kj}, Y_j) \\ &S_{\mathcal{A}_j} + T_{\mathcal{A}_j} + K_j + L_j + S_{ij} + T_{ij} \le \sum_{a\in\mathcal{A}_j}\log|\mathcal{U}_a| + \log\pi_j + H(V_j) - H(U_{\mathcal{A}_j}, V_j, U_{ij}\oplus U_{kj}|Q, U_{\mathcal{A}_j^c}, Y_j) \\ &S_{\mathcal{A}_j} + T_{\mathcal{A}_j} + K_j + L_j + S_{kj} + T_{kj} \le \sum_{a\in\mathcal{A}_j}\log|\mathcal{U}_a| + \log\pi_j + H(V_j) - H(U_{\mathcal{A}_j}, V_j, U_{ij}\oplus U_{kj}|Q, U_{\mathcal{A}_j^c}, Y_j), \end{aligned} \tag{5.34}$$

for every $\mathcal{A}_j \subseteq \{ji, jk\}$ with distinct indices i, j, k in $\{1, 2, 3\}$, where $S_{\mathcal{A}_j} := \sum_{a\in\mathcal{A}_j} S_a, T_{\mathcal{A}_j} := \sum_{a\in\mathcal{A}_j} T_a, U_{\mathcal{A}_j} = (U_a : a \in \mathcal{A}_j)$. Let

$$\beta^f(W_{\underline{Y}|X}, \kappa, \tau) = cocl\left(\bigcup_{\substack{p_{Q\underline{UV}X\underline{Y}}\in \\ \mathbb{D}^f(W_{\underline{Y}|X},\kappa,\tau)}} \beta^f(p_{Q\underline{UV}X\underline{Y}})\right).$$

[26] *Recall $\mathcal{F}_{\pi_j}$ is the finite field of cardinality π_j.*

Theorem 5.7.7 *For* $3-DBC$ $(\mathcal{X}, \underline{\mathcal{Y}}, W_{\underline{Y}|X}, \kappa)$, $\beta^f(W_{\underline{Y}|X}, \kappa, \tau)$ *is achievable, i.e.,* $\beta^f(W_{\underline{Y}|X}, \kappa, \tau) \subseteq \mathbb{C}(W_{\underline{Y}|X}, \kappa, \tau)$.

□

All the non-trivial elements of this proof being illustrated in considerable detail in the context of proof of theorem 5.7.3, we omit a proof of theorem 5.7.7.

5.8 Enlarging Marton's rate region using partitioned coset codes

The natural question that arises is whether the achievable rate region using partitioned coset codes $\beta^f(W_{\underline{Y}|X}, \kappa, \tau)$ contains $\alpha_{\mathscr{U}}(W_{\underline{Y}|X}, \kappa, \tau)$. It is our belief that coding techniques based on structured codes do not substitute their counterparts based on traditional unstructured independent codes, but enhance the same. Indeed, the technique proposed by Körner and Marton [18] is strictly suboptimal to that studied by Berger and Tung [64] if the function is not sufficiently compressive, i.e., entropy of the sum is larger than one half of the joint entropy of the sources.[27] The penalty paid in terms of the binning rate for endowing structure is not sufficiently compensated for by the function. This was (recognized)/(hinted at) by Ahlswede and Han [48, Section VI] for the problem studied by Körner and Marton.

We follow the approach of Ahlswede and Han [48, Section VI] to enlarge $\alpha_{\mathscr{U}}(W_{\underline{Y}|X}, \kappa, \tau)$ by gluing together $\mathscr{U}\mathcal{M}-$technique and the coding technique based on PCC. The resulting rate region will contain $\alpha_{\mathscr{U}}(W_{\underline{Y}|X}, \kappa, \tau) \cup \beta^f(W_{\underline{Y}|X}, \kappa, \tau)$ and will strictly enlarge $\alpha_{\mathscr{U}}(W_{\underline{Y}|X}, \kappa, \tau)$. Indeed, a description of the resulting rate region is quite involved and we spare the reader of these details. The resulting coding technique will involve each user split it's message into six parts - one public and private part each, two semi-private and *bivariate* parts each. This can be understood by splitting the message as proposed in sections 5.5.2 and 5.7.3 and identifying the private parts. In essence each user decodes a univariate component of every other user's transmission particularly set apart for it, and furthermore decodes a bivariate component of the other two user's transmissions.[28]

5.9 Concluding remarks : Common parts of random variables and the need for structure

Let us revisit Marton's coding technique for 2-BC. Define the pair $\overline{V_j} := (W, V_j) : j = 1, 2$ of random variables decoded by the two users and let $\overline{\mathcal{V}_j} := \mathcal{W} \times \mathcal{V}_j : j = 1, 2$. Let us stack the collection of compatible codewords over $\overline{\mathcal{V}_1}^n \times \overline{\mathcal{V}_2}^n$. The encoder can work with this stack, being oblivious to the distinction between $\mathcal{W}$ and $\mathcal{V}_j : j = 1, 2$. In other words, it does not recognize that a symbol over $\overline{V_j}$ is indeed a pair of symbols. A few key observations of this

[27] If X and Y are the distributed binary sources whose modulo-2 sum is to be reconstructed at the decoder, then Körner and Marton technique is strictly suboptimal if $H(X \oplus Y) > \frac{H(X,Y)}{2}$.

[28] An informed and inquisitive reader may begin to see a relationship emerge between the several layers of coding and common parts of a collection of random variables. Please refer to section 6.5 for a discussion.

stack of codewords is in order. Recognize that many pairs of compatible codewords agree in their '$\mathcal{W}-$coordinate'. In other words, they share the same codeword on the $\mathcal{W}-$codebook. W is a common part [65] of the pair $(\overline{V_1}, \overline{V_2})$. Being a common part, it can be realized through univariate functions. Let us say $W = f_1(V_1) = f_2(V_2)$. This indicates, *$\mathcal{W}-$codebook is built such that, the range of these univariate functions when applied on the collection of codewords in this stack, is contained.*

How did Marton accomplish this containment? Marton proposed building the $W-$codebook first, followed by conditional codebooks over V_1, V_2. Conditional coding with a careful choice of order therefore contained the range under the action of univariate function. How is all of this related to the need for containing bivariate functions of a pair of random variables. The fundamental underlying thread is the notion of common part [65]. What are the common parts of a triple of random variables? Clearly, one can simply extend the notion of common part defined for a pair of random variables. This yields four common parts - one part that is simultaneously to common to all three random variables and one common part each, corresponding to each pair in the triple. Indeed, if $\overline{V_1} = (W, U_{12}, U_{31}, V_1), \overline{V_2} = (W, U_{12}, U_{23}, V_2), \overline{V_3} = (W, U_{23}, U_{31}, V_3)$, then W is the part simultaneously to common to $\overline{V_1}, \overline{V_2}, \overline{V_3}$ and $U_{ij} : ij \in \{12, 23, 31\}$ are the pairwise common parts. A simple extension of Marton's coding suggests a way to handle these common parts.

This does not yet answer the need for containment under bivariate function. We envision a fundamentally richer notion of common part for a triple of random variables. Indeed, three nontrivial binary random variables $X, Y, Z = X \oplus Y$ have no common parts as defined earlier, since each pair has no common part and the triple does not admit a simultaneous common part. Yet, the degeneracy in the joint probability matrix hints at a common part. Indeed, they possess a *conferencing* common part. For example, the pair $(X, Y), Z$ have a common part. In other words, there exists a *bivariate* function of X, Y and a univariate function of Z that agree with probability 1. Containment of this bivariate function brings in the need for structured codes. Indeed, the resemblance to the problem studied by Körner and Marton [18] is striking. We therefore believe the need for structured codes for three (multi) user communication problems is closely linked to the notion of common parts of a triple (collection) of random variables. Analogous to conditional coding that contained univariate functions, endowing codebooks with structure is an inherent need to carefully handle additional degrees of freedom prevalent in larger dimensions.

5.10 Strict sub-optimality of $\mathscr{U}\mathcal{M}-$technique

In this section, we prove strict sub-optimality of $\mathscr{U}\mathcal{M}-$technique for the 3−DBC presented in example 5.6.1. In particular, we prove that if parameters $\tau, \delta_1, \delta_2, \delta_3$ are such that $1 + h_b(\delta_1 * \tau) > h_b(\delta_2) + h_b(\delta_3)$ and $(R_1, 1 - h_b(\delta_2), 1 - h_b(\delta_3)) \in \alpha_{\mathscr{U}}(\tau)$, then $R_1 < h_b(\tau * \delta_1) - h_b(\delta_1)$.

Why is $\mathscr{U}\mathcal{M}-$technique suboptimal for the case described above. As mentioned in section 5.6, in this case, receiver 1 is unable to decode the pair of codewords transmitted to users 2 and 3. Furthermore, based on unstructured

independent coding, it does not attempt to decode a function of transmitted codewords - in this case the modulo-2 sum. This forces decoder 1 to be content by decoding only individual components of user 2 and 3's transmissions, leaving residual uncertainty in the interference. The encoder helps out by precoding for this residual uncertainty. However, as a consequence of the cost constraint on X_1, it is forced to live with a rate loss.

Since our proof traces through the above arguments in three stages, it is instructive. In the first stage, we characterize all test channels $p_{QW\underline{UV}X\underline{Y}}$ for which $(R_1, 1-h_b(\delta_2), 1-h_b(\delta_3)) \in \alpha_{\mathscr{U}}(p_{QW\underline{UV}X\underline{Y}})$. This stage enables us identify 'active' codebooks, their corresponding rates and characterize two upper bounds on R_1. One of these contains the rate loss due to precoding. In the second stage, we therefore characterize the condition under which there is no rate loss. As expected, it turns out that there is no rate loss only if decoder 1 has decoded codewords of users 2 and 3. This gets us to the third stage, where we conclude that $1 + h_b(\delta_1 * \tau) > h_b(\delta_2) + h_b(\delta_3)$ precludes this possibility. The first stage is presented in lemma 5.10.1, second stage is stated in lemma G.0.13 and proved in appendices G, H. Third stage can be found in arguments following lemma G.0.13.

We begin with a characterization of a test channel $p_{QW\underline{UV}X\underline{Y}}$ for which $(R_1, 1-h_b(\delta_2), 1-h_b(\delta_3)) \in \alpha_{\mathscr{U}}(p_{QW\underline{UV}X\underline{Y}})$. Since independent information needs to be communicated to users 2 and 3 at their respective point to point capacities, it is expected that their codebooks are not precoded for each other's signal, and moreover none of users 2 and 3 decode a part of the other users' signal. The following lemma establishes this. We remind the reader that $X_1X_2X_3 = X$ denote the three binary digits at the input, where Y_2, the output at receiver 2 is obtained by passing X_2 through a BSC with cross over probability δ_2, Y_3, the output at receiver 3 is obtained by passing X_3 through a BSC with cross over probability δ_3 and Y_1 is obtained by passing $X_1 \oplus X_2 \oplus X_3$ through a BSC with cross over probability δ_1. Moreover, the binary symmetric channels (BSC's) are independent. Input symbol X_1 is constrained with respect to a Hamming cost function and the constraint on the average cost per symbol is τ. Formally, $\kappa(x_1x_2x_3) = 1_{\{x_1=1\}}$ is the cost function and the average cost per symbol is not to exceed τ.

Lemma 5.10.1 *If there exists a test channel $p_{QW\underline{UV}X\underline{Y}} \in \mathbb{D}_{\mathscr{U}}(\tau)$ and nonnegative numbers $K_i, S_{ij}, K_{ij}, L_{ij}, S_i, T_i$ that satisfy (5.1)-(5.11) for each triple $(i,j,k) \in \{(1,2,3),(2,3,1),(3,1,2)\}$ such that $R_2 = K_2 + K_{23} + L_{12} + T_2 = 1 - h_b(\delta_2), R_3 = K_3 + K_{31} + L_{23} + T_3 = 1 - h_b(\delta_3)$, then*

(i) $K_1 = K_2 = K_3 = K_{23} = L_{23} = K_{12} = L_{31} = S_2 = S_3 = 0$ *and* $I(U_{31}V_1V_3; Y_2|QWU_{23}U_{12}V_2) = 0$,

(ii) $S_{31} = I(U_{31}; U_{23}|QW), S_{12} = I(U_{12}; U_{23}|QW)$, $S_{23} = I(U_{12}; U_{31}|QWU_{23}) = 0$,

(iii) $I(V_2U_{12}; V_3U_{31}|QWU_{23}) = 0$, $I(WU_{23}; Y_j|Q) = 0 : j = 2,3, I(V_2U_{12}; Y_2|QWU_{23}) = 1 - h_b(\delta_2)$ *and* $I(V_3U_{31}; Y_3|QWU_{23}) = 1 - h_b(\delta_3)$,

(iv) $(V_3, X_3, V_1, U_{31}) - (QWU_{23}U_{12}V_2) - (X_2, Y_2)$ *and* $(V_2, X_2, V_1, U_{12}) - (QWU_{23}U_{31}V_3) - (X_3, Y_3)$ *are Markov chains,*

(v) $X_2 - QWU_{12}U_{23}U_{31} - X_3$ *is a Markov chain,*

(vi) $U_{12} - QWU_{23}U_{31} - X_3$ *and* $U_{31} - QWU_{23}U_{12} - X_2$ *are Markov chains.*

□

Proof: Substituting (i) $(2,3,1)$ for (i,j,k) in (5.11), (ii) $(1,2,3)$ for (i,j,k) in (5.2) and combining the resulting bounds yields

$$I(WU_{23}U_{12}V_2;Y_2|Q) \geq R_2 + K_3 + K_1 + L_{23} + K_{12} + S_2 \geq R_2 = 1 - h_b(\delta_2), \tag{5.35}$$

where the second inequality follows from non-negativity of $K_3, K_1, L_{23}, K_{12}, S_2$. Moreover,

$$\begin{aligned} 1 - h_b(\delta_2) &\geq I(X_2;Y_2) = I(QW\underline{UV}X_1Y_1X_3Y_3X_2;Y_2) \geq I(WU_{23}U_{12}V_2;Y_2|Q) &(5.36)\\ &\geq R_2+K_3+K_1+L_{23}+K_{12}+S_2 \geq R_2 = 1-h_b(\delta_2), &(5.37)\end{aligned}$$

where (i) equality in (5.36) follows from Markov chain $QW\underline{UV}X_1Y_1X_3Y_3 - X_2 - Y_2$, and (ii) (5.37) follows from substituting (5.35). Since all the terms involved are non-negative, equality holds through the above chain of inequalities to yield

$$S_{12} + S_{23} = I(U_{12};U_{23}|QW), K_1 = K_3 = L_{23} = K_{12} = S_2 = I(Q;Y_2) = 0 \tag{5.38}$$

$$I(U_{31}V_1X_1Y_1V_3X_3Y_3X_2;Y_2|QWU_{12}U_{23}V_2) = 0 \tag{5.39}$$

$$\text{and therefore } (V_1,V_3,X_3,U_{31}) - (QWU_{12}U_{23}V_2) - Y_2 \text{ is a Markov chain} \tag{5.40}$$

where the first equality in (5.38) follows from condition for equality in the first inequality of (5.35). The above sequence of steps are repeated by substituting (i) $(3,1,2)$ for (i,j,k) in (5.11), (ii) $(2,3,1)$ for (i,j,k) in (5.2). It can be verified that

$$S_{31} + S_{23} = I(U_{31};U_{23}|QW), K_1 = K_2 = L_{31} = K_{23} = S_3 = I(Q;Y_3) = 0, \tag{5.41}$$

$$I(U_{12}V_1X_1Y_1V_2X_2Y_2X_3;Y_3|QWU_{23}U_{31}V_3) = 0 \tag{5.42}$$

$$\text{and therefore } (V_1,V_2,X_2,U_{12}) - (QWU_{23}U_{31}V_3) - Y_3 \text{ is a Markov chain.} \tag{5.43}$$

The second set of equalities in (5.38), (5.41) lets us conclude

$$R_1 = T_1, R_2 = L_{12} + T_2 \text{ and } R_3 = K_{31} + T_3. \tag{5.44}$$

From $I(U_{12};U_{23}|QW)+I(U_{31};U_{23}|QW) = S_{12}+S_{23}+S_{31}+S_{23}$, and (5.3), we have $I(U_{12};U_{23}|QW)+I(U_{31};U_{23}|QW) \geq I(U_{12};U_{23};U_{31}|QW)+S_{23}$. The non-negativity of S_{23} (5.1) implies $S_{23} = 0$ and $I(U_{31};U_{12}|QWU_{23}) = 0$. We therefore

conclude

$$S_{12} = I(U_{12}; U_{23}|QW), S_{31} = I(U_{31}; U_{23}|QW), S_{23} = 0, I(U_{31}; U_{12}|QWU_{23}) = 0 \tag{5.45}$$

Substituting (5.38), (5.41), (5.45) in (5.4) for $(i, j, k) = (2, 3, 1)$ and $(i, j, k) = (3, 1, 2)$ and (5.5) for $(i, j, k) = (2, 3, 1)$, we obtain

$$I(V_2; U_{31}|QWU_{12}U_{23}) = I(V_3; U_{12}|QWU_{23}U_{31}) = I(V_2; V_3|QWU_{12}U_{23}U_{31}) = 0. \tag{5.46}$$

(5.46) and last equality in (5.45) yield

$$I(V_2U_{12}; V_3U_{31}|QWU_{23}) = 0. \tag{5.47}$$

Substituting (5.44), (5.45) in (5.8) with $(i, j, k) = (2, 3, 1)$ yields the upper bound $R_2 \leq I(U_{12}V_2; Y_2|QWU_{23})$. Since

$$1 - h_b(\delta_2) = R_2 \leq I(U_{12}V_2; Y_2|QWU_{23}) \leq I(WU_{12}U_{23}V_2; Y_2|Q) \leq 1 - h_b(\delta_2),$$

where the last inequality follows from (5.36), equality holds in all of the above inequalities to yield $I(WU_{23}; Y_2|Q) = 0$ and $I(U_{12}V_2; Y_2|QWU_{23}) = 1 - h_b(\delta_2)$. A similar argument proves $I(WU_{23}; Y_3|Q) = 0$ and $I(U_{31}V_3; Y_3|QWU_{23}) = 1 - h_b(\delta_3)$.

We have proved the Markov chains in (5.40), (5.43). In order to prove Markov chains in item 4, we prove the following lemma.

Lemma 5.10.2 *If A, B, X, Y are discrete random variables such that (i) X, Y take values in $\{0, 1\}$ with $P(Y = 0|X = 1) = P(Y = 1|X = 0) = \eta \in (0, \frac{1}{2})$, (ii) $A - B - Y$ and $AB - X - Y$ are Markov chains, then $A - B - XY$ is also a Markov chain.* □

Please refer to appendix I for a proof. Markov chains in (5.40), (5.43) in conjunction with lemma 5.10.2 establishes Markov chains in item 4.

(5.47) and (5.39) imply $I(U_{31}V_3; U_{12}V_2Y_2|QWU_{23}) = 0$. This in conjunction with (5.42) implies

$$I(U_{31}V_3Y_3; U_{12}V_2Y_2|QWU_{23}) = 0 \text{ and thus } U_{31}V_3Y_3 - QWU_{23} - U_{12}V_2Y_2 \text{ is a Markov chain.} \tag{5.48}$$

(5.48) implies $U_{31}Y_3 - QWU_{23} - U_{12}Y_2$ is a Markov chain, and therefore $Y_3 - QWU_{12}U_{23}U_{31} - Y_2$ is a Markov chain. Employing lemma 5.10.2 twice we observe $Y_3X_3 - WU_{12}U_{23}U_{31} - X_2Y_2$ is a Markov chain and furthermore $X_3 - QWU_{12}U_{23}U_{31} - X_2$ is a Markov chain, thus proving item 5.

Finally, we prove Markov chains in item 6. From Markov chain $(V_3, X_3, V_1, U_{31}) - (QWU_{23}U_{12}V_2) - (X_2, Y_2)$ proved in item 4, we have $I(X_2; U_{31}|QWU_{23}U_{12}V_2) = 0$. From (5.47), we have $I(V_2; U_{31}|QWU_{23}U_{12}) = 0$. Summing these two, we have $I(X_2V_2; U_{31}|QWU_{23}U_{12}) = 0$ and therefore $I(X_2; U_{31}|QWU_{23}U_{12}) = 0$ implying the Markov

chain $X_2 - QWU_{23}U_{12} - U_{31}$. Similarly, Markov chain $(V_2, X_2, V_1, U_{12}) - (QWU_{23}U_{31}V_3) - (X_3, Y_3)$ proved in item 4 implies $I(X_3; U_{12}|WU_{23}U_{31}V_3Q) = 0$. From (5.47), we have $I(V_3; U_{12}|QWU_{23}U_{31}) = 0$. Summing these two, we have $I(X_3V_3; U_{12}|QWU_{23}U_{31}) = 0$ and therefore $I(X_3; U_{12}|QWU_{23}U_{31}) = 0$ implying the Markov chain $X_3 - QWU_{23}U_{31} - U_{12}$. ■

Lemma 5.10.1 enables us simplify the bounds (5.1)-(5.11) for the particular test channel under consideration. Substituting (5.38)-(5.46) in (5.1)-(5.11) and employing statements of lemma 5.10.1, we conclude that if $(R_1, 1-h_b(\delta_2), 1-h_b(\delta_3)) \in \alpha_{\mathscr{U}}(p_{QW\underline{UV}X\underline{Y}})$, then there exists nonnegative numbers S_1, T_1, L_{12}, K_{31} that satisfy $R_1 = T_1, R_2 = L_{12} + T_2 = 1 - h_b(\delta_2), R_3 = K_{31} + T_3 = 1 - h_b(\delta_3)$,

$$S_1 \geq I(V_1; U_{23}V_2V_3|QWU_{12}U_{31}), \quad T_1 + S_1 \leq I(V_1; Y_1|QWU_{12}U_{31}) \tag{5.49}$$

$$L_{12} + K_{31} + T_1 + S_1 \leq I(U_{12}; U_{31}|QW) - I(U_{23}; U_{12}|QW) + I(V_1U_{12}U_{31}; Y_1|QW) - I(U_{23}; U_{31}|QW) \tag{5.50}$$

$$0 \leq T_2 \leq I(V_2; Y_2|QWU_{12}U_{23}), \quad 1 - h_b(\delta_2) = T_2 + L_{12} = I(U_{12}V_2; Y_2|QWU_{23}) \tag{5.51}$$

$$0 \leq T_3 \leq I(V_3; Y_3|QWU_{31}U_{23}), \quad 1 - h_b(\delta_3) = T_3 + K_{31} = I(U_{31}V_3; Y_3|QWU_{23}). \tag{5.52}$$

(5.51), (5.52) imply

$$L_{12} \geq I(U_{12}; Y_2|QWU_{23}), \qquad K_{31} \geq I(U_{31}; Y_3|QWU_{23}), \tag{5.53}$$

(5.49) implies

$$\begin{aligned} T_1 = R_1 &\leq I(V_1; Y_1|QWU_{12}U_{31}) - I(V_1; U_{23}V_2V_3|QWU_{12}U_{31}), \\ &\leq I(V_1; Y_1U_{23}|QWU_{12}U_{31}) - I(V_1; U_{23}V_2V_3|QWU_{12}U_{31}) = I(V_1; Y_1|QW\underline{U}) - I(V_1; V_2V_3|QW\underline{U}), \end{aligned} \tag{5.54}$$

and (5.50) in conjunction with (5.53), and lower bound on S_1 in (5.49) imply

$$\begin{aligned} R_1 &\leq I(U_{12}U_{31}V_1; Y_1|QW) - I(V_1; U_{23}V_2V_3|QWU_{12}U_{31}) - I(U_{12}; Y_2|QWU_{23}) - I(U_{31}; Y_3|QWU_{23}) \\ &\quad + I(U_{12}; U_{31}|QW) - I(U_{23}; U_{12}|QW) - I(U_{23}; U_{31}|QW) \\ &\leq I(U_{12}U_{31}V_1; Y_1U_{23}|QW) - I(V_1; U_{23}V_2V_3|QWU_{12}U_{31}) - I(U_{12}; Y_2|QWU_{23}) - I(U_{31}; Y_3|QWU_{23}) \\ &\quad + I(U_{12}; U_{31}|QW) - I(U_{23}; U_{12}|QW) - I(U_{23}; U_{31}|QW) \\ &= I(V_1; Y_1|QW\underline{U}) - I(V_1; V_2V_3|QW\underline{U}) + I(U_{12}U_{31}; Y_1|QWU_{23}) - I(U_{12}; Y_2|QWU_{23}) - I(U_{31}; Y_3|QWU_{23}) \end{aligned} \tag{5.55}$$

where (5.55) follows from the last equality in (5.45). We have thus obtained (5.54) and (5.55), two upper bounds on R_1 we were seeking, and this concludes the first stage of our proof. In the sequel, we prove the minimum of the above upper bounds on R_1 is strictly lesser than $h_b(\tau * \delta_1) - h_b(\delta_1)$. Towards, that end, note that upper bound (5.54) contains the rate loss due to precoding. In the second stage, we work on (5.54) and derive conditions under which

there is *no* rate loss.

Markov chains of lemma 5.10.1 item 4 imply $V_1 - QW\underline{U}V_2V_3 - X_2$ and $V_1 - QW\underline{U}V_2V_3X_2 - X_3$ are Markov chains. Therefore, $I(V_1; X_2|QW\underline{U}V_2V_3) = 0$ and $I(V_1; X_3|QW\underline{U}V_2V_3X_2) = 0$. Summing these, we have $I(V_1; X_2X_3|QW\underline{U}V_2V_3) = 0$. Employing this in (5.54), we note

$$
\begin{aligned}
R_1 &\leq I(V_1;Y_1|QW\underline{U}) - I(V_1;V_2V_3|QW\underline{U}) = I(V_1;Y_1|QW\underline{U}) - I(V_1;V_2V_3X_2X_3|QW\underline{U}) && (5.56)\\
&\leq I(V_1;Y_1|QW\underline{U}) - I(V_1;X_2,X_3|QW\underline{U}) \leq I(V_1;Y_1|QW\underline{U}) - I(V_1;X_2\oplus X_3|QW\underline{U}) && (5.57)\\
&= \sum_{\substack{(q,w,\underline{u})\in\\ \mathcal{Q}\times\mathcal{W}\times\underline{U}}} p_{QW\underline{U}}(q,w,\underline{u})\left[I(V_1;Y_1|(Q,W,\underline{U}) = (q,w,\underline{u})) - I(V_1;X_2\oplus X_3|(Q,W,\underline{U}) = (q,w,\underline{u}))\right] && (5.58)\\
&\leq \sum_{\substack{(q,w,\underline{u})\in\\ \mathcal{Q}\times\mathcal{W}\times\underline{U}}} p_{QW\underline{U}}(q,w,\underline{u}) I(X_1X_2X_3V_1;Y_1|(Q,W,\underline{U}) = (q,w,\underline{u}))\\
&\leq \sum_{\substack{(q,w,\underline{u})\in\\ \mathcal{Q}\times\mathcal{W}\times\underline{U}}} p_{QW\underline{U}}(q,w,\underline{u})\left[H(Y_1|(Q,W,\underline{U}) = (q,w,\underline{u})) - H(Y_1|X_1X_2X_3V_1,(Q,W,\underline{U}) = (q,w,\underline{u}))\right]\\
&= \sum_{\substack{(q,w,\underline{u})\in\\ \mathcal{Q}\times\mathcal{W}\times\underline{U}}} p_{QW\underline{U}}(q,w,\underline{u})\left[H(X_1\oplus N_1|(Q,W,\underline{U}) = (q,w,\underline{u})) - h_b(\delta_1)\right]\\
&= \sum_{\substack{(q,w,\underline{u})\in\\ \mathcal{Q}\times\mathcal{W}\times\underline{U}}} p_{QW\underline{U}}(q,w,\underline{u}) h_b(\tau_{q,w,\underline{u}} * \delta_1) - h_b(\delta_1),\ \text{where } \tau_{q,w,\underline{u}} = p_{X_1|QW\underline{U}}(1|q,w,\underline{u}) && (5.59)\\
&= \mathbb{E}_{QW\underline{U}}\left\{h_b(\tau_{q,w,\underline{u}} * \delta_1)\right\} - h_b(\delta_1) \leq h_b(\mathbb{E}_{QW\underline{U}}\left\{\tau_{q,w,\underline{u}} * \delta_1\right\}) - h_b(\delta_1) \leq h_b(\tau * \delta_1) - h_b(\delta_1) && (5.60)
\end{aligned}
$$

where (5.60) follows from application of Jensen's inequality to the strictly concave function $h_b(\cdot)$, and second inequality in (5.60) follows from $\delta \in (0, \frac{1}{2})$. We conclude that $R_1 = h_b(\tau * \delta_1) - h_b(\delta_1)$ if and only if equality holds in the above chain of inequalities, and in particular, equality holds in (5.60), which by the condition for equality in Jensen's inequality implies $\tau_{q,w,\underline{u}} = \tau$ for every $(q,w,\underline{u}) \in \mathcal{Q}\times\mathcal{W}\times\underline{U}$ that satisfies $p_{QW\underline{U}}(q,w,\underline{u}) > 0$. This in conjunction with

$$I(V_1;Y_1|(Q,W,\underline{U}) = (q,w,\underline{u})) - I(V_1;X_2\oplus X_3|(Q,W,\underline{U}) = (q,w,\underline{u})) \leq h_b(\tau_{q,w,\underline{u}} * \delta_1) - h_b(\delta_1)$$

which follows from the chain of inequalities from (5.58) through (5.59) implies

$$I(V_1;Y_1|QW\underline{U}) - I(V_1;V_2V_3|QW\underline{U}) \leq h_b(\tau * \delta_1) - h_b(\delta_1) \tag{5.61}$$

with equality if and only if

$$\text{for every } (q,w,\underline{u}) \in \mathcal{Q}\times\mathcal{W}\times\underline{U} \text{ that satisfies } p_{QW\underline{U}}(q,w,\underline{u}) > 0, p_{X_1|QW\underline{U}}(1|q,w,\underline{u}) = \tau_{q,w,\underline{u}} = \tau, \tag{5.62}$$

$$\text{and } I(V_1;Y_1|(Q,W,\underline{U}) = (q,w,\underline{u})) - I(V_1;X_2\oplus X_3|(Q,W,\underline{U}) = (q,w,\underline{u})) = h_b(\tau_{q,w,\underline{u}} * \delta_1) - h_b(\delta_1). \tag{5.63}$$

An informed reader, by now must have made the connection to capacity of the point to point channel with non-causal state [7]. We develop this connection in appendix G. For now, we provide a characterization for (5.63) to hold. This will require us to define a few mathematical objects that may initially seem unrelated to a reader unaware of findings in [7]. Very soon, we argue the relevance. An informed reader will find the following development natural.

Let $\mathbb{D}_T(\tau,\delta,\epsilon)$ denote the collection of all probability mass functions $p_{\tilde{V}\tilde{S}\tilde{X}\tilde{Y}}$ defined on $\tilde{\mathcal{V}}\times\{0,1\}\times\{0,1\}\times\{0,1\}$, where $\tilde{\mathcal{V}}$ is an arbitrary finite set such that (i) $p_{\tilde{Y}|\tilde{X}\tilde{S}\tilde{V}}(x\oplus s|x,s,v)=p_{\tilde{Y}|\tilde{X}\tilde{S}}(x\oplus s|x,s)=1-\delta$, where $\delta\in(0,\frac{1}{2})$, (ii) $p_{\tilde{S}}(1)=\epsilon\in[0,1]$, and (iii) $p_{\tilde{X}}(1)\leq\tau\in(0,\frac{1}{2})$. For $p_{\tilde{V}\tilde{S}\tilde{X}\tilde{Y}}\in\mathbb{D}_T(\tau,\delta,\epsilon)$, let $\alpha_T(p_{\tilde{V}\tilde{S}\tilde{X}\tilde{Y}})=I(\tilde{V};\tilde{Y})-I(\tilde{V};\tilde{S})$ and $\alpha_T(\tau,\delta,\epsilon)=\sup_{p_{\tilde{V}\tilde{S}\tilde{X}\tilde{Y}}\in\mathbb{D}_T(\tau,\delta,\epsilon)}\alpha_T(p_{\tilde{V}\tilde{S}\tilde{X}\tilde{Y}})$.

For every $(q,w,\underline{u})\in\mathcal{Q}\times\mathcal{W}\times\underline{U}$ that satisfies $p_{QW\underline{U}}(q,w,\underline{u})>0$, we note $p_{Y_1|X_1,X_2\oplus X_3V_1QW\underline{U}}(x_1\oplus x_2\oplus x_3|x_1,x_2\oplus x_3,v_1,q,w,\underline{u})=p_{Y_1|X_1,X_2\oplus X_3QW\underline{U}}(x_1\oplus x_2\oplus x_3|x_1,x_2\oplus x_3,q,w,\underline{u})=1-\delta_1$. In other words, conditioned on the event $\{(Q,W,\underline{U})=(q,w,\underline{u})\}$, $V_1-X_1,X_2\oplus X_3-Y_1$ is a Markov chain. We conclude $p_{V_1X_2\oplus X_3X_1Y_1|QW\underline{U}}(\cdots|q,w,\underline{u})\in\mathbb{D}_T(\tau_{q,w,\underline{u}},\delta_1,\epsilon_{q,w,\underline{u}})$, where $\epsilon_{q,w,\underline{u}}=p_{X_2\oplus X_3|QW\underline{U}}(1|q,w,\underline{u})$, and hence

$$I(V_1;Y_1|(Q,W,\underline{U})=(q,w,\underline{u}))-I(V_1;X_2\oplus X_3|(Q,W,\underline{U})=(q,w,\underline{u}))\leq\alpha_T(\tau_{q,w,\underline{u}},\delta_1,\epsilon_{q,w,\underline{u}}).$$

Therefore, (5.63) holds only if $\alpha_T(\tau_{q,w,\underline{u}},\delta_1,\epsilon_{q,w,\underline{u}})=h_b(\tau_{q,w,\underline{u}}*\delta_1)-h_b(\delta_1)$, where $\tau_{q,w,\underline{u}}=\tau\in(0,\frac{1}{2})$. The following lemma characterizes conditions under which this is the case. Please refer to appendices G,H for a proof.

Lemma 5.10.3 *If $\tau,\delta\in(0,\frac{1}{2})$ and $\epsilon\in(0,1)$, then $\alpha_T(\tau,\delta,\epsilon)<h_b(\tau*\delta)-h_b(\delta)$. Alternatively, if $\tau,\delta\in(0,\frac{1}{2})$ and $\epsilon\in[0,1]$, then either $\alpha_T(\tau,\delta,\epsilon)<h_b(\tau*\delta)-h_b(\delta)$ or $\epsilon\in\{0,1\}$.* □

Recall that arguments in relation to (5.63) imply that if for any $(q,w,\underline{u})\in\mathcal{Q}\times\mathcal{W}\times\underline{\mathcal{U}}$ that satisfies $P((Q,W,\underline{U})=(q,w,\underline{u}))>0$, $I(V_1;Y_1|(Q,W,\underline{U})=(q,w,\underline{u}))-I(V_1;X_2\oplus X_3|(Q,W,\underline{U})=(q,w,\underline{u}))<h_b(\tau_{q,w,\underline{u}}*\delta_1)-h_b(\delta_1)$ where $\tau_{q,w,\underline{u}}=p_{X_1|Q,W,\underline{U}}(1|q,w,\underline{u})$, then $R_1<h_b(\tau*\delta_1)-h_b(\delta_1)$ and we have proved strict sub-optimality of $\mathscr{UM}-$technique. We therefore assume (5.62), (5.63) hold for every $(q,w,\underline{u})\in\mathcal{Q}\times\mathcal{W}\times\underline{\mathcal{U}}$ that satisfies $P((Q,W,\underline{U})=(q,w,\underline{u}))>0$. From lemma 5.10.3, we conclude for every such $(q,w,\underline{u})\in\mathcal{Q}\times\mathcal{W}\times\underline{\mathcal{U}}$, $\epsilon_{q,w,\underline{u}}=p_{X_2\oplus X_3|QW\underline{U}}(1|q,w,\underline{u})\in\{0,1\}$. We therefore assume

$$I(V_1;Y_1|QW\underline{U})-I(V_1;X_2\oplus X_3|QW\underline{U})=h_b(\tau*\delta_1)-h_b(\delta_1)\text{ and }H(X_2\oplus X_3|QW\underline{U})=0.\tag{5.64}$$

This has got us to the third and final stage. Here we argue (5.64) implies RHS of (5.55) is strictly smaller than $h_b(\tau*\delta_1)-h_b(\delta_1)$. Towards that end, note that Markov chain $X_2-QWU_{23}U_{12}U_{31}-X_3$ proved in lemma 5.10.1 item 5 and (5.64) imply $H(X_2|QW\underline{U})=H(X_3|QW\underline{U})=0$.[29] Furthermore, Markov chains $U_{12}-WU_{23}U_{31}-X_3$

[29]Indeed, for any $(q,w,\underline{u})\in\mathcal{Q}\times\mathcal{W}\times\underline{\mathcal{U}}$ that satisfies $P((Q,W,\underline{U})=(q,w,\underline{u}))>0$, if $P(X_j=1|(Q,W,\underline{U})=(q,w,\underline{u}))=\alpha_j:j=2,3$, then $0=H(X_2\oplus X_3|(Q,W,\underline{U})=(q,w,\underline{u}))=h_b(\alpha_2*\alpha_3)\geq\alpha_2h_b(\alpha_3)+(1-\alpha_2)h_b(1-\alpha_3)=\alpha_2h_b(\alpha_3)+(1-\alpha_2)h_b(\alpha_3)=h_b(\alpha_3)\geq0$, where the first inequality follows from concavity of binary entropy function, and similarly, interchanging the roles of α_2,α_3, we obtain $0=H(X_2\oplus X_3|(Q,W,\underline{U})=(q,w,\underline{u}))\geq h_b(\alpha_2)\ geq0$.

and $U_{31} - WU_{23}U_{12} - X_2$ proved in lemma 5.10.1 item 6 imply

$$H(X_2|WU_{23}U_{12}) = H(X_3|WU_{23}U_{31}) = 0. \tag{5.65}$$

Observe that

$$\begin{aligned} h_b(\tau * \delta_1) - h_b(\delta_1) &= I(V_1; Y_1|W\underline{U}) - I(V_1; X_2 \oplus X_3 W\underline{U}) = I(V_1; Y_1|W\underline{U}) = I(V_1; Y_1|W\underline{U}, X_2, X_3) && (5.66)\\ &= H(Y_1|W\underline{U}X_2X_3) - H(Y_1|W\underline{U}V_1X_2X_3) \leq H(Y_1|W\underline{U}X_2X_3) - H(Y_1|W\underline{U}V_1X_1X_2X_3) \\ &= H(Y_1|W\underline{U}, X_2, X_3) - h_b(\delta_1) && (5.67)\end{aligned}$$

where the first two equalities in (5.66) follows from (5.64) and the last equality follows from (5.65). (5.67) and first equality in (5.66) enables us conclude

$$H(Y_1|W\underline{U}, X_2, X_3) \geq h_b(\tau * \delta_1) \tag{5.68}$$

We now upper bound RHS of (5.55). Note that it suffices to prove $I(U_{12}U_{31}; Y_1|WU_{23}) - I(U_{12}; Y_2|WU_{23}) - I(U_{31}; Y_3|WU_{23})$ is negative. Observe that

$$\begin{aligned} &I(U_{12}U_{31}; Y_1|WU_{23}) - I(U_{12}; Y_2|WU_{23}) - I(U_{31}; Y_3|WU_{23}) \\ &\quad = H(Y_1|WU_{23}) - H(Y_1|W\underline{U}) - H(Y_2|WU_{23}) + H(Y_2|WU_{23}U_{12}) - H(Y_3|WU_{23}) + H(Y_3|WU_{23}U_{31}) \\ &\quad = H(Y_1|WU_{23}) - H(Y_1|W\underline{U}) - H(Y_2) + H(Y_2|WU_{23}U_{12}) - H(Y_3) + H(Y_3|WU_{23}U_{31}) \\ &\quad = H(Y_1|WU_{23}) - H(Y_1|WX_2X_3\underline{U}) - H(Y_2) + H(Y_2|WU_{23}U_{12}X_2) - H(Y_3) + H(Y_3|WU_{23}U_{31}X_3) && (5.69)\\ &\quad = H(Y_1|WU_{23}) - H(Y_1|WX_2X_3\underline{U}) - 2 + h_b(\delta_2) + h_b(\delta_3) \\ &\quad \leq 1 - H(Y_1|WX_2X_3\underline{U}) - 2 + h_b(\delta_2) + h_b(\delta_3) \leq h_b(\delta_2) + h_b(\delta_3) - h_b(\delta_1 * \tau) - 1 && (5.70)\end{aligned}$$

where (5.69) follows from (5.64) and (5.65), second inequality in (5.70) follows from (5.68). If $\tau, \delta_1, \delta_2, \delta_3$ are such that $h_b(\delta_2) + h_b(\delta_3) < 1 + h_b(\delta_1 * \tau)$, then $R_1 < h_b(\tau * \delta_1) - h_b(\delta_1)$ and RHS of (5.70) is negative. We summarize our findings in the following theorem and corollary.

Theorem 5.10.4 *Consider the 3−DBC in example 5.6.1. If $h_b(\delta_2) + h_b(\delta_3) < 1 + h_b(\delta_1 * \tau)$, then $(h_b(\tau * \delta_1) - h_b(\delta_1), 1 - h_b(\delta_2), 1 - h_b(\delta_3)) \notin \alpha_{\mathscr{U}}(\tau)$.* □

Corollary 5.10.5 *Consider the 3−DBC in example 5.6.1 with $\delta = \delta_2 = \delta_3$. If $h_b(\tau * \delta_1) \leq h_b(\delta) < \frac{1+h_b(\delta_1 * \tau)}{2}$, then $(h_b(\tau * \delta_1) - h_b(\delta_1), 1 - h_b(\delta), 1 - h_b(\delta)) \notin \alpha_{\mathscr{U}}(\tau)$ but $(h_b(\tau * \delta_1) - h_b(\delta_1), 1 - h_b(\delta), 1 - h_b(\delta)) \in \mathbb{C}(\tau)$ and thus $\alpha_{\mathscr{U}}(\tau) \neq \mathbb{C}(\tau)$. In particular, if $\delta_1 = 0.01$ and $\delta_2 \in (0.1325, 0.21)$, then $\alpha_{\mathscr{U}}(\frac{1}{8}) \neq \mathbb{C}(\frac{1}{8})$.*

Chapter 6

Multiple access channel with distributed states

Consider a multiple access channel with distributed states (MAC-DSTx) depicted in figure 6.1. This is a simple MAC analogue of the PTP-STx studied in chapter 3. It should be evident to an informed reader that a natural extension of the Gel'fand Pinsker technique of binning followed by joint decoding yields an achievable rate region for the MAC-DSTx.[1] As a matter of fact, this is the currently known largest achievable rate region for an arbitrary MAC-DSTx. Gel'fand and Pinsker's technique of binning being optimal for PTP-STx, it is natural to ask the question whether it's extension to MAC-DSTx is optimal.

[1] In particular, each of encoders build codes over an auxiliary alphabet and partition the same into bins. From the bin indexed by the message, they choose codewords jointly typical with the state sequence and a function of this chosen codeword and the state sequence, evaluated letter-wise, is input on the channel. The decoder employs joint typical decoding to disambiguate the pair of codewords chosen by the encoder and thereby decodes the pair of messages.

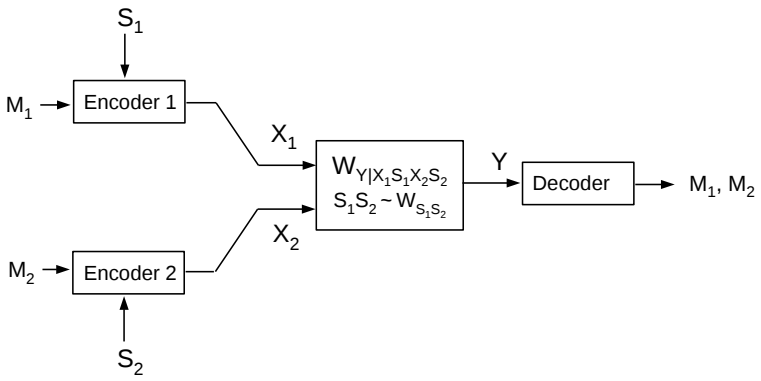

Figure 6.1: Multiple access channel with distributed states

Philosof and Zamir [15] propose an alternate technique for communicating over a particular symmetric additive binary doubly dirty MAC (BDD-MAC). As against to Gel'fand Pinsker's technique of partitioning the channel codes uniformly and independently into bins, they propose a partition of the two channel codes using cosets of a common linear code, thereby building dependency across the two codebooks and their codewords. Crucially exploiting the property of closure under addition, of bins in the two codebooks, they propose a coding technique, henceforth referred to as PZ-technique that achieves capacity of BDD-MAC. Furthermore, they prove strict sub-optimality of natural extension of Gel'fand Pinsker's technique of independent and unstructured binning.

Nevertheless ingenious, PZ-technique [15] is very specific to the additive and symmetric nature of the BDD-MAC studied therein. This technique being strictly more efficient than the currently known best strategy based on independent unstructured codes raises the following question. Is there a general coding framework for communicating over an *arbitrary* discrete MAC-DSTx, that reduces to the PZ-technique for the BDD-MAC, and that would yield an achievable rate region strictly larger than the best known achievable rate region using unstructured independent codes even for non-additive and non-symmetric MAC-DSTx?

In this chapter, we propose an algebraic framework based on nested coset codes for communication over an arbitrary MAC-DSTx and thereby answer the above questions in the affirmative. We present our framework in three pedagogical stages. We begin by identifying two key elements of PZ-technique 1) decoding mod-2 sum, instead of the pair of codewords chosen by the two transmitters and 2) choosing the bins of each user's code to be cosets of a common linear code to enable containment of the range of this mod-2 sum. The first stage, presented in section 6.2.2, captures all of the nontrivial elements of our framework in it's simplest setting. In this stage we employ nested coset codes built on finite fields, to decode the sum of codewords. The analysis of this technique enables us to derive a new achievable rate region for MAC-DSTx. The key elements of the first stage are (i) the use of nested coset codes to induce non-uniform input distributions, (ii) the use of joint typical encoding and decoding that enables us to analyze the probability of error over an *arbitrary* MAC-DSTx that is not constrained to be additive or symmetric, and (iii) an analysis of decoding of the sum of the pair of transmitted codewords chosen from two dependent codebooks. Indeed, the analysis of joint typical encoding and decoding of correlated codebooks with statistically dependent codewords involves several new elements. The reader is encouraged to peruse these in the proof of theorem 6.2.2.

The significance of the rate region proved achievable in the first stage is illustrated through examples in section 6.2.3.[2] In particular, we provide a simple modification of the BDD-MAC for which it is necessary to induce non-uniform input distributions and is more efficient to decode the sum of transmitted codewords. We also randomly

[2]The coding technique proposed in the first stage reduces to that proposed in [15] for BDD-MAC and moreover Philosof and Zamir have proved strict sub-optimality of unstructured independent coding for BDD-MAC. This in itself establishes significance of theorem 6.2.2. Notwithstanding this, it is easy to argue significance of our generalization by appealing to continuity. An additive channel can be perturbed slightly to result in a non-additive channel for which the technique proposed in [15] may not be applicable as is. By continuity of the rate regions as a function of the channel parameters, one can see why the proposed coding scheme must perform strictly better than unstructured independent coding. Example 6.2.5 presented in section 6.2.3 corroborates this.

perturb the BDD-MAC and demonstrate that coding framework proposed herein can outperform unstructured independent codes. The channels being non-additive, it is significantly harder to provide analytical comparisons, and hence we resort to direct computation of rate regions achievable using unstructured independent and nested coset codes. These examples illustrate that structured code based strategies do not hinge on the channel being additive but would benefit as long as the optimizing test channel from the auxiliary inputs to the channel output is not far from additive.

Does the rate region proved achievable using nested coset codes subsume the largest known achievable rate region using unstructured independent codes? It is our belief that strategies based on structured codes are not in lieu of their counterparts based on unstructured codes. In most cases, structured codes enable efficient decoding of a 'compressive'[3] function of the two codewords. However, for decoding both the codewords, it turns out the strategy of using a common linear code to effect partition of the two codebooks is not optimal, instead one has to employ two independent linear codes. The rate region achieved using the latter strategy is equivalent to that achieved using unstructured independent codes.[4]. This leads us to the second stage of our coding scheme which is presented in section 6.3. Following the approach of Ahlswede and Han [48, Section VI], we glue together structured and unstructured coding techniques to derive the largest known achievable rate region for communicating over a MAC-DSTx that combines structured and unstructured coding techniques. We present another simple modification of BDD-MAC to illustrate how the gluing of unstructured and structured coding techniques can yield a rate region larger than either one, and their union. We remark that in spite of our inability to compute the achievable rate region proposed in section 6.3, we are able to demonstrate the significance of the same through an example.

If the channel is far from additive, it may not be efficient to decode the sum, with respect to a finite field, of codewords. For example, if the MAC-DSTx is doubly dirty with field addition replaced by addition of an Abelian group, referred to as group addition or group sum, then it is natural to decode group sum of codewords. In other words, the technique of decoding sum of codewords must be generalized to decoding any arbitrary bivariate function of the auxiliary inputs. In the third stage of our coding scheme, presented in section 6.4, we consider decoding the group sum of the codewords. Specifically, codebooks are built over Abelian group alphabets and each encoder is provided with codebooks that possess a certain group structure. Analogous to the first stage, we propose joint typical encoding and decoding of group codes. Though essential elements of this analysis are similar to that of decoding sum of codewords chosen from nested coset codes over an *arbitrary* MAC-DSTx, the algebraic structure of a Abelian group being looser, leads to several new elements.

The importance of (i) decoding an appropriate bivariate function of codewords, and (ii) endowing codebooks with

[3] $f(U_1, U_2)$ is 'compressive' if $H(f(U_1, U_2))$ is significantly lower than $H(U_1, U_2)$.

[4] Indeed, for the problem of distributed reconstruction of modulo-2 sum of binary sources, Körner Marton strategy [18] based on common linear codes is outperformed by Slepian-Wolf [5] strategy (or equivalently the strategy of Csiszár based on independent linear codes [66].) for the class of source distributions for which the modulo-2 sum is not sufficiently compressive. More precisely, if $H(X \oplus Y) > \frac{H(X,Y)}{2}$, then it is better to reconstruct $X \oplus Y$ using the technique of Slepian-Wolf or Csiszár.

the appropriate algebraic structure is illustrated through an example discussed in section 6.4. Specifically, we indicate using numerical computation that for a quaternary doubly dirty MAC-DSTx (QDD-MAC) wherein the operation is mod-4 addition, decoding mod-4 sum, which is the group operation in the quaternary alphabet, of the codewords strictly outperforms both independent unstructured and nested coset codes based strategies. In fact, significant gains for this problem are achievable using Abelian group codes. The reader is encouraged to peruse details in section 6.4.

Several findings in the context of multi-terminal communication problems point to efficient strategies based on structured codes. Nazer and Gastpar [16] propose a strategy based on linear codes for computing the sum of sources over additive multiple access channels that outperforms earlier known strategies. Building on this technique, we develop a framework for computing sum of sources over an arbitrary multiple access channel in [67]. Sridharan et. al. [19] propose a coding technique based on lattices for communicating over a $K-$user Gaussian interference channel ($K \geq 3$) that outperforms a natural extension of Han-Kobayashi technique [13] under the Gaussian input distribution. We propose an analogous coding technique based on nested linear codes [68] for the general discrete 3$-$user interference channel and identify an example for which the proposed technique outperforms the natural extension of Han-Kobayashi technique [13]. Krithivasan and Pradhan [23] propose a framework based on structured codes for the distributed source coding problem that outperforms the best known strategy based on unstructured independent codes due to Berger and Tung [64]. The reader is also referred to [69], wherein lattices are employed to efficiently reconstruct linear functions of Gaussian sources.

6.1 MAC-DSTx: Definitions, largest known achievable rate region

In this section, we lay the necessary groundwork. In particular, we describe MAC-DSTx and precisely state relevant notions such as code, achievability in section 6.1.1. In section 6.1.2, we provide a characterization of the currently known largest achievable rate region. We illustrate this rate region for BDD-MAC in section 6.1.3 and highlight the reasons for it's suboptimality. This will set the stage for it's enlargement in subsequent sections.

6.1.1 Definitions : MAC-DSTx, code and achievability

Consider the two user multiple access analogue of PTP-STx [7]. Let $\mathcal{X}_1$ and $\mathcal{X}_2$ denote finite input alphabet sets and $\mathcal{Y}$, the output alphabet set. Transition probabilities depend on a random vector parameter $\boldsymbol{S} := (S_1, S_2)$, called state, that takes values in a finite set $\mathcal{S} := \mathcal{S}_1 \times \mathcal{S}_2$. The discrete time channel is (i) time invariant, i.e., pmf of Y_i, the output at time i, conditioned on inputs $\boldsymbol{X_i} := (X_{1i}, X_{2i})$ and state $\mathbf{S_i} := (S_{1i}, S_{2i})$ at time i, is invariant with i, (ii) memoryless, i.e., Y_i is conditionally independent of $(\boldsymbol{X_t}, \boldsymbol{S_t}) : 1 \leq t < i$ given $\boldsymbol{X_i}, \boldsymbol{S_i}$, and (iii) used without feedback. Let $W_{Y|\boldsymbol{XS}}(y|\boldsymbol{x}, \boldsymbol{s})$ be the probability of observing $y \in \mathcal{Y}$ at the output given $\boldsymbol{x} := (x_1, x_2) \in \mathcal{X} := \mathcal{X}_1 \times \mathcal{X}_2$ is input to the channel in state $\boldsymbol{s} := (s_1, s_2) \in \mathcal{S}$. The state at time i, $\boldsymbol{S_i}$ is (i) independent of $(\boldsymbol{S_t}, \boldsymbol{X_t}, Y_t) : 1 \leq t < i$, and (ii) identically distributed for all i. Let $W_{\boldsymbol{S}}(\boldsymbol{s})$ be the probability of MAC-DSTx being in state $\boldsymbol{s} \in \mathcal{S}$. We assume S_j^n

is non-causally known to encoder j. Input X_j is constrained with respect to a cost function $\kappa_j : \mathcal{X}_j \times \mathcal{S}_j \to [0, \infty)$. We assume that the cost is time-invariant and additive i.e., cost of input X_j^n at input j to the channel in state $\boldsymbol{S}^n$ is $\bar{\kappa}_j^n(X_j^n, S_j^n) := \frac{1}{n}\sum_{i=1}^n \kappa_j(X_{ji}, S_{ji})$. We refer to this channel as MAC-DSTx $(\mathcal{S}, W_{\boldsymbol{S}}, \mathcal{X}, \kappa, \mathcal{Y}, W_{Y|\boldsymbol{X},\boldsymbol{S}})$. Towards characterizing a new inner bound for the capacity region of a MAC-DSTx, we begin with definitions of relevant notions such as achievability and capacity.

Definition 6.1.1 *A MAC-DSTx code $(n, \mathscr{M}_1, \mathscr{M}_2, e_1, e_2, d)$ consists of (i) index sets $\mathcal{M}_j$ of messages, of cardinality $\mathscr{M}_j$ for $j = 1, 2$ (ii) encoder maps $e_j : \mathcal{M}_j \times \mathcal{S}_j^n \to \mathcal{X}_j^n$ for $j = 1, 2$, and (iii) a decoder map $d : \mathcal{Y}^n \to \mathcal{M}_1 \times \mathcal{M}_2$.*

We let $\mathscr{M} := (\mathscr{M}_1, \mathscr{M}_2)$, $\mathbf{e} := (e_1, e_2)$ and refer to above as MAC-DSTx code $(n, \mathscr{M}, \mathbf{e}, d)$. Assuming the pair of messages to be uniformly distributed, we define the average error probability and the cost of a MAC-DSTx code as follows.

Definition 6.1.2 *The average error probability of MAC-DSTx code $(n, \mathscr{M}, \mathbf{e}, d)$ conditioned on message $\mathbf{m} := (m_1, m_2) \in \mathcal{M} := \mathcal{M}_1 \times \mathcal{M}_2$ is*

$$\xi(\mathbf{e}, d|\mathbf{m}) := \sum_{\boldsymbol{s}^n \in \mathcal{S}^n} W_{\boldsymbol{S}^n}(\boldsymbol{s}^n) \sum_{y^n : d(y^n) \neq \mathbf{m}} W_{Y^n|\boldsymbol{X}^n, \boldsymbol{S}^n}(y^n | e_1(m_1, s_1^n), e_2(m_2, s_2^n), \boldsymbol{s}^n).$$

The average error probability is $\bar{\xi}(\mathbf{e}, d) := \sum_{\mathbf{m} \in \mathcal{M}} \frac{1}{\mathscr{M}_1 \mathscr{M}_2} \xi(\mathbf{e}, d|\mathbf{m})$. The average cost of transmitting message pair $\mathbf{m}$ is $\tau(\mathbf{e}|\mathbf{m}) := (\tau_1(e_1|m_1), \tau_2(e_2|m_2))$, where

$$\tau_j(e_j|m_j) := \sum_{s_j^n \in \mathcal{S}_j^n} W_{S_j^n}(s_j^n) \bar{\kappa}_j^n(e_j(m_j, s_j^n), s_j^n).$$

The average cost of the code is $\tau(\mathbf{e}) := \sum_{\mathbf{m} \in \mathcal{M}} \frac{1}{\mathscr{M}_1 \mathscr{M}_2} \tau(\mathbf{e}|\mathbf{m})$, where $\tau(\mathbf{e}) = (\tau(e_1), \tau(e_2))$.

Definition 6.1.3 *A rate cost quadruple $(\mathbf{R}, \boldsymbol{\tau}) \in [0, \infty)^4$ is achievable if for every $\eta > 0$, there exists $N(\eta) \in \mathbb{N}$ such that for all $n > N(\eta)$, there exists a MAC-DSTx code $(n, \mathcal{M}^{(n)}, \mathbf{e}^{(n)}, d^{(n)})$ such that (i) $\frac{\log \mathcal{M}_j^{(n)}}{n} \geq R_j - \eta$ for $j = 1, 2$, (ii) $\bar{\xi}(\mathbf{e}^{(n)}, d^{(n)}) \leq \eta$, and (iii) $\tau_j(e_j^{(n)}) \leq \tau_j + \eta$, for $j = 1, 2$. The capacity region $\mathbb{C}(\boldsymbol{\tau}) := cocl\big(\{\mathbf{R} \in [0, \infty)^2 : (\mathbf{R}, \boldsymbol{\tau}) \text{ is achievable}\}\big)$.*

The coding technique that achieves capacity of PTP-STx [7] can be generalized to obtain an achievable rate region for MAC-DSTx. For a general MAC-DSTx this is the largest known inner bound to $\mathbb{C}(\boldsymbol{\tau})$. We provide a characterization of the same in the following section.

6.1.2 Largest known achievable rate region using unstructured codes

Definition 6.1.4 *Let $\mathbb{D}(\boldsymbol{\tau})$ be collection of pmfs $p_{\mathbf{U}\boldsymbol{X}\boldsymbol{S}Y}$ on $\mathcal{U}^2 \times \mathcal{X} \times \mathcal{S} \times \mathcal{Y}$, where $\mathbf{U}$ denotes U_1, U_2 and $\mathcal{U}^2$ is a two fold Cartesian product of a finite set $\mathcal{U}$, such that (i) $p_{\boldsymbol{S}} = W_{\boldsymbol{S}}$, (ii) $p_{Y|\boldsymbol{X}\boldsymbol{S}\mathbf{U}} = p_{Y|\boldsymbol{X}\boldsymbol{S}} = W_{Y|\boldsymbol{X}\boldsymbol{S}}$, (iii) $p_{U_j|\boldsymbol{S}U_{\bar{j}}} =$*

$p_{U_j|\boldsymbol{S}} = p_{U_j|S_j}$ *and* $p_{X_j|\boldsymbol{S}\mathbf{U}X_{\not j}} = p_{X_j|\boldsymbol{S}\mathbf{U}} = p_{X_j|S_jU_j}$ *for any distinct elements* $j, \not j \in \{1,2\}$, *(iv)* $p_{X_j|S_jU_j}(x_j|s_j,u_j) \in \{0,1\}$ *for all* $(u_j,s_j,x_j), j=1,2$ *and (v)* $\mathbb{E}\{\kappa_j(X_j,S_j)\} \leq \tau_j$ *for* $j=1,2$. *For* $p_{\mathbf{U}\boldsymbol{X}\boldsymbol{S}Y} \in \mathbb{D}(\boldsymbol{\tau})$, *let* $\alpha(p_{\mathbf{U}\boldsymbol{X}\boldsymbol{S}Y})$ *be defined as the set*

$$\left\{ \begin{array}{ll} (R_1,R_2) \in [0,\infty)^2 : & R_1 \leq I(U_1;YU_2) - I(U_1;S_1), R_2 \leq I(U_2;YU_1) - I(U_2;S_2), \\ & R_1 + R_2 \leq I(\mathbf{U};Y) + I(U_1;U_2) - \sum_{j=1}^{2} I(U_j;S_j) \end{array} \right\}$$

and

$$\alpha(\tau) := cocl\left(\bigcup_{p_{\mathbf{U}\boldsymbol{X}\boldsymbol{S}Y} \in \mathbb{D}(\tau)} \alpha(p_{\mathbf{U}\boldsymbol{X}\boldsymbol{S}Y}) \right).$$

Theorem 6.1.5 $\alpha(\tau) \subseteq \mathbb{C}(\boldsymbol{\tau})$. □

Achievability of $\alpha(p_{\mathbf{U}\boldsymbol{X}\boldsymbol{S}Y})$ can be proved by employing the encoding technique proposed by Gel'fand and Pinsker [7] at each encoder and joint decoding proposed by Ahlswede [3], Liao [4]. In the sequel, we provide an illustration of this coding technique for BDD-MAC.

6.1.3 Rate region achievable using unstructured codes for BDD-MAC

Philosof and Zamir characterize $\mathbb{C}(\boldsymbol{\tau})$ for BDD-MAC using PZ-technique and prove $\alpha(\boldsymbol{\tau}) \subsetneq \mathbb{C}(\boldsymbol{\tau})$ for the same. In order to identify the key elements of PZ-technique, we briefly analyze unstructured coding (this section), PZ-technique (section 6.2.1) and set the stage for a new coding scheme.

BDD-MAC is a MAC-DSTx with binary alphabets $\mathcal{S}_j = \mathcal{X}_j = \mathcal{Y} = \{0,1\}$, $j=1,2$. The state sequences are independent Bernoulli-$\frac{1}{2}$ processes, i.e., $W_{\boldsymbol{S}}(\boldsymbol{s}) = \frac{1}{4}$ for all $\boldsymbol{s} \in \mathcal{S}$. The channel transition is described by the relation $Y = X_1 \oplus_2 S_1 \oplus_2 X_2 \oplus_2 S_2$. An additive Hamming cost is assumed on the input, i.e., $\kappa_j(1,s_j) = 1$ and $\kappa_j(0,s_j) = 0$ for any $s_j \in \mathcal{S}_j$, $j=1,2$ and the input is subject to a symmetric cost constraint $\boldsymbol{\tau} = (\tau,\tau)$.

We describe the test channel $p_{\mathbf{U}\boldsymbol{S}\boldsymbol{X}Y} \in \mathbb{D}(\boldsymbol{\tau})$ that achieves $\alpha(\boldsymbol{\tau})$. For each user j, consider the test channel that achieves the Gel'fand-Pinsker capacity treating the other user as noise i.e., $p_{U_jS_jX_j}(0,1,1) = p_{U_jS_jX_k}(1,0,1) = \frac{\tau}{2}$, $p_{U_jS_jX_j}(0,0,0) = p_{U_jS_jX_j}(1,1,0) = \frac{1-\tau}{2}$. Philosof and Zamir prove $p_{\mathbf{U}\boldsymbol{S}\boldsymbol{X}} = p_{U_1S_1X_1}p_{U_2S_2X_2}$ achieves $\alpha(\boldsymbol{\tau}) = \{\mathbf{R} : R_1 + R_2 \leq |2h_b(\tau) - 1|^+\}$, where $|\cdot|^+$ denotes upper convex envelope.

Let us take a closer look at achievability of the vertex $(2h_b(\tau)-1, 0)$ using the above test channel. Since user 2 has no message to transmit, it picks a single bin with roughly $2^{nI(U_2;S_2)} = 2^{n(1-h_b(\tau))}$ codewords independently and uniformly from the entire space of binary vectors. User 1 picks 2^{nR_1} bins each with roughly $2^{nI(U_1;S_1)} = 2^{n(1-h_b(\tau))}$ independently and uniformly distributed binary vectors. Encoder 2 observes S_2^n and chooses a codeword, say U_2^n, that is within a Hamming distance of roughly $n\tau$ from S_2^n and transmits $X_2^n = U_2^n \oplus_2 S_2^n$. Encoder 1 performs a similar encoding, except that it restricts the choice of U_1^n to the bin indexed by user 1's message, and transmits $X_1^n = U_1^n \oplus_2 S_1^n$.

What is the maximum rate R_1 at which user 1 can transmit it's message? Decoder receives $Y^n = U_1^n \oplus_2 U_2^n$ and looks for all pairs of codewords that are jointly typical with Y^n. Since any pair of binary $n-$length vectors are jointly typical (U_1 and U_2 are independent and uniform), the decoding rule reduces to finding all pairs of binary $n-$length vectors in the pair of codebooks that sum to the received vector Y^n. All bins chosen independently without structure imply that any bin of user 1's codebook when added to the user 2's codebook (a single bin) results in roughly $2^{n(2-2h_b(\tau))}$ distinct vectors. Therefore, we cannot hope to pack more than roughly $\frac{2^n}{2^{n(2-2h_b(q))}} = 2^{n(2h_b(q)-1)}$ bins in user 1's codebook. We remark that *an explosion in the range of sum of transmitted codewords severely limits achievable rate.*

We make a few observations. Effectively, communication occurs over the $(U_1, U_2) - Y$ channel and the test channel induces the Markov chain $(U_1, U_2) - U_1 \oplus_2 U_2 - Y$. It would therefore be more efficient to communicate information over the $U_1 \oplus_2 U_2 - Y$ channel which suggests an efficient utilization of $U_1 \oplus_2 U_2-$space. Having chosen codewords in each bin independently and moreover the two users' bins independently, each message pair utilizes $2^{n(2-2h_b(\tau))}$ vectors in the $U_1 \oplus_2 U_2-$space. In section 6.2.1, we summarize PZ-technique, wherein the algebraic structure in the codebooks is exploited for more efficient utilization of $U_1 \oplus_2 U_2-$space.

6.2 An achievable rate region using nested coset codes

6.2.1 Nested linear codes for BDD-MAC

We present PZ-technique proposed for BDD-MAC. The encoding and decoding techniques are similar to that stated in 6.1.3 except for one key difference. The bins of user 1 and 2's codebooks are cosets of a common linear code. In particular, let λ_I denote a linear code of rate roughly equal to $1 - h_b(\tau)$ that can quantize a uniform source, state S_j^n in our case, within an average Hamming distortion of τ. Since user 2 has no message to transmit, it employs λ_I as it's only bin. Encoder 1 employs 2^{nR_1} cosets of λ_I within a larger linear code, called λ_O, as it's bins. Note that rate of λ_O is roughly $R_1 + 1 - h_b(\tau)$. Encoding rule is as described in section 6.1.3.

The codebook of user 2 when added to any bin of user 1's code results in a coset of λ_I, and therefore contains approximately at most $2^{n(1-h_b(\tau))}$ codewords. Moreover, since U_1^n lies in λ_I, user 2's codeword U_2^n and the received vector $Y^n = U_1^n \oplus_2 U_2^n$ lie in the same coset.[5] Since the channel is noiseless, user 1 may employ all cosets of λ_I and therefore communicate at rate $h_b(\tau)$ which is larger than $2h_b(\tau) - 1$ for all $\tau \in (0, \frac{1}{2})$.

Let us identify key elements of PZ-technique. Each message pair corresponds to roughly $2^{n(1-h_b(\tau))}$ vectors in $U_1 \oplus_2 U_2-$space, resulting in a more efficient utilization of this space. This indeed is the difference in the sum rate achievable using independent unstructured codes and PZ-technique. We also note the *decoder does not attempt to disambiguate the pair* (U_1^n, U_2^n) *and restricts to decoding* $U_1^n \oplus_2 U_2^n$. *This is motivated by the Markov chain*

[5]This is also because the channel is noiseless.

$(U_1, U_2) - U_1 \oplus_2 U_2 - Y$ *induced by the test channel and the use of structured codebooks that contain the sum.*

It is instructive to investigate the efficacy of this technique if users 1 and 2 employ distinct linear codes $\lambda_{I1}, \lambda_{I2}$ of rate $1 - h_b(\tau)$ instead of a common linear code λ_I. In this case, each message of user 1 can result in $2^{2-2h_b(\tau)}$ received vectors which restricts user 1's rate to $2h_b(q) - 1$ and provides no improvement over the unstructured coding technique. We conclude that *if the bins of the MAC channel code are nontrivial, as in this case due to the presence of a state, then it maybe beneficial to endow the bins with an algebraic structure that restricts the range of a bivariate function, and enable the decoder decode this function of chosen codewords.*

6.2.2 Stage I : An achievable rate region for MAC-DSTx using nested coset codes

In this section, we present the first stage of our coding scheme that uses joint typical encoding and decoding and nested coset codes over an arbitrary MAC-DSTx. The technique proposed by Philosof and Zamir is specific to the binary doubly dirty MAC - Hamming cost constraint that induces additive test channels between the auxiliary and state random variables, and additive and symmetric nature of the channel. Moreover, linear codes only achieve the symmetric capacity, and therefore if the output were obtained by passing $(X_1^n \oplus_2 S_1^n, X_2^n \oplus_2 S_2^n)$ through an asymmetric MAC, linear codes though applicable, might not be optimal.

We begin with a characterization of test channels followed by achievability.

Definition 6.2.1 *Let $\mathbb{D}_f(\boldsymbol{\tau}) \subseteq \mathbb{D}(\boldsymbol{\tau})$ be the collection of distributions $p_{\boldsymbol{VSXY}}$ on $\mathcal{V}^2 \times \mathcal{S} \times \mathcal{X} \times \mathcal{Y}$ where $\mathcal{V}$ is a finite field. For $p_{\boldsymbol{VXSY}} \in \mathbb{D}_f(\boldsymbol{\tau})$, let $\beta_f(p_{\boldsymbol{VXSY}})$ be defined as the set*

$$\left\{(R_1, R_2) \in [0, \infty)^2 : R_1 + R_2 \leq \min\left\{H(V_1|S_1), H(V_2|S_2)\right\} - H(V_1 \oplus V_2|Y)\right\}. \tag{6.1}$$

Let

$$\beta_f(\boldsymbol{\tau}) := \mathit{cocl}\left(\bigcup_{p_{\boldsymbol{VXSY}} \in \mathbb{D}_f(\boldsymbol{\tau})} \beta_f(p_{\boldsymbol{VXSY}})\right)$$

Theorem 6.2.2 $\beta_f(\boldsymbol{\tau}) \subseteq \mathbb{C}(\boldsymbol{\tau})$. □

Before we provide a proof, we state the coding technique and indicate achievability of promised rates. As stated in section 6.2.1, the key aspect is to employ cosets of a common linear as a bin for quantizing the state. We employ three nested coset codes -one each for the two encoders and the decoder- that share a common inner (sparser) code. We begin by describing the encoding rule. The nested coset code provided to encoder j is described through a pair of generator matrices $g_I \in \mathcal{V}^{k \times n}$ and $g_{Oj/I} \in \mathcal{V}^{l_j \times n}$ where (i) g_I and $g_{Oj}^T := \begin{bmatrix} g_I^T & g_{Oj/I}^T \end{bmatrix}$ are generator matrices of

inner (sparser) and complete (denser) codes respectively, (ii)

$$\frac{k}{n} > 1 - \frac{\min\{H(V_1|S_1), H(V_2|S_2)\}}{\log \pi} \tag{6.2}$$

$$\frac{k+l_1+l_2}{n} < 1 - \frac{H(V_1 \oplus V_2|Y)}{\log \pi}. \tag{6.3}$$

with $\pi := |\mathcal{V}|$ and (iii) bias vector b_j^n. Let λ_I and λ_{Oj} denote linear codes corresponding to generator matrices g_I and g_{Oj} respectively. User j's message $M_j^{l_j} \in \mathcal{V}^{l_j}$ indexes the coset $(a^k g_I \oplus M_j^{l_j} g_{Oj/I} \oplus b_j^n : a^k \in \mathcal{V}^k)$. Encoder j observes state S_j^n and looks for a codeword in the coset indexed by the message that is jointly typical with the state sequence S_j^n according to $p_{S_jV_j}$. If it finds one such codeword, say V_j^n, a vector X_j^n is generated according $\prod_{t=1}^n p_{X_j|S_jV_j}(\cdot|S_{jt}V_{jt})$ and X_j^n is fed as input to the channel. Otherwise, it declares an error.

Now to the decoding rule. Let λ_O denote the complete code provided to the decoder, i.e., the coset code whose (i) generator matrix is $g_O^T := \begin{bmatrix} g_I^T & g_{O/I}^T \end{bmatrix}$, where $g_{O/I}^T := \begin{bmatrix} g_{O1/I}^T & g_{O2/I}^T \end{bmatrix}$ and (ii) bias vector $b_1^n \oplus b_2^n$. Having received Y^n, it lists all codewords in λ_O that are jointly typical with Y^n with respect to $p_{V_1 \oplus V_2, Y}$. If all such codewords belong to a unique coset (of λ_I in λ_O) say $(a^k g_I \oplus m_1^{l_1} g_{O1/I} \oplus m_2^{l_2} g_{O2/I} : a^k \in \mathcal{V}^k)$, it declares $(m_1^{l_1}, m_2^{l_2})$ as the pair of decoded messages. Otherwise, it declares an error.

We pick entries of each of the constituent generator matrices $g_I, g_{O1/I}, g_{O2/I}$ independently and uniformly from $\mathcal{V}$. Lower bound (7.1) enable us to drive down the probability of encoder not finding a jointly typical codeword in the indexed coset. This bound can be interpreted easily. If we picked codewords according to $\prod_{t=1}^n p_V$, then we need the bin to be of rate roughly $H(V_1) - H(V_1|S_1)$. Since we average uniformly over the ensemble of coset codes, each codeword of a linear code is uniformly distributed over $\mathcal{V}^n$. Hence the bin must of rate at least $\log \pi - H(V_1|S_1)$. The decoder makes an error with arbitrarily small probability if (6.3) is satisfied. This bound can also be interpreted intuitively. If the codewords were picked according to $p_{V_1 \oplus V_2}$, the upper bound would have been $H(V_1 \oplus V_2) - H(V_1 \oplus V_2|Y)$. In this case, the codewords in the sum of nested linear codes are also uniformly distributed over $\mathcal{V}^n$, and this explains the bound in (6.3). From (7.1), (6.3) it can be verified that $R_1 + R_2 = \frac{l_1+l_2}{n} \leq \min\{H(V_1|S_1), H(V_2|S_2) - H(V_1 \oplus V_2|Y)\}$ is achievable.

We emphasize that joint typical encoding and decoding enables us to decode the sum over an arbitrary MAC-DSTx. The informed reader will recognize the need to prove statistical independence of a codeword in a competing sum coset and the pair of cosets indexed by the messages. The dependence built across the codewords and cosets as a consequence of the algebraic structure exemplifies the interplay of algebra and probability. The following proof details these elements.

Proof: Let pmf $p_{\boldsymbol{VXS}Y} \in \mathbb{D}_f(\boldsymbol{\tau})$, rate pair $\mathbf{R} \in \beta_f(p_{\boldsymbol{VXS}Y})$ and $\eta > 0$. We prove existence of a MAC-DSTx code $(n, \mathscr{M}, \mathbf{e}, d)$ whose rate $\frac{\log \mathscr{M}_j}{n} \geq R_j - \eta$, average error probability $\overline{\xi}(\mathbf{e}, d) \leq \eta$, and average cost $\tau(e_j) \leq \tau_j + \eta$ for $j = 1, 2$.

We begin with a description of the structure of the MAC-DSTx code whose existence we seek to prove. Let

$\pi := |\mathcal{V}|$ and we assume $H(V_1|S_1) \geq H(V_2|S_2)$ without loss of generality. Consider a pair of nested coset codes $(n, k_j, l_j, g_{I_j}, g_{O_j/I_j}, b_j^n) : j = 1, 2$ built over $\mathcal{V}$, denoted $\lambda_{O_j}/\lambda_{I_j} : j = 1, 2$ with parameters

$$\begin{aligned}
k_1 &:= \lceil n\left(1 - \frac{H(V_1|S_1)}{\log \pi} + \frac{\eta_1(\eta)}{\log \pi}\right)\rceil, && (6.4)\\
k_2 = k_1 + k_+, \text{ where } k_+ &:= \lceil n\left(1 - \frac{H(V_2|S_2)}{\log \pi} + \frac{\eta_1(\eta)}{\log \pi}\right)\rceil - k_1, && (6.5)\\
l_1 &:= \lfloor n\left(\frac{R_1}{\log \pi} - \frac{\eta_2(\eta)}{\log \pi}\right)\rfloor && (6.6)\\
l_2 &:= \lfloor n\left(1 + \frac{R_2}{\log \pi} - \frac{H(V_2|S_2)}{\log \pi} - \frac{\eta_3(\eta)}{\log \pi}\right)\rfloor - k_2, \text{ and,} && (6.7)\\
\text{the first } k_1 \text{ rows of } g_{I_1} &\text{ and } g_{I_2} \text{ are identical i.e., } g_{I_1,t} = g_{I_2,t} \text{ for } t \in [k_1]. && (6.8)
\end{aligned}$$

A few remarks on the structure of $\lambda_{O_j}/\lambda_{I_j} : j = 1, 2$ and the relationship between their parameters are in order. For $n \geq N_1(\eta) := \max\left\{\frac{\log \pi}{\eta_1(\eta)}, \frac{\log \pi}{\eta_2(\eta)}, \frac{\log \pi}{\eta_3(\eta)}\right\}$, we have

$$\begin{aligned}
\frac{n}{\log \pi}(\log \pi - H(V_j|S_j) + \eta_1(\eta)) \leq \quad & k_j \quad && \leq \frac{n}{\log \pi}(\log \pi - H(V_j|S_j) + 2\eta_1(\eta)) && (6.9)\\
\frac{n}{\log \pi}(R_1 - 2\eta_2(\eta)) \leq \quad & l_1 \quad && \leq \frac{n}{\log \pi}(R_1 - \eta_2(\eta)) && (6.10)\\
\frac{n}{\log \pi}(R_2 + \log \pi - H(V_2|S_2) - 2\eta_3(\eta)) \leq \quad & k_2 + l_2 \quad && \leq \frac{n}{\log \pi}(R_2 + \log \pi - H(V_2|S_2) - \eta_3(\eta)) && (6.11)
\end{aligned}$$

Combining the lower bound in (6.11) and the upper bound for k_2 in (6.9), we have

$$\frac{l_2 \log \pi}{n} \geq R_2 - 2\eta_3(\eta) - 2\eta_1(\eta) \tag{6.12}$$

and similarly, combining the upper bound for $k_2 + l_2$ in (6.11) and the upper bound for l_1 in (6.10), we have

$$\begin{aligned}
k_2 + l_1 + l_2 &\leq \frac{n}{\log \pi}\left(R_1 + R_2 + \log \pi - H(V_2|S_2) - \eta_3(\eta) - \eta_2(\eta)\right)\\
&\leq \frac{n}{\log \pi}\left(\log \pi - H(V_1 \oplus V_2|Y) - \eta_3(\eta) - \eta_2(\eta)\right), && (6.13)
\end{aligned}$$

where (6.13) follows from $\mathbf{R} \in \beta_f(p_{\boldsymbol{VXS}Y})$.

We now specify encoding and decoding rules that map this pair $\lambda_{O_j}/\lambda_{I_j} : j = 1, 2$ of nested coset codes into a MAC-DSTx code. User j is provided with the nested coset code $\lambda_{O_j}/\lambda_{I_j}$. User j's message is used to index one among π^{l_j} cosets of $\lambda_{O_j}/\lambda_{I_j}$. We assume that the set of messages $\mathcal{M}_j := \mathcal{V}^{l_j}$, and $M_j^{l_j} \in \mathcal{V}^{l_j}$ to be the uniformly distributed random variable representing user j's message. We let $v_j^n(a_j^{k_j}, m_j^{l_j}) := a_j^{k_j} g_{I_j} \oplus m_j^{l_j} g_{O_j/I_j} \oplus b_j^n$ denote a generic codeword in $\lambda_{O_j}/\lambda_{I_j}$ and $c_j(m_j^{l_j}) := (v_j^n(a_j^{k_j}, m_j^{l_j}) : a_j^{k_j} \in \mathcal{V}^{k_j})$ denote the coset corresponding to message $m_j^{l_j}$. Encoder j observes the state sequence S_j^n and populates the list $L_j(M_j^{l_j}, S_j^n) =$

$\left\{v_j(a_j^{k_j}, M_j^{l_j}) : (S_j^n, v_j(a_j^{k_j}, M_j^{l_j})) \in T_{\eta_4(\eta)}(S_j, V_j)\right\}$ of codewords in the coset corresponding to the message that are jointly typical with the state sequence. If $L_j(M_j^{l_j}, S_j^n)$ is empty, it picks a codeword uniformly at random from coset $c_j(M_j^{l_j})$. Otherwise, it picks a codeword uniformly at random from $L_j(M_j^{l_j}, S_j^n)$. Let $V_j(A_j^{k_j}, M_j^{l_j})$ denote the picked codeword in either case. The encoder computes $X_j^n(M_j^{l_j}, S_j^n) := f_j^n(V_j^n(A_j^{k_j}, M_j^{l_j}), S_j^n)$, where $f_j : \mathcal{V}_j \times \mathcal{S}_j \to \mathcal{X}_j$ is any map that satisfies $p_{X_j|V_jS_j}(f_j(v_j, s_j)|v_j, s_j) = 1$ for all pairs $(v_j, s_j) \in \mathcal{V}_j \times \mathcal{S}_j$. $X_j^n(M_j^{l_j}, S_j^n)$ is fed as input to the channel.

We now specify the decoding rule. The decoder is provided with nested coset code $(n, k, l, g_I, g_{O/I}, b^n)$ denoted λ_O/λ_I where $k = k_2$, $l = l_1 + l_2$, $g_I = g_{I_2}$, $g_{O/I}^T := \begin{bmatrix} g_{O_1/I_1}^T & g_{O_2/I_2}^T \end{bmatrix}$ and $b^n := b_1^n \oplus b_2^n$. With a slight abuse of notation, we let $m^l := (m_1^{l_1}, m_2^{l_2}) \in \mathcal{V}^l := \mathcal{V}^{l_1} \times \mathcal{V}^{l_2}$ represent a pair of messages and analogously random variable $M^l := (M_1^{l_1}, M_2^{l_2})$ denote the pair of user messages. For $a^k \in \mathcal{V}^k$ and $m^l \in \mathcal{V}^l$, let $v^n(a^k, m^l) := a^k g_I \oplus m^l g_{O/I} \oplus b^n$ and $c(m^l) := (v^n(a^k, m^l) : a^k \in \mathcal{V}^k)$ denote a generic codeword in λ_O/λ_I and the coset corresponding to the message pair m^l respectively. The decoder observes the received vector Y^n and populates $D(Y^n) := \left\{m^l \in \mathcal{V}^l : \exists v^n(a^k, m^l) \text{ such that } (v^n(a^k, m^l), Y^n) \in T_{\eta_5(\eta)}(V_1 \oplus V_2, Y)\right\}$. If $D(Y^n)$ is a singleton, the decoder declares the content of $D(Y^n)$ as the decoded message pair. Otherwise, it declares an error.

The above encoding and decoding rules map every pair $\lambda_{O_j}/\lambda_{I_j} : j = 1, 2$ of nested coset codes that satisfy (6.4)-(6.8) into a corresponding MAC-DSTx code $(n, \mathscr{M}^{(n)}, \mathbf{e}^{(n)}, d^{(n)})$ of rate $\frac{\log \mathscr{M}_j^{(n)}}{n} \geq R_j - 2\eta_1(\eta) - 2\eta_2(\eta)$, thus characterizing an ensemble, one for each n, of MAC-DSTx codes. We average the error probability over this ensemble of MAC-DSTx codes by letting the bias vectors $B_j^n : j = 1, 2$ and generator matrices $G_{I_2}, G_{O_j/I_j} : j = 1, 2$ mutually independent and uniformly distributed over their respective range spaces. Let $\Lambda_{O_j}/\Lambda_{I_j} : j = 1, 2$ and Λ_O/Λ_I denote the random nested coset codes $(n, k_j, l_j, G_{I_j}, G_{O_j/I_j}, B_j^n) : j = 1, 2$ and $(n, k, l, G_I, G_{O/I}, B^n)$ respectively. For $a_j^{k_j} \in \mathcal{V}^{k_j}$, $m_j^{l_j} \in \mathcal{V}^{l_j}$, $a^k \in \mathcal{V}^k$, $m^l \in \mathcal{V}^l$, let $V_j^n(a_j^{k_j}, m_j^{l_j}) := a_j^{k_j} G_{I_j} \oplus m_j^{l_j} G_{O_j/I_j} \oplus B_j^n : j = 1, 2$, $V^n(a^k, m^l) := a^k G_I \oplus m^l G_{O/I} \oplus B^n$ denote corresponding random codewords in $\Lambda_{O_j}/\Lambda_{I_j} : j = 1, 2$ and Λ_O/Λ_I respectively. Let $C_j(m_j^{l_j}) := (V_j^n(a_j^{k_j}, m_j^{l_j}) : a_j^{k_j} \in \mathcal{V}^{k_j})$ and $C(m^l) := (V^n(a^k, m^l) : a^k \in \mathcal{V}^k)$ denote random cosets in $\Lambda_{O_j}/\Lambda_{I_j} : j = 1, 2$ and Λ_O/Λ_I corresponding to message $m_j^{l_j} : j = 1, 2$ and m^l respectively.

Our next goal is to derive an upper bound on the probability of error. Towards this end, we begin with a characterization of related events. Let

$$
\begin{aligned}
\epsilon_{1j} &:= \{S_j^n \notin T_{\frac{\eta_4(\eta)}{2}}(S_j)\}, \quad \epsilon_1 := \left\{\boldsymbol{S}^n \notin T_{\frac{\eta_4(\eta)}{2}}(\boldsymbol{S})\right\} \\
\epsilon_{2j} &:= \{\phi_j(S_j^n, M_j^{l_j}) = 0\}, \text{ where } \phi_j(s_j^n, m_j^{l_j}) := \sum_{a_j^{k_j} \in \mathcal{V}^{k_j}} 1_{\left\{\left(V_j^n(a_j^{k_j}, m_j^{l_j}), s_j^n\right) \in T_{\eta_4(\eta)}(V_j, S_j)\right\}} \\
\epsilon_4 &:= \bigcup_{a^k \in \mathcal{V}^k} \left\{(V^n(a^k, M^l), Y^n) \in T_{\eta_5(\eta)}(V_1 \oplus V_2, Y)\right\} \\
\epsilon_5 &:= \bigcup_{\hat{m}^l \neq M^l} \bigcup_{a^k \in \mathcal{V}^k} \left\{\left(V^n(a^k, \hat{m}^l), Y^n\right) \in T_{\eta_5(\eta)}(p_{V_1 \oplus V_2, Y})\right\}.
\end{aligned}
$$

Note that $\epsilon_1 \cup \epsilon_{21} \cup \epsilon_{22} \cup \epsilon_4^c \cup \epsilon_5$ contains the error event and hence $P(\epsilon_1) + P(\epsilon_{11}^c \cap \epsilon_{21}) + P(\epsilon_{12}^c \cap \epsilon_{22}) + P((\epsilon_1 \cup \epsilon_{21} \cup \epsilon_{22})^c \cap \epsilon_4^c) + P(\epsilon_5)$ is an upper bound on the probability of error. In the sequel, we provide an upper bound on each of the above terms.

Lemma 2.3.1 guarantees the existence of $N_2(\eta) \in \mathbb{N}$ such that $P(\epsilon_1) \leq \frac{\eta}{8}$ for all $n \geq N_1(\eta)$. Lemma A.0.2(3) in appendix A implies the existence of $N_3(\eta) \in \mathbb{N}$ such that for all $n \geq N_3(\eta)$

$$P(\epsilon_{1j}^c \cap \epsilon_{2j}) \leq \exp\left\{-n \log \pi \left(\frac{k_j}{n} - \left(1 - \frac{H(V_j|S_j)}{\log \pi} + \frac{3\eta_4(\eta)}{2 \log \pi}\right)\right)\right\}.$$

Substituting the lower bound in (6.9) for $\frac{k_j}{n}$, we obtain

$$P(\epsilon_{1j}^c \cap \epsilon_{2j}) \leq \exp\left\{-n \left(\eta_1(\eta) - \frac{3\eta_4(\eta)}{2}\right)\right\}. \tag{6.14}$$

for all $n \geq \max\{N_1(\eta), N_3(\eta)\}$. We now derive an upper bound on $P((\epsilon_1 \cup \epsilon_{21} \cup \epsilon_{22})^c \cap \epsilon_4^c)$. The encoding rule ensures $(\epsilon_1 \cup \epsilon_{21} \cup \epsilon_{22})^c \subseteq (\epsilon_1 \cup \epsilon_2)^c$, where

$$\epsilon_2 = \bigcup_{j=1}^{2} \left\{\left(S_j^n, V_j^n(A_j^{k_j}, M_j^{l_j})\right) \notin T_{\eta_4(\eta)}(S_j, V_j)\right\},$$

and $V_j^n(A_j^{k_j}, M_j^{l_j})$ denotes codeword in $L_j(M_j^{l_j}, S_j^n)$ chosen by encoder j. Our first step is to provide an upper bound on $P((\epsilon_1 \cup \epsilon_2)^c \cap \epsilon_3)$ for sufficiently large n, where

$$\epsilon_3 = \left\{\left(S_j^n, V_j^n(A_j^{k_j}, M_j^{l_j}) : j = 1, 2\right) \notin T_{\frac{\eta_5(\eta)}{2}}(S_1, V_1, S_2, V_2)\right\}.$$

In the second step, we employ the result of conditional frequency typicality to provide an upper bound on $P((\epsilon_1 \cup \epsilon_2 \cup \epsilon_3)^c \cap \epsilon_4^c)$.

As an astute reader might have guessed, the proof of first step will employ the Markov chain $V_1 - S_1 - S_2 - V_2$. The proof is non-trivial because of statistical dependence of the codebooks. We begin with the definition

$$\Theta(\boldsymbol{s}^n) := \left\{\boldsymbol{v}^n \in \mathcal{V}^n : (s_j^n, v_j^n) \in T_{\eta_4(\eta)}(S_j, V_j) : j = 1, 2, (\boldsymbol{s}^n, \boldsymbol{v}^n) \notin T_{\frac{\eta_5(\eta)}{2}}(\boldsymbol{S}, \boldsymbol{V})\right\}$$

for any $\boldsymbol{s}^n \in \mathcal{S}^n$. Observe that,

$$\begin{aligned}
P((\epsilon_1 \cup \epsilon_2)^c \cap \epsilon_3) &= \sum_{\boldsymbol{s}^n \in T_{\frac{\eta_4(\eta)}{2}}(\boldsymbol{S})} \sum_{\boldsymbol{v}^n \in \Theta(\boldsymbol{s}^n)} P(\boldsymbol{S}^n = \boldsymbol{s}^n, V_j^n(A_j^{k_j}, M_j^{l_j}) = v_j^n : j = 1,2) \\
&= \sum_{\boldsymbol{s}^n \in T_{\frac{\eta_4(\eta)}{2}}(\boldsymbol{S})} \sum_{\boldsymbol{v}^n \in \Theta(\boldsymbol{s}^n)} P\left(\bigcup_{a_1^{k_1} \in \mathcal{V}_1^{k_1}} \bigcup_{a_2^{k_2} \in \mathcal{V}_2^{k_2}} \left\{ \boldsymbol{S}^n = \boldsymbol{s}^n, \begin{smallmatrix} V_j^n(A_j^{k_j}, M_j^{l_j}) = v_j^n : j=1,2, \\ V_j^n(a_j^{k_j}, M_j^{l_j}) = v_j^n : j=1,2 \end{smallmatrix} \right\} \right) \\
&\leq \sum_{\boldsymbol{s}^n \in T_{\frac{\eta_4(\eta)}{2}}(\boldsymbol{S})} \sum_{\boldsymbol{v}^n \in \Theta(\boldsymbol{s}^n)} \sum_{a_1^{k_1} \in \mathcal{V}_1^{k_1}} \sum_{a_2^{k_2} \in \mathcal{V}_2^{k_2}} P\left(\left\{ \boldsymbol{S}^n = \boldsymbol{s}^n, \begin{smallmatrix} V_1^n(a_1^{k_1}, M_1^{l_1}) = v_1^n, \\ V_2^n(a_2^{k_2}, M_2^{l_2}) = v_2^n \end{smallmatrix} \right\} \right) \\
&= \sum_{\boldsymbol{s}^n \in T_{\frac{\eta_4(\eta)}{2}}(\boldsymbol{S})} \sum_{\boldsymbol{v}^n \in \Theta(\boldsymbol{s}^n)} \sum_{a_1^{k_1} \in \mathcal{V}_1^{k_1}} \sum_{a_2^{k_2} \in \mathcal{V}_2^{k_2}} P\left(\boldsymbol{S}^n = \boldsymbol{s}^n\right) P\left(\begin{smallmatrix} V_1^n(a_1^{k_1}, M_1^{l_1}) = v_1^n, \\ V_2^n(a_2^{k_2}, M_2^{l_2}) = v_2^n \end{smallmatrix} \right) \qquad (6.15) \\
&= \sum_{\boldsymbol{s}^n \in T_{\frac{\eta_4(\eta)}{2}}(\boldsymbol{S})} \sum_{\boldsymbol{v}^n \in \Theta(\boldsymbol{s}^n)} P\left(\boldsymbol{S}^n = \boldsymbol{s}^n\right) \frac{1}{\pi^{n-k_1}} \frac{1}{\pi^{n-k_2}} \qquad (6.16)
\end{aligned}$$

where $V_j^n(A_j^{k_j}, M_j^{l_j})$ is defined as the random codeword chosen by the encoder, (6.15) follows from independence of random variables $(M^l, G_I, G_{O/I}, B_1^n, B_2^n)$ that characterize $V_j^n(a_j^{k_j}, M_j^{l_j})$ and $\boldsymbol{S}^n$. We now employ the upper bound on k_j in (6.9) to substitute for $\frac{1}{\pi^{n-k_j}}$. For $n \geq N_1(\eta)$, we have $k_j \leq n - \frac{H(V_j|S_j)}{\log \pi} + \frac{2\eta_1(\eta)}{\log \pi}$ and hence

$$\frac{1}{\pi^{n-k_j}} \leq \exp\left\{-n\left(H(V_j|S_j) - 2\eta_1(\eta)\right)\right\}. \qquad (6.17)$$

Furthermore, by Lemma 2.2.3, for every $\boldsymbol{s}^n \in T_{\frac{\eta_4(\eta)}{2}}(\boldsymbol{S})$ and $\boldsymbol{v}^n \in \Theta(\boldsymbol{s}^n)$,

$$\exp\left\{-n\left(H(V_j|S_j) - 2\eta_4(\eta)\right)\right\} \leq p_{V_j^n|S_j^n}(v_j^n|s_j^n) = p_{V_j^n|\boldsymbol{S}^n}(v_j^n|\boldsymbol{s}^n) = p_{V_j^n|\boldsymbol{S}^n V_{\not j}^n}(v_j^n|\boldsymbol{s}^n, v_{\not j}^n), \qquad (6.18)$$

where the last equalities is a consequence of Markov chain $V_1 - S_1 - S_2 - V_2$. Substituting the upper bounds in (6.17) and (6.18) for $\frac{1}{\pi^{n-k_j}}$ in (6.16), we obtain

$$\begin{aligned}
P((\epsilon_1 \cup \epsilon_2)^c \cap \epsilon_3) &\leq \exp\left\{n(4\eta_1(\eta) + 4\eta_4(\eta))\right\} \cdot \sum_{\boldsymbol{s}^n \in T_{\frac{\eta_4(\eta)}{2}}(\boldsymbol{S})} \sum_{\boldsymbol{v}^n \in \Theta(\boldsymbol{s}^n)} p_{\boldsymbol{S}^n \boldsymbol{V}^n}(\boldsymbol{s}^n, \boldsymbol{v}^n) \\
&\leq \exp\left\{n(4\eta_1(\eta) + 4\eta_4(\eta))\right\} \cdot \sum_{(\boldsymbol{s}^n, \boldsymbol{v}^n) \notin T_{\eta_5(\eta)}(\boldsymbol{S}, \boldsymbol{V})} p_{\boldsymbol{S}^n \boldsymbol{V}^n}(\boldsymbol{s}^n, \boldsymbol{v}^n) \qquad (6.19)
\end{aligned}$$

for all $n \geq N_1(\eta)$. We now employ the exponential upper bound provided in Lemma 2.3.1. In particular, Lemma

2.3.1 guarantees the existence of $N_4(\eta) \in \mathbb{N}$ such that for every $n \geq N_4(\eta)$,

$$\sum_{\substack{(\boldsymbol{s}^n,\boldsymbol{v}^n)\in \\ T_{\eta_5(\eta)}(\boldsymbol{S},\boldsymbol{V})}} p_{\boldsymbol{S}^n\boldsymbol{V}^n}(\boldsymbol{s}^n,\boldsymbol{v}^n) \leq \exp\left\{-n\lambda\eta_5^2(\eta)\right\}, \text{ where } \lambda := \frac{\min_{(\boldsymbol{s},\boldsymbol{v})\in\mathcal{S}\times\mathcal{V}}\left\{p^2_{\boldsymbol{S}\boldsymbol{V}}(\boldsymbol{s},\boldsymbol{v}) : p_{\boldsymbol{S}\boldsymbol{V}}(\boldsymbol{s},\boldsymbol{v})>0\right\}}{\left(\log|\mathcal{S}||\mathcal{V}|\right)^2}. \tag{6.20}$$

Substituting (6.20) in (6.19), we conclude

$$P((\epsilon_1\cup\epsilon_2)^c\cap\epsilon_3) \leq \exp\left\{-n\left(\lambda\eta_5^2(\eta) - 4\eta_1(\eta) - 4\eta_4(\eta)\right)\right\} \tag{6.21}$$

for every $n \geq \max\{N_1(\eta), N_4(\eta)\}$. This gets us to the second step. We begin with two observations. Firstly, note that $V(a_1^{k_1}0^{k_+}\oplus a_2^{k_2}, m_1^{l_1}m_2^{l_2}) = V_1(a_1^{k_1}, m_1^{l_1})\oplus V_2(a_2^{k_2}, m_2^{l_2})$. This follows from the definition of the codewords involved. Secondly,

$$P\left(\begin{matrix}V(A_1^{k_1}0^{k_+}\oplus A_2^{k_2}, M_1^{l_1}M_2^{l_2})=v^n, \\ X_j^n(M_j^{l_j}, S_j^n)=x_j^n : j=1,2, Y^n=y^n\end{matrix}\middle| \begin{matrix}V_j^n(A_j^{k_j}, M_j^{l_j})=v_j^n, \\ :j=1,2, \boldsymbol{S}^n=\boldsymbol{s}^n\end{matrix}\right) = P\left(\begin{matrix}V_1(A_1^{k_1}, M_1^{l_1})\oplus V_2(A_2^{k_2}, M_2^{l_2})=v^n, \\ X_j^n(M_j^{l_j}, S_j^n)=x_j^n : j=1,2, Y^n=y^n\end{matrix}\middle| \begin{matrix}V_j^n(A_j^{k_j}, M_j^{l_j})=v_j^n, \\ :j=1,2, \boldsymbol{S}^n=\boldsymbol{s}^n\end{matrix}\right)$$

$$= \prod_{t=1}^n \left[p_{V_1\oplus V_2|V_1V_2}(v_t|v_{1t}, v_{2t})\left(\prod_{j=1}^2 p_{X_j|V_jS_j}(x_{jt}|v_{jt}, s_{jt})\right) W_{Y|\boldsymbol{X}\boldsymbol{S}}(y_t|\boldsymbol{x}_t, \boldsymbol{s}_t)\right] \tag{6.22}$$

$$= \prod_{t=1}^n P(V_1\oplus V_2 = v_t, \boldsymbol{X}=\boldsymbol{x}_t, Y_t = y_t|\boldsymbol{S}_t=\boldsymbol{s}_t, \boldsymbol{V}_t=\boldsymbol{v}_t), \tag{6.23}$$

where we have employed 1) encoding rule and Markov chains $\mathbf{U}-(\boldsymbol{X},\boldsymbol{S})-Y$ in arriving at (6.22) and 2) the identity $p_{X_j|\boldsymbol{S}\mathbf{U}X_{\not j}} = p_{X_j|\boldsymbol{S}\mathbf{U}} = p_{X_j|S_jU_j}$ for any distinct elements $j, \not j \in \{1,2\}$ in arriving at (6.23). Since

$$P((\epsilon_1\cup\epsilon_2\cup\epsilon_3)^c\cap\epsilon_4^c) \leq P\left((\epsilon_1\cup\epsilon_2\cup\epsilon_3)^c\cap\left\{(V(\substack{A_1^{k_1}0^{k_+}\\ \oplus A_2^{k_2}}, M_1^{l_1}M_2^{l_2}), Y^n)\notin T_{\eta_5(\eta)}(V_1\oplus V_2, Y)\right\}\right)$$

$$\leq P\left((S_j^n, V_j^n(A_j^{k_j}, M_j^{l_j}) : j=1,2)\in T_{\frac{\eta_5(\eta)}{2}}(\boldsymbol{S},\boldsymbol{V}), (V(\substack{A_1^{k_1}0^{k_+}\\ \oplus A_2^{k_2}}, M_1^{l_1}M_2^{l_2}), Y^n)\notin T_{\eta_5(\eta)}(V_1\oplus V_2, Y)\right),$$

and the above two observations imply that $(V(A_1^{k_1}0^{k_+}\oplus A_2^{k_2}, M_1^{l_1}M_2^{l_2}), \boldsymbol{X}^n, Y^n)$ is distributed according to $\prod_{t=1}^n P(V_1\oplus V_2 = v_t, \boldsymbol{X}=\boldsymbol{x}_t, Y_t=y_t|\boldsymbol{S}_t=\boldsymbol{s}_t, \boldsymbol{V}_t=\boldsymbol{v}_t)$. Lemma 2.4.1 guarantees the existence of $N_5(\eta)\in\mathbb{N}$, such that for all $n\geq N_5(\eta)$, the term on the right hand side of (6.24) is bounded from above by $\frac{\eta}{8}$. Therefore, for all $n\geq N_5(\eta)$

$$P((\epsilon_1\cup\epsilon_2\cup\epsilon_3)^c\cap\epsilon_4^c) \leq \frac{\eta}{8}. \tag{6.24}$$

It remains to provide an upper bound on $P((\epsilon_1\cup\epsilon_{21}\cup\epsilon_{22}\cup\epsilon_4^c)^c\cap\epsilon_5)$. In appendix M, we prove the existence of $N_6(\eta)\in\mathbb{N}$ such that $P(\epsilon_5)\leq\exp\{-n(3\eta_5(\eta)-\eta_2(\eta)-\eta_3(\eta))\}$ for all $n\geq\max\{N_1(\eta), N_6(\eta)\}$. The informed reader will recognize that deriving an upper bound on $P(\epsilon_5)$ will involve proving statistical independence of the pair $(C_j(M_j^{l_j}) : j=1,2)$ of cosets corresponding to the legitimate message pair M_j^l and any codeword $V^n(\hat{a}^k, \hat{m}^l)$

corresponding to a competing message pair $\hat{m}^l \neq M^l$. This is considerably simple for a coding technique based on classical unstructured codes wherein codebooks and codewords in every codebook are independent. The coding technique proposed herein involves correlated codebooks - the first k_1 rows of $G_{I_j} : j = 1, 2$ are identical[6] - and codewords in each codebook are correlated.

To conclude, we put together the upper bounds derived on the probability of events that comprise the error event. For $n \geq N_2(\eta)$, $P(\epsilon_1) \leq \frac{\eta}{8}$. In (6.14), we proved $P(\epsilon_{1j}^c \cap \epsilon_{2j}) \leq \exp\left\{-n\left(\eta_1(\eta) - \frac{3\eta_4(\eta)}{2}\right)\right\}$ for all $n \geq N_3(\eta)$. Combining (6.21) and (6.24), we have

$$P((\epsilon_1 \cup \epsilon_2)^c \cap \epsilon_4^c) \leq \exp\left\{-n\left(\lambda\eta_5^2(\eta) - 4\eta_1(\eta) - 4\eta_4(\eta)\right)\right\} + \frac{\eta}{8}$$

for all $n \geq \max\{N_1(\eta), N_4(\eta), N_5(\eta)\}$. And finally $P(\epsilon_5) \leq \exp\{-n\left(\eta_2(\eta) + \eta_3(\eta) - 3\eta_5(\eta)\right)\}$ for all $n \geq \max\{N_1(\eta), N_6(\eta)\}$ follows from (M.10). By choosing

$$\eta_2(\eta) = \eta_3(\eta) = \frac{\eta}{16}, \eta_5(\eta) = \frac{\eta}{48}, \eta_1(\eta) = \min\left\{\frac{\eta}{16}, \frac{\lambda\eta_5^2(\eta)}{10}\right\} \text{ and } \eta_4(\eta) = \frac{\eta_1(\eta)}{4} \tag{6.25}$$

it can be verified that for $n \geq \overline{N}(\eta) := \max\{N_i(\eta) : i \in [6]\}$,

- $2\eta_1(\eta) + 2\eta_3(\eta) < \frac{\eta}{2}$ and thus $\frac{l_2 \log \pi}{n} \geq R_2 - \frac{\eta}{2}$ from (6.12),
- $\eta_2(\eta) < \frac{\eta}{2}$ and thus $\frac{l_1 \log \pi}{n} > R_1 - \frac{\eta}{2}$ from (6.10),
- $\eta_1(\eta) - \frac{3\eta_4(\eta)}{2} = \frac{5\eta_1(\eta)}{8}$ and thus $P(\epsilon_{1j}^c \cap \epsilon_{2j}) \leq \exp\left\{-n\left(\frac{5\eta_1(\eta)}{8}\right)\right\}$,
- $\lambda\eta_5^2(\eta) - 4\eta_1(\eta) - 4\eta_4(\eta) \geq \frac{\lambda\eta_5^2(\eta)}{2}$ and thus $P((\epsilon_1 \cup \epsilon_2)^c \cap \epsilon_4^c) \leq \exp\left\{-n\left(\frac{\lambda\eta_5^2(\eta)}{2}\right)\right\} + \frac{\eta}{8}$, and
- $\eta_2(\eta) + \eta_3(\eta) - 3\eta_5(\eta) = \frac{\eta}{16}$ and therefore $P(\epsilon_5) \leq \exp\left\{-n\left(\frac{\eta}{16}\right)\right\}$.

For $n \geq \overline{N}(\eta)$, $P(\epsilon_1) + P(\epsilon_{11}^c \cap \epsilon_{21}) + P(\epsilon_{12}^c \cap \epsilon_{22}) + P((\epsilon_1 \cup \epsilon_{21} \cup \epsilon_{22})^c \cap \epsilon_4^c) + P(\epsilon_5) \leq \frac{\eta}{4} + 3\exp\left\{-n(\frac{5\eta_1}{8})\right\}$. Thus for $n \geq N(\eta) := \max\left\{\overline{N}(\eta), \frac{1}{\eta_1(\eta)} \log\lceil\frac{4}{\eta}\rceil\right\}$, the error event has probability at most η. ∎

We conclude this section with two remarks.

Remark 6.2.3 *For BDD-MAC described in section 6.2.2, $\beta_f(\boldsymbol{\tau}) = \mathbb{C}(\boldsymbol{\tau})$. Indeed, the test channel $p_{\boldsymbol{VSXY}} \in \mathbb{D}_f(\boldsymbol{\tau})$ defined as $p_{\boldsymbol{VSX}} = \prod_{j=1}^2 p_{V_j S_j X_j}$ where V_j takes values over $\mathcal{V}_j = \{0, 1\}$ with*

$$p_{V_j, X_j|S_j}(x_j \oplus_2 s_j, x_j|s_j) = \begin{cases} 1 - \tau & \text{if } x_j = 0 \\ \tau & \text{otherwise} \end{cases}$$

for each $j = 1, 2$ and $s_j \in \{0, 1\}$ achieves $\mathbb{C}(\boldsymbol{\tau}) = \{(R_1, R_2) : R_1 + R_2 \leq h_b(\tau)\}$.

[6]If $H(V_1|S_1) = H(V_2|S_2)$, users 1 and 2 share the same generator matrix G_I. Indeed, channel codes of users' 1 and 2 are partitioned into cosets of the same linear code.

We have thus presented a coding technique based on decoding the sum of codewords chosen by the encoders and analyzed the same to derive an achievable rate region for an arbitrary MAC-DSTx. One might attempt a generalization of PZ-technique along the lines of modulo lattice transformation proposed by Haim, Kochman and Erez [70]. The rate region proposed herein subsumes that achievable using modulo-lattice transformation using test channels identified through the virtual channel in a natural way.

6.2.3 Examples

A key element of the coding framework proposed herein lies in characterizing achievable rate regions for arbitrary test channels, i.e., test channels that are not restricted to be uniform or additive in nature using structured codes. Our first example (example 6.2.4) illustrates that such test channels indeed optimize the achievable rate region for certain MAC-DSTx. Through a simple modification of BDD-MAC, we illustrate the same.

In several practical scenarios, the channel[7] is not perfectly additive but nearly so. We therefore randomly perturb BDD-MAC and study the efficacy of linear codes for such a channel in example 6.2.5. We do not expect the resulting test channel to be either additive or uniform. Yet, our results indicate that by employing nested coset codes and exploiting the algebraic structure yields larger achievable rate regions. In [47], we have presented results for a few more channels that have been obtained by a random perturbation of the BDD-MAC.

A few remarks on our study of the following examples are in order. The examples needing to be non-additive lends it considerably hard to provide analytical upper bounds for the rate region achievable using unstructured codes.[8] We therefore resort to computation. It can be noted that the problem of computing the sum rate bound achievable using unstructured codes is a non-convex optimization problem. The only approach is direct enumeration, i.e., sampling the probability matrix of the auxiliary random variables.[9] Sampling the probability matrix with any reasonable step size beyond the auxiliary alphabets of size 2 is infeasible with currently available computation resources. The sum rate bound for the unstructured coding technique projected below is therefore obtained through computation involving binary auxiliary alphabet sets followed by convexification (timesharing between different cost). The resulting space of probability distributions that respect the cost constraints was sampled with a step size of 0.015 in each dimension. The resulting bound on the sum rate achievable using unstructured codes (without time sharing) is marked with blue crosses (denoted α in the legend) in the plots. The resulting upper bound is obtained as an upper convex envelope. Similarly, sum rate achievable using nested coset codes is marked with red circles (denoted β in the legend) in the plots.

For examples 6.2.4 and 6.2.5, we assume the alphabet sets to be binary $\mathcal{S}_j = \mathcal{X}_j = \{0, 1\}$, $j = 1, 2$, (ii) uniform and independent states, i.e., $W_{\boldsymbol{S}}(\boldsymbol{s}) = \frac{1}{4}$ for all $\boldsymbol{s} \in \mathcal{S}$, (iii) a Hamming cost function $\kappa_j(1, s_j) = 1$ and $\kappa_j(0, s_j) = 0$

[7] Usually the channel is tracked through pilot or training waveforms and the presence feedback link.
[8] We recognize that the analytical upper bound derived in [15] is a key element of the findings therein.
[9] This holds even for the case of multiple access without states for which a computable characterization of the capacity region is known.

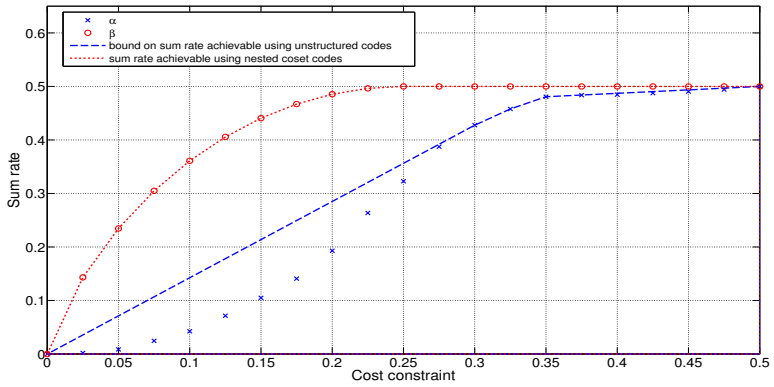

Figure 6.2: Bounds on sum rate for example 6.2.4

for any $s_j \in \mathcal{S}_j$, $j = 1, 2$. We compute the sum rate achievable using unstructured codes and sum rate achievable using nested coset codes. These are plotted in figures 6.2 and 6.3.

Example 6.2.4 *Let $Y = (X_1 \vee S_1) \oplus (X_2 \vee S_2)$, where $\vee$ denotes logical OR operator. Having studied the BDD-MAC it is natural to conjecture that the test channel that optimizes the sum rate achievable using linear codes to be $p_{U_j X_j|S_j}(0,0|0) = 1 - 2\tau, p_{U_j X_j|S_j}(1,1|0) = 2\tau, p_{U_j X_j|S_j}(1,0|1) = 1$, for $j = 1, 2$ when the cost constraint $\tau \in [0, \frac{1}{4}]$. Indeed, our numerical computation asserts this. In other words, the sum rate achievable using linear codes for a cost $\tau \in (0, \frac{1}{4})$ is $\frac{h_b(2\tau)}{2}$ and 0.5 for $\tau \in [0.25, 0.5]$. We highlight significant gains achievable using nested coset codes.*

A preliminary look at this channel may lead the reader to conclude that PZ-technique appropriately modified can achieve the same sum rate as that achievable using nested coset codes, since the above test channel is additive, i.e., $U_j = S_j \oplus X_j$ for $j = 1, 2$ and $Y = U_1 \oplus U_2$. However, a careful analysis will reveal the significance of the coding framework proposed herein. The induced pmf on U_j, $p_{U_j}(1) = \frac{1}{2} + 2\tau$ for $\tau \in (0, \frac{1}{4})$ is not uniform, and the PZ-technique of choosing a codeword in the indexed bin with an average Hamming distance of τ does not yield the sum rate guaranteed by nested coset codes. Nesting of codes enables achieving non-uniform distributions that are necessary as exemplified herein.

Example 6.2.5 *The channel transition matrix is given in table 6.1. 1) An upper bound on sum rate achievable using unstructured codes and 2) sum rate achievable using structured are plotted in figure 6.3. This channel is obtained by randomly perturbing the BDD-MAC. In the space of channel transition probability matrices, this channel is in a neighborhood of the BDD-MAC. Since the rate regions are continuous functions over this space of channels, the coding technique proposed herein outperforms unstructured coding technique in this neighborhood. This example validates the same. As in the previous example, we note that the optimizing distribution of the auxiliary random variables is non-uniform for certain cost values. Furthermore, note that $\beta_f(\boldsymbol{\tau})$ does not contain $\alpha(\boldsymbol{\tau})$ and therefore it helps to incorporate both unstructured and structured coding techniques as will be studied in the following section.*

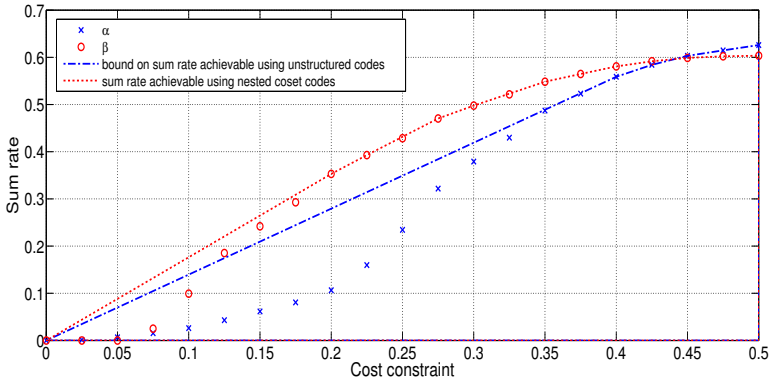

Figure 6.3: Bounds on sum rate for example 6.2.5

$S_2X_2S_1X_1$	$w_{Y\|\boldsymbol{SX}}(0\|\cdot)$	$S_2X_2S_1X_1$	$w_{Y\|\boldsymbol{SX}}(0\|\cdot)$	$S_2X_2S_1X_1$	$w_{Y\|\boldsymbol{SX}}(0\|\cdot)$	$S_2X_2S_1X_1$	$w_{Y\|\boldsymbol{SX}}(0\|\cdot)$
0000	0.92	1000	0.07	0100	0.10	1100	0.88
0001	0.08	1001	0.92	0101	0.92	1101	0.08
0010	0.06	1010	0.96	0110	0.95	1110	0.11
0011	0.94	1011	0.10	0111	0.06	1111	0.91

Table 6.1: Channel transition matrix Example 6.2.5

$U_1S_1X_1$	$p_{U_1S_1X_1}$	$U_1S_1X_1$	$p_{U_1S_1X_1}$	$U_2S_2X_2$	$p_{U_2S_2X_2}$	$U_2S_2X_2$	$p_{U_2S_2X_2}$
000	0.1472	101	0.3528	000	0.1472	101	0.3528
011	0.50			011	0.50		

Table 6.2: Test channel for example 6.2.6 for which nested coset code over $\mathcal{F}_3$ performs better than unstructured code

Example 6.2.6 *Consider the channel $Y = (S_1 \oplus X_1) \vee (S_2 \oplus X_2)$. Observe that the information available at the encoders is fused through a logical OR operation by the channel. Moreover, $U_1, U_2 - U_1 \oplus_3 U_2 - U_1 \vee U_2$ is a Markov chain and hence, although channel input, state and output alphabets are binary, we expect that for certain choice of auxiliary distributions, the sum rate achievable using codes over $\mathcal{F}_3$ is larger than that achievable using unstructured codes. Through an exhaustive search, we have identified such distributions, an example of which is given in table 6.2.*

For the above distribution, the rate achievable using nested coset codes over $\mathcal{F}_3$ is 0.0017, while that achievable using unstructured code is negative. For an appropriate choice of cost function, the above might be the optimizing distribution for the unstructured coding scheme thus resulting in larger sum rate using nested coset codes over $\mathcal{F}_3$. We do not as of yet have a precise analytical characterization of such a cost function[10] and we are in pursuit of the same. Nevertheless, the above lends credence to the use of nested coset codes for arbitrary channels.

6.3 Stage II: Combining unstructured and structured coding techniques

In this section, we put together the techniques of unstructured and structured random coding to derive a larger achievable rate region for a general MAC-DSTx. Our approach is similar to that proposed by Ahlswede and Han [48, Section VI] for the problem of reconstructing mod-2 sum of distributed binary sources. We begin with a characterization of valid test channels.

Definition 6.3.1 *Let $\mathbb{D}_{sf}(\boldsymbol{\tau}) \subseteq \mathbb{D}(\boldsymbol{\tau})$ be the collection of distributions $p_{\mathbf{UVSXY}}$ on $(\mathcal{U} \times \mathcal{V})^2 \times \mathcal{S} \times \mathcal{X} \times \mathcal{Y}$ where $\mathcal{U}$ is a finite set and $\mathcal{V}$ is a finite field. For $p_{\mathbf{UVSXY}} \in \mathbb{D}_{sf}(\boldsymbol{\tau})$, let $\mathcal{R}(p_{\mathbf{UVXSY}})$ be defined as*

$$\left\{ \begin{array}{ll} (R_1, R_2): & 0 \leq R_1 \leq I(U_1; U_2 Y) - I(U_1; S_1) + \min \left\{ \begin{smallmatrix} H(V_1|U_1,S_1), \\ H(V_2|U_2,S_2) \end{smallmatrix} \right\} - H(V_1 \oplus V_2 | U_1, U_2, Y) \\ & 0 \leq R_2 \leq I(U_2; U_1 Y) - I(U_2; S_2) + \min \left\{ \begin{smallmatrix} H(V_1|U_1,S_1), \\ H(V_2|U_2,S_2) \end{smallmatrix} \right\} - H(V_1 \oplus V_2 | U_1, U_2, Y) \\ & R_1 + R_2 \leq I(U_1 U_2; Y) + I(U_1; U_2) - \sum_{j=1}^{2} I(U_j; S_j) + \min \left\{ \begin{smallmatrix} H(V_1|U_1,S_1), \\ H(V_2|U_2,S_2) \end{smallmatrix} \right\} \\ & \qquad - H(V_1 \oplus V_2 | U_1, U_2, Y) \end{array} \right\},$$

[10] Such a characterization of cost function is available for point-to-point channels with state available at both encoder and decoder [2], [71], [72].

where $\oplus$ is addition in $\mathcal{V}$. Let

$$\mathcal{R}(\boldsymbol{\tau}) := cocl\left(\bigcup_{p_{\mathbf{UVXS}Y}\in\mathbb{D}_{sf}(\boldsymbol{\tau})}\mathcal{R}(p_{\mathbf{UVXS}Y})\right) \tag{6.26}$$

Theorem 6.3.2 $\mathcal{R}(\boldsymbol{\tau}) \subseteq \mathcal{C}(\boldsymbol{\tau})$. □

Remark 6.3.3 *$\alpha(\boldsymbol{\tau}) \subseteq \mathcal{R}(\boldsymbol{\tau})$ and moreover $\mathcal{R}(\boldsymbol{\tau})$ is the largest known achievable rate region for the general MAC-DSTx.*

Before we provide an outline of the proof, we briefly state the coding technique. Each user builds an unstructured code over $\mathcal{U}$ and a nested coset code over $\mathcal{V}$. The nested linear codes share a common inner (sparser) code. The unstructured code is partitioned into bins (Gel'fand-Pinsker binning). Each user's message is split into two parts - one indexing a bin in the unstructured code and one indexing a coset in the nested linear code. Encoder j picks a pair of codewords, say (U_j^n, V_j^n) jointly typical with the observed state sequence from the indexed pair of bins and transmits an input vector generated according to $p_{X_j|V_jU_jS_j}$. Having received Y^n, decoder looks for all triples $(U_1^n, U_2^n, V_1^n \oplus V_2^n)$ of codewords in the corresponding codebooks that is jointly typical with Y^n according to $p_{\mathbf{U},V_1\oplus V_2,Y}$ and declares the quadruple of bin indices as the decoded message. Achievability is proved by providing an upper bound on the probability of error by averaging over the ensemble of codes. We now provide an outline of the proof.

Proof: Achievability of $\mathcal{R}(\boldsymbol{\tau})$ is proved by gluing together unstructured and structured coding techniques. Each encoder splits it's message M_j into two parts $M_{j,1}$ and $M_j^{l_j}$. $M_{j,1}$ is communicated to the decoder using an unstructured random code built over $\mathcal{U}^n$. $M_j^{l_j}$ is communicated to the decoder using a nested coset code identical to that proposed in proof of theorem 6.2.2. With regard to nested coset codes, we employ the notation proposed in the proof of theorem 6.2.2 and do not restate the same.

Encoder j is provided a codebook built over $\mathcal{U}^n$ that contains $2^{n\bar{R}_j}$ bins each with 2^{nB_j} codewords. For $1 \leq b_j \leq 2^{nB_j}$, let $u_j(r_j, b_j)$ denote a generic codeword in bin r_j $(1 \leq r_j \leq 2^{n\bar{R}_j})$. Encoder j is also provided with the nested coset code $\lambda_{O_j/I}$. Without loss of generality, we assume $M_j^{l_j} \in \mathcal{V}^{l_j}$. Encoder j observes state sequence S_j^n and declares error if $S_j^n \notin T_{\frac{\delta}{8}}(W_{S_j})$. Otherwise it looks for a pair $(u_j^n(M_{j,1}, b_j), v^n(a^k, M_j^{l_j})) \in T_{\frac{\delta}{4}}(p_{U_jV_j|S_j^n}|S_j^n)$. If it finds at least one such pair, one of them say, $(u_j^n(M_{j,1}, b_j), v^n(a^k, M_j^{l_j}))$ is chosen uniformly at random and $e_j^n(M_j, S_j^n)$ is transmitted, where $e_j^n(M_j, S_j^n)$ is a function of $u_j^n(M_{j,1}, b_j), v^n(a^k, M_j^{l_j}), S_j^n$ that is determined upfront. Otherwise, an error is declared.

We now specify the decoding rule. The decoder receives Y^n and declares error if $Y^n \notin T_{\frac{\delta}{2}}(p_Y)$. Otherwise, decoding is performed in two stages. In the first stage it lists all codewords $(u_j^n(m_{j,1}, b_j) : j = 1, 2) \in T_\delta^n(p_{U_1,U_2|Y}|y^n)$. If it finds exactly one such pair, say $(u_j^n(m_{j,1}, b_j) : j = 1, 2)$, then the decoding proceeds to the next stage. Otherwise, an error is declared and decoding halts. In the second stage, the decoder looks for all codewords $v^n(a^k, \mathbf{m^l}) \in \lambda_O$ such that $(u_j^n(m_{j,1}, b_j) : j = 1, 2, v^n(a^k, \mathbf{m^l}), Y^n) \in T_\delta^n(p_{U,V_1\oplus V_2,Y})$. If it finds all such codewords in a unique bin, say corresponding to $\mathbf{m^l}$, then it declares $m_{j,1}, m_j^{l_j} : j = 1, 2$ as the decoded pair of messages. Otherwise, an error

is declared. We derive an upper bound on probability of error by averaging the error probability over the ensemble codes. A pmf is induced over the ensemble of codes by letting $U_j^n(r_j, b_j) : 1 \leq r_{j \leq 2^{n\bar{R}_j}}, 1 \leq b_j \leq 2^{nB_j}, j = 1, 2$ be mutually independent and distributed according to $\prod_{t=1}^n p_{U_j}$. The pmf induced on the ensemble of nested coset codes is identical to that in proof of theorem 6.2.2. Moreover, $(G_I, G_{O_j/I}, B_j^n : j = 1, 2)$ is independent of the unstructured random code on $\mathcal{U}^n$. Analyzing the error events, we obtain the following sufficient conditions for the average probability of error to decay exponentially.

$$\begin{aligned}
B_1 &\geq I(U_1; S_1) & B_2 &\geq I(U_2; S_2) \\
\bar{R}_1 + B_1 &\leq I(U_1; U_2 Y) & \bar{R}_2 + B_2 &\leq I(U_2; U_1 Y) \\
\frac{k}{n} &\geq 1 - H(V_1|U_1 S_1) & \tfrac{k}{n} &\geq 1 - H(V_2|U_2 S_2) \\
\sum_{j=1}^{2} \bar{R}_j + B_j &\leq I(\mathbf{U}; Y) + I(U_1; U_2) & \tfrac{l_1+l_2}{n} &\leq 1 - H(V_1 + V_2|\mathbf{U}Y).
\end{aligned}$$

For each $j = 1, 2$, substituting $R_j - \frac{l_j}{n}$ for $\bar{R}_j$ in the above bounds and eliminating $B_j, \frac{k}{n}, \frac{l_j}{n} : j = 1, 2$ using the technique of Fourier-Motzkin [26, Appendix D], $\mathcal{R}(\boldsymbol{\tau})$ is proved achievable. ■

Remark 6.3.4 *The above rate region is obtained by analyzing sequential typicality encoding and decoding, i.e., encoding and decoding of unstructured codes precedes that of structured codes. The informed reader will recognize that performing joint typicality encoding and decoding of unstructured and structured codes might enlarge the achievable rate region. While this might be true, Fourier-Motzkin elimination of the resulting bounds does not yield a compact description of the resulting achievable rate region. We therefore chose to present the above rate region.*

We conclude with an illustrative example.

Example 6.3.5 *For $j = 1, 2$, let $\mathcal{S}_j = \mathcal{X}_j = \mathcal{Y} = \{0, 1\}$. The channel transition is described as $W_{Y|\boldsymbol{X}\boldsymbol{S}}(y|\boldsymbol{x}, \boldsymbol{s}) = W^*_{Y|g(\boldsymbol{X},\boldsymbol{S})}(y|g(\boldsymbol{x}, \boldsymbol{s}))$, where $g(\boldsymbol{x}, \boldsymbol{s}) = [(s_2 \wedge \bar{x_2}) \wedge (\bar{s_1} \vee x_1)] \vee [(s_1 \wedge \bar{x_1}) \wedge (\bar{s_2} \vee x_2)]$ and $W^*_{Y|g(\boldsymbol{X},\boldsymbol{S})}(1|0) = 0.02$, $W^*_{Y|g(\boldsymbol{X},\boldsymbol{S})}(0|1) = 0.04$. The function $g(\cdot, \cdot)$ can be alternatively described as $g(\boldsymbol{X}, \boldsymbol{S}) = [S_1 \wedge (S_1 \oplus X_1)] \oplus [S_2 \wedge (S_2 \oplus X_2)]$.*

This channel is inspired by Blackwell's broadcast channel and in particular the coding technique proposed by Gel'fand [8].[11] *The bounds on the sum rate achievable with unstructured and nested coset codes are plotted in figure 6.4. The above plots unequivocally indicate $\mathcal{R}(\boldsymbol{\tau})$ to be strictly larger than $\alpha(\boldsymbol{\tau}) \cup \beta_f(\boldsymbol{\tau})$ and in particular either one of $\alpha(\boldsymbol{\tau})$, $\beta_f(\boldsymbol{\tau})$. It is therefore desirable to compute $\mathcal{R}(\boldsymbol{\tau})$, however the presence of two additional auxiliary random variables lends computation infeasible with current computational resources. We remark that the structure of this example enables us to argue the strict containment $\alpha(\boldsymbol{\tau}) \cup \beta_f(\boldsymbol{\tau}) \subsetneq \mathcal{R}(\boldsymbol{\tau})$ in spite of not being able to compute $\mathcal{R}(\boldsymbol{\tau})$.*

[11] *Analogous to the defect masking the written bits, here the states mask the corresponding channel.*

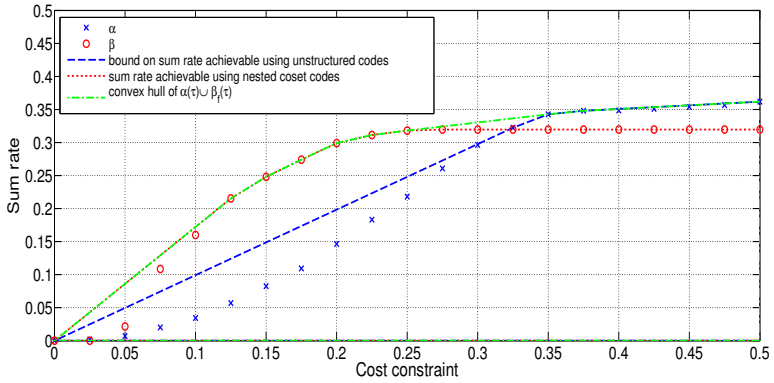

Figure 6.4: Bounds on sum rate for example 6.3.5

6.4 Stage III: Achievable rate region using codes over Abelian groups

Consider a quaternary doubly dirty MAC-DSTx (QDD-MAC), with $\mathcal{S}_j = \mathcal{X}_j = \mathcal{Y} = \{0, 1, 2, 3\}$, $j = 1, 2$. The state sequences are independent and uniformly distributed, i.e., $W_{\boldsymbol{S}}(\boldsymbol{s}) = \frac{1}{16}$ for all $\boldsymbol{s} \in \mathcal{S}$. The channel transition is described by the relation $Y = X_1 \oplus S_1 \oplus X_2 \oplus S_2$, where $\oplus$ denotes addition mod-4. All nonzero symbols have equal cost, i.e., $\kappa_j(x, s_j) = 1$ for all $x \in \{1, 2, 3\}$ and $\kappa_j(0, s_j) = 0$ for all $s_j \in \mathcal{S}_j$, $j = 1, 2$ and the input is subject to a symmetric cost constraint $\boldsymbol{\tau} = (\tau, \tau)$.

What would be the achievable rate region for QDD-MAC using unstructured codes? It is natural to guess the optimizing test channel to be

$$p_{X_j U_j | S_j}(x_j, x_j \oplus s_j | s_j) = \begin{cases} 1 - \tau & \text{for } x_j = 0 \\ \frac{\tau}{3} & \text{otherwise.} \end{cases} \tag{6.27}$$

In appendix D of [47], with the aid of numerical computation, we argue that this is indeed the case. The sum rate achievable using unstructured codes can be evaluated to be the upper convex envelope of the function $\alpha : [0, \frac{3}{4}] \to [0, \infty)$ defined as $\alpha(\tau) = \max\left\{-2\tau \log(\frac{\tau}{3}) - 2(1-\tau)\log(1-\tau)) - 2, 0\right\}$. Since 4 is a prime power, there exists a unique field $\mathcal{F}_4$ of cardinality 4. Do nested coset codes built over $\mathcal{F}_4$ achieve a larger sum rate?

We are unable to characterize the sum rate achievable using nested coset codes and the dimensionality of the space of probability distributions lends computation infeasible. We conjecture that the above test channel optimizes the sum rate achievable using nested coset codes. In any case, computing the sum rate achievable using nested coset codes for the above test channel is instructive. It can be verified that the sum rate achievable using the above test channel with nested coset codes is the upper convex envelope of the function $\beta_f : [0, \frac{3}{4}] \to [0, \infty)$ defined as $\beta_f(\tau) = \max\left\{-\tau \log(\frac{\tau}{3}) - (1-\tau)\log(1-\tau)) - \frac{1}{2}, 0\right\}$.

The sum rate achievable for the above test channel using unstructured and nested coset codes are plotted in figure 6.5. It is no surprise that nested coset codes perform poorly. The channel operation is *not* the field addition $\oplus_4$ in

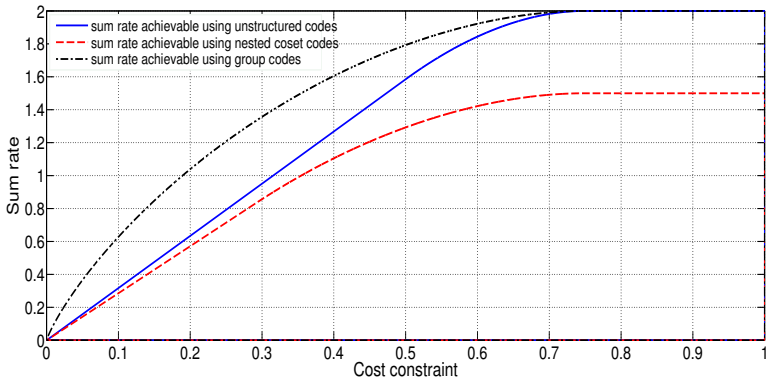

Figure 6.5: Sum rate achievable using unstructured, nested coset and Abelian group codes for test channel (6.27)

$\mathcal{F}_4$. Instead, $\oplus$ is the group addition[12] in the Abelian group $\mathbb{Z}_4$. This suggests that we build codes over Abelian groups that are closed under group addition and decode the group sum $\oplus$ of codewords.

Linear codes are kernels of field homomorphisms. This lends them the property of closure under field addition. We build *Abelian group codes* that are kernels of group homomorphisms. Abelian group codes are closed under group addition. As was proposed with nested coset codes, we employ bins of each user's code to be cosets of a common Abelian group code. The encoder chooses a codeword from the bin indexed by the message and the decoder attempts to localize the group sum of chosen codewords. The bins of each users' codebook is chosen such that the decoder can decode the pair of messages by identifying the group sum of transmitted codewords.

In the interest of brevity, we only describe the results and omit proofs. Recall that any Abelian group $\mathcal{U}$ can be decomposed as sum of $\mathbb{Z}_{p^r}-$cyclic groups, i.e.,

$$\mathcal{U} = \bigoplus_{i=1}^{I} \mathbb{Z}_{p_i^{r_i}}, \tag{6.28}$$

where p_i is a prime and r_i is a positive integer for each $i = 1, \cdots, I$. We therefore state our findings in two stages. The first stage, described in section 6.4.1 describes the coding technique and achievable rate region for a $\mathbb{Z}_{p^r}-$ group. This is extended to an arbitrary Abelian group in section 6.4.2

6.4.1 Achievable rate region for MAC-DSTx using group codes : The $\mathbb{Z}_{p^r}$-case

In the discussion following proof of theorem 3.5.1, we noted that if the auxiliary alphabet $\mathcal{U}$ is a field and the bins are constrained to be closed under field addition then with respect to a test channel $p_{U|S}$, the bins need to be of rate at least $\log|\mathcal{U}| - H(U|S)$. This enlargement of the bins was compensated by the ability to pack more bins. In particular, the rate of the composite code could be as large as $\log|\mathcal{U}| - H(U|Y)$ with respect to the induced distribution $p_{U|Y}$, and this enabled us to achieve the capacity of PTP-STx.

[12]We refer to group operation of an Abelian group as group addition.

If the auxiliary alphabet $\mathcal{U} = \mathbb{Z}_{p^r}$ is an Abelian group of order p^r, and the bins are restricted to be closed under group addition, then with respect to a test channel $p_{U|S}$, using the results of [73], the bins have to be of rate at least

$$\overline{I}^{\mathcal{U}}_{s.c}(U;S) = \max_{\theta=1}^{r} \left[r \log p - \frac{r}{\theta} H([U]_\theta | S) \right] = \max_{\theta=1}^{r} \frac{r}{\theta} I([U]_\theta; S), \tag{6.29}$$

where $\mathcal{H}_\theta$ is the sub-group $p^\theta \mathbb{Z}_{p^r}$ and $[U]_\theta := U \oplus \mathcal{H}_\theta$ is the random variable taking values from cosets of subgroup $\mathcal{H}_\theta$ of $\mathcal{U}$, denoted $\mathcal{H}_\theta \preceq \mathcal{U}$. We note that $\overline{I}^{\mathcal{U}}_{s.c}(U;S) \geq \log q - H(U|S) \geq I(U;S)$. The natural question to ask is whether we can pack sufficient number of bins to achieve capacity of PTP-STx. It turns out that if we constrain the composite code, i.e., the union of bins, to be a coset of a group code, then the rate of this union can be at most

$$\overline{I}^{\mathcal{U}}_{c.c}(U;Y) = \min_{\theta=0}^{r-1} \left[r \log p - \frac{r}{r-\theta} H(U|Y[U]_\theta) \right] = \min_{\theta=0}^{r-1} \frac{r}{r-\theta} I(U;Y|[U]_\theta).$$

with respect to the induced distribution $p_{U|Y}$. Since $\log |\mathcal{U}| - H(U|Y)$ corresponds to $\theta = 0$ in the above expression, $\overline{I}^{\mathcal{U}}_{c.c}(U;Y)$ is in general smaller than $\log |\mathcal{U}| - H(U|Y)$. Therefore, $\overline{I}^{\mathcal{U}}_{c.c}(U;Y) - \overline{I}^{\mathcal{U}}_{s.c}(U;S)$ is in general strictly smaller than the capacity of PTP-STx, implying the constraint of closure under group addition results in a rate penalty. This indicates that the use of group codes will in general result in rate penalties for multi-terminal communication problems.[13]

With the objective of increasing $\overline{I}^{\mathcal{U}}_{c.c}(U;Y)$ and therefore minimizing the rate penalty, we take a closer look at the coding technique proposed in section 6.2.2. While we exploited the property of bins being closed under field addition, we did not need the union of bins to be a coset. We therefore relax this and only require the bins to have an algebraic structure, i.e., a coset of a group code, but the composite code of each user is not required to be a coset of a group code. In other words, we employ union coset codes (UCC) (section 3.4.3) built over groups. While this relaxation does not yield gains in achievable rate for the field case, we do obtain larger achievable rates while coding over groups. In particular, the rate of the composite code, or the union of bins can be as large as $\log |\mathcal{U}| - H(U|Y)$ which is in general larger than $\overline{I}^{\mathcal{U}}_{c.c}(U;Y)$. Therefore, if we were to communicate over a PTP-STx $(\mathcal{S}, W_S, \mathcal{X}, \kappa, \mathcal{Y}, W_{Y|XS})$ using codes over an Abelian $\mathbb{Z}_{p^r}-$group $\mathcal{U} = \mathbb{Z}_{p^r}$ and we constrained the bins to be closed under group addition, then the test channel $p_{USXY} \in \overline{\mathbb{D}}(\tau)$ yields an achievable rate $\log |\mathcal{U}| - H(U|Y) - (\overline{I}^{\mathcal{U}}_{s.c}(U;S)) = \overline{H}^{\mathcal{U}}_{s.c}(U|S) - H(U|Y)$, where

$$\overline{H}^{\mathcal{U}}_{s.c}(U|S) = \log |\mathcal{U}| - \overline{I}^{\mathcal{U}}_{s.c}(U;S), \tag{6.30}$$

is defined as *source coding group entropy* of group $\mathcal{U} = \mathbb{Z}_{p^r}$ and $\overline{H}^{\mathcal{U}}_{s.c}(U) = \overline{H}^{\mathcal{U}}_{s.c}(U|0)$.

The diligent reader will now be able to characterize an achievable rate region for a MAC-DSTx based on UCC built over groups (group UCC). As mentioned earlier, the encoding and decoding techniques are identical to that

[13]The interested reader is referred to [74], [75], [30] for early work on rates achievable using group codes for point-to-point channels. [76] provides bounds on rates achievable using Abelian group codes for point-to-point source and channel coding problems.

proposed in section 6.2.2 except for group addition replacing field addition. Consider a distribution $p_{\mathbf{USXY}} \in \mathbb{D}(\tau)$ defined over $\mathcal{U}^2 \times \mathcal{S} \times \mathcal{X} \times \mathcal{Y}$ where $\mathcal{U}$ is an Abelian group of order p^r. Cosets of a common group code is employed as bins of each user's code. Following an analysis similar to that performed in proof of theorem 6.2.2, one can prove the probability of the encoders not finding a codeword jointly typical with the state sequence decays exponentially with block length if the bins are of rate at least $\max\left\{\log|\mathcal{U}| - \overline{H}^{\mathcal{U}}_{s.c}(U_j|S_j) : j = 1,2\right\}$. The decoder decodes the group sum of chosen codewords from the group sum of the two users' codebooks. The codebooks of the two users are chosen to be union of arbitrary cosets of a common group code and therefore the the group sum of the two users codebooks will also be a union of arbitrary cosets of this group code. The probability of error at the decoders decays exponentially if the rate of the group sum of the two users' codebooks is at most $\log|\mathcal{U}| - H(U_1 \oplus U_2|Y)$. We conclude that a rate pair (R_1, R_2) is achievable if $R_1 + R_2 \leq \min\left\{\overline{H}^{\mathcal{U}}_{s.c}(U_j|S_j) : j = 1,2\right\} - H(U_1 \oplus U_2|Y)$. The following is a formal characterization of achievable rate region for MAC-DSTx using group codes over a $\mathbb{Z}_{p^r}-$group.

Definition 6.4.1 *Let $\mathbb{D}_G(\boldsymbol{\tau}) \subseteq \mathbb{D}(\boldsymbol{\tau})$ be the collection of distributions $p_{\mathbf{USXY}}$ on $\mathcal{U}^2 \times \mathcal{S} \times \mathcal{X} \times \mathcal{Y}$ where $\mathcal{U}$ is an Abelian group of order p^r, where p is a prime. For $p_{\mathbf{USXY}} \in \mathbb{D}_G(\boldsymbol{\tau})$, let $\beta_g(p_{\mathbf{USXY}})$ be defined as the set*

$$\left\{(R_1, R_2) \in [0,\infty)^2 : R_1 + R_2 \leq \min\left\{\overline{H}^{\mathcal{U}}_{s.c}(U_1|S_1), \overline{H}^{\mathcal{U}}_{s.c}(U_2|S_2)\right\} - H(U_1 \oplus U_2|Y)\right\} \tag{6.31}$$

where $\oplus$ denotes group addition in group $\mathcal{U} = \mathbb{Z}_{p^r}$, and

$$\beta_g(\boldsymbol{\tau}) := cocl\left(\bigcup_{p_{\mathbf{USXY}} \in \mathbb{D}_G(\boldsymbol{\tau})} \beta_g(p_{\mathbf{USXY}})\right). \tag{6.32}$$

Theorem 6.4.2 $\beta_g(\boldsymbol{\tau}) \in \mathbb{C}(\boldsymbol{\tau})$. □

Example 6.4.3 *Let us now compute achievable rate region using group UCC for QDD-MAC. $\mathcal{U} = \{0,1,2,3\}$ has two sub-groups - the group itself and $\{0,2\}$. It can be verified that*

$$\overline{I}^{\mathcal{U}}_{s.c}(U;S) = \max\left\{\log_2 4 - 2h_b(\frac{2\tau}{3}), \log 4 + \tau\log(\frac{\tau}{3}) + (1-\tau)\log(1-\tau)\right\}$$

yielding $\beta_{\mathcal{G}}(\boldsymbol{\tau}) = \left\{(R_1, R_2) \in [0,\infty)^2 : R_1 + R_2 \leq |\beta_g(\tau)|^+\right\}$, where

$$\beta_g(\tau) = \max\left\{\min\left\{-\tau\log(\frac{\tau}{3}) - (1-\tau)\log(1-\tau), 2h_b(\frac{2\tau}{3})\right\}, 0\right\}.$$

In figure 6.5, the sum rate achievable using group UCC for the above test channel is plotted. We highlight significant gains achievable using group UCC for QDD-MAC thus emphasizing the need to build codes with appropriate algebraic structure that matches the channel.

6.4.2 Achievable rate region for MAC-DSTx using group UCC : The general Abelian group

We now let the auxiliary alphabet $\mathcal{U}$ be a general Abelian group and build group UCC over $\mathcal{U}$ to enable the decoder to reconstruct the group sum of chosen codewords. The discussion in section 6.4.1 indicates that we only need to characterize the minimum rate of a bin in the code with respect to a generic test channel $p_{U|S}$ under the constraint that the bin has to be a coset of a group code. Essentially, this will involve characterizing fundamental group information theoretic quantity $\overline{I}^{\mathcal{U}}_{s.c}(U;S)$ and the related source coding group entropy $\overline{H}^{\mathcal{U}}_{s.c}(U|S)$ in the context of a general Abelian group $\mathcal{U}$.

Let $\mathcal{U}$ be the Abelian group in (6.28). Let $\theta = (\theta_1, \cdots, \theta_r)$ be such that $0 \leq \theta_i \leq r_i$ for $i = 1, 2, \cdots, I$ and let $\mathcal{H}_\theta$ be a subgroup of $\mathcal{U}$ defined as

$$\mathcal{H}_\theta = \bigoplus_{i=1}^{I} p^{\theta_i} \mathbb{Z}_{p_i^{r_i}},$$

and random variable $[U]_\theta$ taking values from cosets of $\mathcal{H}_\theta$ in $\mathcal{U}$ as $[U]_\theta = U \oplus \mathcal{H}_\theta$. If the state has a pmf p_S and the bins over $\mathcal{U}$ are constrained to be cosets of a group code, then for a test channel $p_{U|S}$, the rate of a bin has to be at least

$$\overline{I}^{\mathcal{U}}_{s.c}(U;S) \quad := \quad \min_{\substack{w_1,\cdots,w_I \\ w_1+\cdots+w_I=1}} \max_{\substack{\mathcal{H} \trianglelefteq \mathcal{U} \\ \mathcal{H} \neq \mathcal{U}}} \frac{1}{1-w_\theta} I([U]_\theta; S) \tag{6.33}$$

where

$$w_\theta = \sum_{i=1}^{I} \frac{r_i - \theta_i}{r_i} w_i.$$

Alternatively, one might express the minimum rate of the bin as $\log|\mathcal{U}| - \overline{H}^{\mathcal{U}}_{s.c}(U|S)$, where, as before

$$\overline{H}^{\mathcal{U}}_{s.c}(U|S) = \log|\mathcal{U}| - \overline{I}^{\mathcal{U}}_{s.c}(U;S), \tag{6.34}$$

is defined as the *source coding group entropy* of an Abelian group $\mathcal{U}$ and $\overline{H}^{\mathcal{U}}_{s.c}(U) = \overline{H}^{\mathcal{U}}_{s.c}(U|0)$. We note that definitions (6.33) and (6.34) defined for an arbitrary Abelian group reduces to that in (6.29) and (6.30) for a $\mathbb{Z}_{p^r}-$group. This enables us to characterize an achievable rate region for MAC-DSTx based on Abelian group codes using $\beta_g(\boldsymbol{\tau})$.

Definition 6.4.4 *Let $\mathbb{D}_G(\boldsymbol{\tau}) \subseteq \mathbb{D}(\boldsymbol{\tau})$ be the collection of distributions $p_{\mathbf{USXY}}$ on $\mathcal{U}^2 \times \mathcal{S} \times \mathcal{X} \times \mathcal{Y}$ where $\mathcal{U}$ is an Abelian group. For $p_{\mathbf{USXY}} \in \mathbb{D}_G(\boldsymbol{\tau})$, let $\beta_g(p_{\mathbf{USXY}})$ be defined as the set in (6.31) and $\beta_g(\boldsymbol{\tau})$ as in (6.32).*

We conclude by stating that $\beta_g(\boldsymbol{\tau})$ is indeed achievable.

Theorem 6.4.5 $\beta_g(\boldsymbol{\tau}) \in \mathbb{C}(\boldsymbol{\tau})$. □

Remark 6.4.6 *The persistent reader will recognize that the achievable rate region based on group UCC hinges on the characterization of the minimum rate of a bin that is closed under group addition with respect to a test channel* $p_{U|S}$. *For the general Abelian group we stated this to be (6.33). Recent pursuit has resulted in further reduction of this quantity and is available in [76].*

Remark 6.4.7 *The results in this section, though preliminary, point to a rich theory of strategies for multi-terminal communication systems based on structured code ensembles. Gains crucially rely on the compressive nature of the bivariate function and the ability to build efficient codes with rich algebraic structure. It is therefore no surprise that all of earlier findings were based on exploiting modulo*-2 *sum - the simplest compressive function with binary arguments - using linear codes - an ensemble that has been studied at length from different perspectives.*

6.5 Concluding Remarks

We have provided a single letter characterization of a new achievable rate region for the general MAC-DSTx. The reader will recognize that our findings are aimed at developing a new framework for enlarging achievable rate region for multi-terminal communication problems based on algebraic tools. We proposed achievable rate regions for an arbitrary MAC-DSTx based on two algebraic structures - fields and groups. It should now be clear to a persistent reader that a general rate region will involve a closure over all algebraic structures of which fields and groups are just two of them. Furthermore, this rate region will also incorporate the unstructured coding as indicated in section 6.3. Indeed, a description of this will be involved, and is justified by the presence of additional degrees of freedom in the multi-terminal communication settings.

Chapter 7

Computation over multiple access channel

Consider a scenario wherein a centralized receiver is interested in evaluating a multi-variate function, the arguments of which are available to spatially distributed transmitters. Traditionally, the technique of computing functions at a centralized receiver is based on it's decoding of the arguments in it's entirety. Solutions based on this technique have been proven optimal for particular instances of distributed source coding. Moreover, this technique lends itself naturally for communication based on separation. Buoyed by this partial success and ease of implementation, the de facto framework for computing at a centralized receiver is by enabling the decoder decode the arguments of the function in it's entirety.

The problem of computing mod-2 sum of distributed binary sources has proved to be an exception. Studied in the context of a source coding problem, Körner and Marton [18] propose an ingenious technique based on linear codes, that circumvent the need to communicate sources to the decoder, and thereby perform strictly better for a class of source distributions. In fact, as proposed in [18], the decoder needs only sum of message indices put out by the source encoder. This fact has been further exploited by Nazer and Gastpar [16] in developing a channel coding technique for a *linear* MAC, henceforth referred to as linear computation coding (LCC), that enables the decoder reconstruct the sum of the message indices input to the channel encoder. Since the decoder does not need to disambiguate individual message indices, this technique, when applicable, outperforms earlier known techniques.

LCC [16] is built around employing the same linear code as a channel code at both encoders. The message indices output by the Körner-Marton (KM) source code is linearly mapped into channel codewords. Since a linear MAC first computes a sum of the transmitted codewords, it is as if the codeword corresponding to the sum of messages was input to the ensuing channel. The first question that comes to mind is the following. If the MAC is not linear, would it be possible to decode sum of message indices without having to decode the individual codewords? In other

words, what would be the generalization of LCC for an arbitrary MAC?[1] If there exist such a generalization, how efficient would it be?

In this chapter, we answer the above question in the affirmative. Firstly, we recognize that in order to decode the sum of transmitted codewords, it is most efficient to employ channel codes that are closed under addition, of which a linear code employed in LCC is the simplest example. Closure under addition contains the range of the sum of transmitted codewords and thereby support a larger range for individual messages. Secondly, typical set decoding circumvents need for the MAC to be linear. Since nested coset codes have been proven to achieve capacity of arbitrary point-to-point channels [63] and are closed under addition, we employ this ensemble for generalizing the technique of LCC. As illustrated by examples 7.2.4,7.2.5 in section 7.2, the generalization we propose (i) outperforms separation based technique for an arbitrary MAC and moreover (ii) outperforms LCC even for examples with a structural match.[2] We remark that analysis of typical set decoding of a function of transmitted codewords with nested coset codes that contain statistically dependent codewords contains new elements and are detailed in proof of theorem 7.2.2.

Even in the case of a structural match, separation based schemes could outperform LCC [16, Example 4]. This raises the following question. What then would be a unified scheme for computing over an arbitrary MAC? Is there such a scheme that reduces to (i) separation when the desired function and MAC are not matched and (ii) LCC when appropriately matched? We recognize that KM technique is indeed suboptimal for a class of source distributions. For such sources, it is more efficient to transmit the sources as is. We therefore take the approach of Ahlswede and Han [48, Section VI], where in a two layer source code accomplishes distributed compression. The first layer generates message indices of those parts that are best reconstructed as is, and the second part employs a KM technique. In section 7.3, we propose a two layer channel code for MAC that is compatible with the above two layer source code. The first layer of the MAC channel code communicates the message indices as is, while the second layer enables the decoder decode the sum of second layer message indices. We therefore develop a unifying strategy that subsumes separation and LCC. Since Ahlswede and Han [48, Example 4] have proved the existence of source pairs for which their scheme outperforms both separation based and KM strategy their findings carry over to the problem studied herein. We highlight the significance of our contribution. Firstly, we propose a strategy based on nested coset codes and derive a set of sufficient conditions for the problem of computing sum of sources over an *arbitrary* MAC. The proposed strategy subsumes all current known strategies and performs strictly better for certain examples (section 7.2). Secondly, our findings highlight the utility of nested coset codes [63] as a generic ensemble of structured codes for communicating over arbitrary multi-terminal communication problems. Thirdly, and perhaps more importantly, our findings hint at a general theory of structured codes. Linear and nested linear codes

[1] The technique of systematic computation coding (SCC) [16] may not be considered as a generalization of LCC. Indeed SCC does not reduce to LCC for a linear MAC.

[2] This is expected since linear codes achieve only symmetric capacity and nested coset codes can achieve capacity of arbitrary point-to-point channels.

have been employed to derive communication strategies for particular symmetric additive source and channel coding problems that outperform all classical unstructured-code based techniques. However the question remains whether these structured code based techniques can be generalized to arbitrary multi-terminal communication problems. Our findings indicate that strategies based on structured codes can be employed to analyze more intelligent encoding and decoding techniques for an arbitrary multi-terminal communication problem.

This chapter is organized as follows. We begin with preliminaries and a brief description of LCC in section 7.1. Section 7.2 contains the main findings of this chapter - a generalization of LCC for an arbitrary MAC using the ensemble of nested coset codes. Theorem 7.2.2 is a statement of this characterization and examples 7.2.4-7.2.5 highlight the significance of theorem 7.2.2. In section 7.3, we propose a unified strategy for computing sum of sources over an arbitrary MAC that subsumes separation and LCC based techniques.

7.1 Preliminaries and Problem statement

Following remarks on notation (7.1.1) and problem statement (7.1.2), we briefly describe LCC for a linear MAC (7.1.3) and set the stage for it's generalization.

7.1.1 Notation

We employ notation that is now widely adopted in the information theory literature supplemented by the following. We let $\mathcal{F}_q$ denote a finite field of cardinality q. While $+$ denotes addition in $\mathbb{R}$, we let $\oplus$ denote addition in a finite field. The particular finite field, which is uniquely determined (up to an isomorphism) by it's cardinality, is clear from context. When ambiguous, or to enhance clarity, we specify addition in $\mathcal{F}_q$ using $\oplus_q$. For elements a, b, in a finite field, $a \ominus b := a \oplus (-b)$, where $(-b)$ is the additive inverse of b. The log and exp functions are taken with respect to the same base. For concreteness, the base may be assumed to be 2, in which case, units for information theoretic quantities such as entropy and mutual information would be bits. If $f : \mathcal{U} \to \mathcal{X}$ is a map, the n-letter extension of f denoted $f^n : \mathcal{U}^n \to \mathcal{X}^n$ is defined $f^n(u^n) := (f(u_i) : i = 1, 2, \cdots, n)$. In this chapter, we repeatedly refer to pairs of objects of similar type. To reduce clutter in notation, we use an underline to refer to aggregates of similar type. For example, (i) $\underline{S}$ abbreviates (S_1, S_2), (ii) if $\mathcal{X}_1, \mathcal{X}_2$ are finite alphabet sets, we let $\underline{\mathcal{X}}$ either denote the Cartesian product $\mathcal{X}_1 \times \mathcal{X}_2$ or abbreviate the pair $\mathcal{X}_1, \mathcal{X}_2$ of sets. More non trivially, if $e_j : \mathcal{S}^n \to \mathcal{X}_j^n : j = 1, 2$ are a pair of maps, we let $\underline{e}(\underline{s}^n)$ abbreviate $(e_1(s_1^n), e_2(s_2^n))$.

7.1.2 Problem statement

Consider a pair (S_1, S_2) of information sources each taking values over a finite field $\mathcal{S}$ of cardinality q. We assume outcome $(S_{1,t}, S_{2,t})$ of the sources at time $t \in \mathbb{N}$, is independent and identically distributed across time, with distribution $W_{\underline{S}}$. We let $(\mathcal{S}, W_{\underline{S}})$ denote this pair of sources. S_j is observed by encoder j that has access to input

j of a two user discrete memoryless multiple access channel (MAC) that is used without feedback. Let $\mathcal{X}_1$, $\mathcal{X}_2$ be the finite input alphabet sets and $\mathcal{Y}$ the finite output alphabet set of MAC. Let $W_{Y|X_1X_2}(y|x_1,x_2)$ denote MAC transition probabilities. We refer to this as MAC $(\underline{\mathcal{X}},\mathcal{Y},W_{Y|\underline{X}})$. The objective of the decoder is to compute $S_1 \oplus S_2$. In this chapter, we provide a characterization of a sufficient condition for computing $S_1 \oplus S_2$ with arbitrary small probability of error. The relevant notions are made precise in the following definitions.

Definition 7.1.1 *A computation code $(n,\underline{e},d)$ for computing sum of sources $(\mathcal{S},W_{\underline{S}})$ over the MAC $(\underline{\mathcal{X}},\mathcal{Y},W_{Y|\underline{X}})$ consists of (i) two encoder maps $e_j : \mathcal{S}^n \to \mathcal{X}_j^n : j = 1,2$ and (ii) a decoder map $d : \mathcal{Y}^n \to \mathcal{S}^n$.*

Definition 7.1.2 *The average error probability $\bar{\xi}(\underline{e},d)$ of a computation code $(n,\underline{e},d)$ is*

$$\sum_{\underline{s}\in\underline{\mathcal{S}}^n} \sum_{\substack{y^n : d(y^n) \neq \\ s_1^n \oplus s_2^n}} W_{Y^n|\underline{X}^n}(y^n|\underline{e}(\underline{s}^n)) W_{\underline{S}^n}(\underline{s}^n).$$

Definition 7.1.3 *The sum of sources $(\mathcal{S},W_{\underline{S}})$ is computable over MAC $(\underline{\mathcal{X}},\mathcal{Y},W_{Y|\underline{X}})$ if for all $\eta > 0$, there exists an $N(\eta) \in \mathbb{N}$ such that for all $n > N(\eta)$, there exists an $(n,\underline{e}^{(n)},d^{(n)})$ computation code such that $\bar{\xi}(\underline{e}^{(n)},d^{(n)}) \leq \eta$.*

The main objective in this chapter is to provide a sufficient condition for computability of sum of sources over a MAC.

7.1.3 Linear Computation Coding

We describe the technique of LCC in a simple setting and highlight the key aspects. Consider binary sources and a binary additive MAC, i.e., $\mathcal{S} = \mathcal{X}_1 = \mathcal{X}_2 = \{0,1\}$ and $Y = X_1 \oplus X_2 \oplus N$, where N is independent of the inputs and $P(N=1) = q$. Furthermore assume sources are symmetric, uniform, i.e., $P(\underline{S} = (0,0)) = \frac{1-p}{2} = P(\underline{S} = (1,1))$ and $P(\underline{S} = (0,1)) = P(\underline{S} = (1,0)) = \frac{p}{2}$ such that $h_b(p) < 1 - h_b(q)$.

By employing a KM source code, the two message indices at rate $h_b(p)$ can be employed to decode $S_1 \oplus S_2$. Let $h \in \mathcal{S}^{k\times n}$ denote a parity check matrix for the KM source code, with $\frac{k}{n}$ arbitrarily close to $h_b(p)$. Nazer and Gastpar observe that the decoder only requires the sum $h(S_1^n \oplus S_2^n) = h(S_1^n) \oplus h(S_2^n)$ of message indices. If the map from message indices to channel code is linear, then the decoder can infer $h(S_1^n) \oplus h(S_2^n)$ by decoding the codeword corresponding to sum of transmitted codewords. Since sum of transmitted codewords passes through a BSC(q), they employ a capacity achieving linear code of rate arbitrarily close to $1 - h_b(q)$ with generator matrix $g \in \mathcal{X}_1^{l\times n}$. Each encoder employs the same linear code and transmits $x_j^n := h(S_j^n)g$. The decoder receives Y^n and decodes as if the channel is a BSC(q). It ends up decoding message corresponding to $x_1^n \oplus x_2^n$ which was precisely what it was looking for.

We note that a separation based scheme will require the sum capacity of the MAC to be greater than $2h_b(p)$ and hence LCC is more efficient. What are key aspects of LCC? Note that (i) the channel code is designed for the

$X_1 \oplus X_2$ to Y channel, i.e., the BSC(q) and (ii) both encoders employ the same linear channel code, thereby ensuring their codes are closed under addition. This contains range of the sum of transmitted codewords to a rate $1 - h_b(q)$. It is instructive to analyze the case when the two users are provided two linear codes of rates R_1 and R_2 spanning disjoint subspaces. Since the range of sum of transmitted codewords is $R_1 + R_2$, the same decoding rule will impose the constraint $R_1 + R_2 < 1 - h_b(q)$ resulting in the constraint $2h_b(p) \leq 1 - h_b(q)$ which is strictly suboptimal. *We conclude that the two users' channel codes being closed under addition is crucial to the optimality of LCC for this problem.* Furthermore, the coupling of (i) a linear map of KM message indices to the channel code at the encoder and (ii) decoding of the sum of transmitted codewords, is central to LCC.

In the following section, we make use of the above observations to propose a generalization of LCC for computing sum of sources over an arbitrary MAC.

7.2 Nested coset codes for computing sum of sources over a MAC

In this section, we propose a technique for computing $S_1 \oplus S_2$ over an *arbitrary* MAC using the ensemble of nested coset codes [63], and derive a set of sufficient conditions under which, sum of sources $(\mathcal{S}, W_{\underline{S}})$ can be computed over a MAC $(\underline{\mathcal{X}}, \mathcal{Y}, W_{Y|\underline{X}})$. Definitions 7.2.1 and theorem 7.2.2 state these sufficient conditions. For certain examples such as example 7.2.5, the technique proposed in theorem 7.2.2 outperforms all known earlier techniques. Indeed, as illustrated by examples 7.2.4, 7.2.5, even in the case of a structural match, the above sufficient conditions are weaker than that imposed by LCC. Nevertheless, we further relax the same in section 7.3, or in other words enrich our technique, by incorporating separation.

Definition 7.2.1 *Let $\mathbb{D}(W_{Y|\underline{X}})$ be collection of distributions $p_{V_1V_2X_1X_2Y}$ defined over $\mathcal{S}^2 \times \underline{\mathcal{X}} \times \mathcal{Y}$ such that (i) $p_{V_1X_1V_2X_2} = p_{V_1X_1}p_{V_2X_2}$, (ii) $p_{Y|\underline{X}V} = p_{Y|\underline{X}} = W_{Y|\underline{X}}$. For $p_{\underline{V}\underline{X}Y} \in \mathbb{D}(W_{Y|\underline{X}})$, let $\alpha(p_{\underline{V}\underline{X}Y})$ be defined as*

$$\{R \geq 0 : R \leq \min\{H(V_1), H(V_2)\} - H(V_1 \oplus V_2|Y)\}, \quad \textit{and} \quad \alpha(W_{Y|\underline{X}}) := \sup \bigcup_{\substack{p_{\underline{V}\underline{X}Y} \in \\ \mathbb{D}(W_{Y|\underline{X}})}} \alpha(p_{\underline{V}\underline{X}Y}).$$

Theorem 7.2.2 *The sum of sources $(\mathcal{S}, W_{\underline{S}})$is computable over a MAC $(\underline{\mathcal{X}}, \mathcal{Y}, W_{Y|\underline{X}})$ if $H(S_1 \oplus S_2) \leq \alpha(W_{Y|\underline{X}})$.* □

Before we provide a proof, we briefly discuss the coding strategy and indicate how we attain the rates promised above. The reader is referred to [77] for a complete proof of theorem 7.2.2.

We begin with a description of the encoding rule. Encoder j employs a KM source code to compress the observed source. Let $M_j^l := hS_j^n$ denote corresponding message index, where $h \in \mathcal{S}^{l\times n}$ is a KM parity check matrix of rate $\frac{l}{n} \approx \frac{H(S_1 \oplus S_2)}{\log|\mathcal{S}|}$. Each encoder is provided with a common nested linear code taking values over $\mathcal{S}$. The nested linear code is described through a pair of generator matrices $g_I \in \mathcal{S}^{k\times n}$ and $g_{O/I} \in \mathcal{S}^{l\times n}$, where g_I and $\begin{bmatrix} g_I \\ g_{O/I} \end{bmatrix}$ are the

generator matrices of the inner (sparser) code and complete (finer) codes respectively, where

$$\frac{k\log|\mathcal{S}|}{n} \overset{(a)}{\geq} \log|\mathcal{S}| - \min\{H(V_1), H(V_2)\} \quad , \quad \frac{(k+l)\log|\mathcal{S}|}{n} \overset{(b)}{\leq} \log|\mathcal{S}| - H(V_1 \oplus V_2). \tag{7.1}$$

Encoder j picks a codeword in coset $\left(a^k g_I \oplus M_j^l g_{O/I} : a^k \in \mathcal{S}^k\right)$ indexed by M_j^l that is typical with respect to p_{V_j}. Based on this chosen codeword X^n is generated according to $p_{X_j|V_j}$ and transmitted.

The decoder is provided with the same nested linear code. Having received Y^n it lists all codewords that are jointly typical with Y^n with respect to distribution $p_{V_1\oplus V_2,Y}$. If it finds all such codewords in a unique coset, say $\left(a^k g_I \oplus m^l g_{O/I} : a^k \in \mathcal{S}^k\right)$, then it declares m^l to be the sum of KM message indices and employs KM decoder to decode the sum of sources. Otherwise, it declares an error.

As is typical in information theory, we derive an upper bound on probability of error by averaging the error probability over the ensemble of nested linear codes. For the purpose of proof, we consider user codebooks to be cosets of nested linear codes.[3] We average uniformly over the entire ensemble of nested *coset* codes. Lower bound (7.1(a)) ensures the encoders find a typical codeword in the particular coset. Upper bound (7.1(b)) enables us derive an upper bound on the probability of decoding error. From (7.1), it can be verified that if $H(S_1 \oplus S_2) \approx \frac{l\log|\mathcal{S}|}{n} \leq \min\{H(V_1), H(V_2)\} - H(V_1 \oplus V_2|Y)$ then the decoder can reconstruct the sum of sources with arbitrarily small probability of error.

Since the ensemble of codebooks contain statistically dependent codewords and moreover user codebooks are closely related, deriving an upper bound on the probability of error involves new elements. The informed reader will recognize that in particular, deriving an upper bound on the probability of decoding error will involve proving statistical independence of the pair of cosets indexed by KM indices (M_1^l, M_2^l) and any codeword in a coset corresponding to $\hat{m}^l \neq M_1^l \oplus M_2^l$. The statistical dependence of the codebooks results in new elements to the proof.

Proof: Given $\eta > 0$, our goal is to identify a computation code $(n, \underline{e}, d)$ such that $P(d(Y^n) \neq S_1^n \oplus S_2^n) \leq \eta$ for all sufficiently large $n \in \mathbb{N}$. The source sequences are mapped to channel input codewords in two stages. In the first stage, a distributed source code proposed by Körner and Marton [18] is employed to map n-length source sequences to message indices that takes values over $\mathcal{S}^l$. The second stage maps these indices to channel input codewords. We begin by stating the main findings of [18] on which our first stage relies.

Lemma 7.2.3 *Given a pair of $(\mathcal{S}, W_{\underline{S}})$ of information sources and $\eta > 0$, there exists an $N(\eta) \in \mathbb{N}$ such that for every $n \in \mathbb{N}$, there exists a parity check matrix $h \in \mathcal{S}^{l(n)\times n}$ and a map $r : \mathcal{S}^{l(n)} \to \mathcal{S}^n$ such that (i) $\frac{l(n)}{n} \leq \frac{H(S_1\oplus S_2)}{\log|\mathcal{S}|} + \frac{\eta}{2}$, and (ii) $P(r(hS_1^n \oplus hS_2^n) \neq S_1^n \oplus S_2^n) \leq \frac{\eta}{2}$.* □

Given $\eta > 0$, let $h \in \mathcal{S}^{l\times n}$ be a parity check matrix that satisfies (i) and (ii) in lemma 7.2.3. Let $M_j^l := hS_j^n : j = 1, 2$ be the message indices output by the source encoder. In the second stage, we identify maps $\mu_j : \mathcal{S}^l \to \mathcal{X}_j^n : j = 1, 2$ that maps these message indices to channel input codewords. The encoder $e_j : \mathcal{S}^n \to \mathcal{X}_j^n$ of the computation code is

[3] This is analogous to the use of cosets of a linear code to prove achievability of symmetric capacity over point to point channels.

therefore defined as $e_j(S_j^n) := \mu_j(hS_j^n)$. The second stage of the encoding is based on nested coset codes. We begin with a brief review of nested coset codes.

An (n,k) coset is a collection of vectors in $\mathcal{F}_q^n$ obtained by adding a constant bias vector to a $k-$dimensional subspace of $\mathcal{F}_q^n$. If $\lambda_O \subseteq \mathcal{F}_q^n$ and $\lambda_I \subseteq \lambda_O$ are $(n,k+l)$ and (n,k) coset codes respectively, then q^l cosets λ_O/λ_I that partition λ_O is a nested coset code.

A couple of remarks are in order. An (n,k) coset code is specified by a bias vector $b^n \in \mathcal{F}_q^n$ and generator matrices $g \in \mathcal{F}_q^{k\times n}$. If $\lambda_O \subseteq \mathcal{F}_q^n$ and $\lambda_I \subseteq \lambda_O$ are $(n,k+l)$ and (n,k) coset codes respectively, then there exists a bias vector $b^n \in \mathcal{F}_q^n$ and generator matrices $g_I \in \mathcal{F}_q^{k\times n}$ and $g_O = \begin{bmatrix} g_I \\ g_{O/I} \end{bmatrix} \in \mathcal{F}_q^{(k+l)\times n}$, such that b^n, g_I specify λ_I and b^n, g_O specify λ_O. Therefore, a nested coset code is specified by a bias vector b^n and any two of the three generator matrices g_I, $g_{O/I}$ and g_O. We refer to this as nested coset code $(n,k,l,g_I,g_{O/I},b^n)$.

We now specify the encoding rule. Encoder j is provided a nested coset code $(n,k,l,g_I,g_{O/I},b_j^n)$ denoted λ_{Oj}/λ_I taking values over the finite field $\mathcal{S}$. Let $v_j^n(a^k,m_j^l) := a^k g_I \oplus m_j^l g_{O/I} \oplus b_j^n$ denote a generic codeword in λ_{Oj}/λ_I and $c_j(m_j^l) := (v_j^n(a^k,m_j^l) : a^k \in \mathcal{S}^k)$ denote coset corresponding to message m_j^l. The message index $M_j^l = hS_j^n$ put out by the source encoder is used to index coset $c_j(M_j^l)$. Encoder j looks for a codeword in coset $c(M_j^l)$ that is typical according to p_{V_j}. If it finds at least one such codeword, one of them, say $v_j^n(a^k,M_j^l)$ is chosen uniformly at random. $\mu_j(M_j^l)$ is generated according $p_{X^n|V^n}(\cdot|v_j^n(a^k,M_j^l)) = \prod_{t=1}^n p_{X_j|V_j}(\cdot|(v_j^n(a^k,M_j^l))_t)$ and $\mu_j(M_j^l)$ is transmitted. Otherwise, an error is declared.

We now specify the decoding rule. The decoder is provided with the nested coset code $(n,k,l,g_I,g_{O/I},b^n)$ denoted λ_O/λ_I, where $b^n = b_1^n \oplus b_2^n$. We employ notation similar to that specified for the encoder. In particular, let $v^n(a^k,m^l) := a^k g_I \oplus m^l g_{O/I} \oplus b^n$ denote a generic codeword and $c(m^l) := (v^n(a^k,m^l) : a^k \in \mathcal{S}^k)$ denote a generic coset in λ_O/λ_I respectively. Decoder receives Y^n and declares error if $Y^n \notin T_{\frac{\eta_1}{2}}(p_Y)$. Else, it lists all codewords $v^n(a^k,m^l) \in \lambda_O$ such that $(v^n(a^k,m^l),Y^n) \in T_{\eta_1}^n(p_{V_1\oplus V_2,Y})$. If it finds all such codewords in a unique coset say $c(m^l)$ of λ_O/λ_I, then it declares $r(\hat{m}^l)$ to be the decoded sum of sources, where $r : \mathcal{S}^l \to \mathcal{S}^n$ is as specified in lemma 7.2.3. Otherwise, it declares an error.

As is typical in information theory, we derive an upper bound on probability of error by averaging the error probability over the ensemble of nested coset codes. We average over the ensemble of nested coset codes by letting the bias vectors $B_j^n : j = 1,2$ and generator matrices $G_I, G_{O/I}$ mutually independent and uniformly distributed over their respective range spaces. Let $\Lambda_{Oj}/\Lambda_I : j = 1,2$ and Λ_O/Λ_I denote the random nested coset codes $(n,k,l,G_I,G_{O/I},B_j^n) : j = 1,2$ and $(n,k,l,G_I,G_{O/I},B^n)$ respectively, where $B^n = B_1^n \oplus B_2^n$. For $a^k \in \mathcal{S}^k$, $m^l \in \mathcal{S}^l$, let $V_j^n(a^k,m_j^l) : j = 1,2$, $V^n(a^k,m^l)$ denote corresponding random codewords in $\Lambda_{Oj}/\Lambda_I : j = 1,2$ and Λ_O/Λ_I respectively. Let $C_j(m_j^l) := (V_j^n(a^k,m_j^l) : a^k \in \mathcal{S}^k)$ and $C(m^l) := (V^n(a^k,m^l) : a^k \in \mathcal{S}^k)$ denote random cosets in $\Lambda_{Oj}/\Lambda_I : j = 1,2$ and Λ_O/Λ_I corresponding to message $m_j^l : j = 1,2$ and m^l respectively. We now analyze error events and upper bound probability of error.

We begin by characterizing error events at encoder. If $\phi(m_j^l) := \sum_{a^k \in \mathcal{S}^k} 1_{\{(V_j^n(a^k,m_j^l)) \in T_{\eta_2}^n(p_{V_j})\}}$ and $\epsilon_{j1} := \{\phi(hS_j^n) = 0\}$, then ϵ_{j1} is the error event at encoder j. An upper bound on $P(\epsilon_{j1})$ can be derived by following the arguments in [Proof of Theorem 1][63]. Findings in [63] imply existence of $N_{j2} \in \mathbb{N}$ such that $\forall n \geq N_{j2}$, $P(\epsilon_{j1}) \leq \frac{\eta}{8}$ if $\frac{k}{n} > 1 - H(V_j)$.

The error event at the decoder is $\epsilon_2 \cup \epsilon_3$, where $\epsilon_2 := \{Y^n \notin T_{\frac{\eta_1}{2}}^n(p_Y)\}$ and

$$\epsilon_3 := \bigcup_{\substack{m^l \neq \\ hS_1^n \oplus hS_2^n}} \bigcup_{a^k \in \mathcal{S}^k} \left\{(V^n(a^k,m^l),Y^n) \in T_{\eta_1}^n(p_{V_1 \oplus V_2,Y})\right\}.$$

In order to upper bound $P(\epsilon_2)$ by conditional frequency typicality, it suffices to upper bound $P((\underline{e}(\underline{S}^n)) \notin T_{\frac{\eta_1}{4}}(p_{\underline{X}}))$. Note that (i) independence of $(V_j, X_j) : j = 1, 2$ implies the Markov chain $X_1 - V_1 - V_2 - X_2$, and (ii) the chosen codeword $V_j^n(a^k, M_j^l)$ and the transmitted vector $e_j(S_j^n) = \mu_j(M_j^l)$ are jointly typical with high probability as a consequence of conditional generation of the latter. By the Markov lemma, it suffices to prove $V_j^n(a^k, M_j^l) : j = 1, 2$ are jointly typical. If the codewords were chosen independently at random according to $\prod_{t=1}^n p_{V_j}$, this would fall out as a consequence of uniformly sampling from the typical set [26,]. However, the generation of nested coset code is different, and the proof of this involves an alternate route. An analogous proof of the Markov lemma is provided in proof of theorem 6.2.2 and omitted here in the interest of brevity.

It remains to upper bound $P((\epsilon_{11} \cup \epsilon_{21} \cup \epsilon_2)^c \cap \epsilon_3)$. In appendix N, we prove that if $\frac{(k+l)\log|\mathcal{S}|}{n} < \log|\mathcal{S}| - H(V_1 \oplus V_2|Y)$, there exists $N_4(\eta) \in \mathbb{N}$ such that $\forall n \geq N_4$, $P(\epsilon_3) \leq \frac{\eta}{8}$. Combining the bounds $\frac{k\log|\mathcal{S}|}{n} > \log|\mathcal{S}| - H(V_j)$ and $\frac{(k+l)\log|\mathcal{S}|}{n} < \log|\mathcal{S}| - H(V_1 \oplus V_2|Y)$, we note that $\frac{l\log|\mathcal{S}|}{n} < \min\{H(V_1), H(V_2)\} - H(V_1 \oplus V_2|Y)$, then the sum of message indices $h(S_1^n \oplus S_2^n)$ can be reconstructed at the decoder. This concludes proof of achievability.

The informed reader will recognize that deriving an upper bound on $P(\epsilon_3)$ will involve proving statistical independence of the pair $(C_j(hS_j^n) : j = 1, 2)$ of cosets and any codeword $V^n(\hat{a}^k, \hat{m}^l)$ corresponding to a competing sum of messages $\hat{m}^l \neq h(S_1^n \oplus S_2^n)$. This is considerably simple for a coding technique based on classical unstructured codes wherein codebooks and codewords in every codebook are independent. The coding technique proposed herein involves correlated codebooks and codewords resulting in new elements to the proof. The reader is encouraged to peruse details of this element presented in appendix N. ■

It can be verified that theorem 7.2.2 subsumes LCC. In particular the rate region presented in theorem 7.2.2 subsumes the rate region presented in [16, Theorem 1, Corollary 2]. This follows by substituting a uniform distribution for V_1, V_2. Therefore examples presented in [16] carry over as examples of rates achievable using nested coset codes.

We now present a sample of examples to illustrate significance of theorem 7.2.2. The reader is referred to [77] for additional examples that illustrate utility of the coding technique proposed herein. As was noted in [16, Example 4] a uniform distribution induced by a linear code maybe suboptimal even for computing functions over a MAC with a structural match. The following example, closely related to the former, demonstrates the ability of nested coset

codes to achieve a nonuniform distribution and thus exploit the structural match better.

Example 7.2.4 *Let S_1 and S_2 be a pair of independent and uniformly distributed sources taking values over the field $\mathcal{F}_5$ of five elements. The decoder wishes to reconstruct $S_1 \oplus_5 S_2$. The two user MAC channel input alphabets $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{F}_5$ and output alphabet $\mathcal{Y} = \{0, 2, 4\}$. The output Y is obtained by passing $W = X_1 \oplus_5 X_2$ through an asymmetric channel whose transition probabilities are given by $p_{Y|W}(y|1) = p_{Y|W}(y|3) = \frac{1}{3}$ for each $y \in \mathcal{Y}$ and $p_{Y|W}(0|0) = p_{Y|W}(2|2) = p_{Y|W}(4|4) = 1$. Let the number of source digits output per channel use be λ. We wish to compute the range of values of λ for which the decoder can reconstruct the sum of sources. This is termed as computation rate in [16].*

It can be verified that the decoder can reconstruct $S_1 \oplus_5 S_2$ using the technique of LCC if $\lambda \leq \frac{3}{5}\frac{\log_2(3)}{\log_2 5} = 0.4096$. A separation based scheme enables the decoder reconstruct the sum if $\lambda \leq \frac{1}{2}\frac{\log_2(3)}{\log_2(5)} = 0.3413$. We now explore the use of nested coset codes. It maybe verified that pmf

$$p_{\underline{V}\underline{X}Y}(\underline{v}, \underline{x}, x_1 \oplus_5 x_2) = \begin{cases} \frac{1}{4} & \text{if } v_1 = x_1, v_2 = x_2 \\ & \text{and } v_1, v_2 \in \{0,2\} \\ 0 & \text{otherwise .} \end{cases} \tag{7.2}$$

defined on $\mathcal{F}_5 \times \mathcal{F}_5$ satisfies (i),(ii) of definition 7.2.1 and moreover $\alpha(p_{\underline{V}\underline{X}Y}) = \{R \geq 0 : R \leq 1\}$. Thus nested coset codes enable reconstructing $S_1 \oplus_5 S_2$ at the decoder if $\lambda \leq \frac{1}{\log_2 5} = .43067$.

The above example illustrates the need for nesting codes in order to achieve nonuniform distributions. However, for the above example, a suitable modification of LCC is optimal. Instead of building codes over $\mathcal{F}_5$, let each user employ the linear code of rate 1[4] built on $\mathbb{F}_2$. The map $\mathbb{F}_2 \to \mathcal{X}_j : j = 1, 2$ defined as $0 \to 0$ and $1 \to 2$ induces a code over $\mathcal{F}_5$ and it can be verified that LCC achieves the rate achievable using nested coset codes. However, the following example precludes such a modification of LCC.

Example 7.2.5 *The source is assumed to be the same as in example 7.2.4. The two user MAC input and output alphabets are also assumed the same, i.e., $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{F}_5$ and output alphabet $\mathcal{Y} = \{0, 2, 4\}$. The output Y is obtained by passing $W = X_1 \oplus_5 X_2$ through an asymmetric channel whose transition probabilities are given by $p_{Y|W}(y|1) = p_{Y|W}(y|3) = \frac{1}{3}$ for each $y \in \mathcal{Y}$ and $p_{Y|W}(0|0) = p_{Y|W}(2|2) = p_{Y|W}(4|4) = 0.90, p_{Y|W}(2|0) = p_{Y|W}(4|0) = p_{Y|W}(0|2) = p_{Y|W}(4|2) = p_{Y|W}(0|4) = p_{Y|W}(2|4) = 0.05$.*

The technique of LCC builds a linear code over $\mathcal{F}_5$. It can be verified that the symmetric capacity for the $X_1 \oplus_5 X_2 (= W) - Y$ channel is 0.6096 and therefore LCC enables decoder reconstruct the sum if $\lambda \leq \frac{0.6096}{\log_2 5} = 0.2625$. A separation based scheme necessitates communicating each of the sources to the decoder and this can be done only if $\lambda \leq \frac{1}{2}\frac{\log_2 3}{\log_2 5} = 0.3413$. The achievable rate region of the test channel in (7.2) is $\alpha(p_{\underline{V}\underline{X}Y}) = \{R \geq 0 : R \leq 0.91168\}$ and therefore nested coset codes enable decoder reconstruct the sum if $\lambda \leq \frac{0.91168}{\log_2 5} = 0.3926$.

[4]This would be the set of all binary $n-$length vectors

Example 7.2.6 *Let S_1 and S_2 be independent sources distributed uniformly over $\{0,1,2\}$. The input alphabets $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{F}_3$ is the ternary field and the output alphabet $\mathcal{Y} = \mathcal{F}_2$ is the binary field. Let $W = 1_{\{X_1 \neq X_2\}}$ and output Y is obtained by passing W through a BSC with crossover probability 0.1. The decoder is interested in reconstructing W. As noted in [16, Example 8], W is 0 if an only if $S_1 \oplus_3 2S_2 = 0$. Therefore, it suffices for the decoder to reconstruct $S_1 \oplus_3 2S_2$. Following the arguments in proof of theorem 7.2.2 it can be proved that $S_1 \oplus_3 2S_2$ can be reconstructed using nested coset codes if there exists a pmf $p_{\underline{VX}Y} \in \mathbb{D}(W_{Y|\underline{X}})$ such that $H(S_1 \oplus_3 2S_2) \leq \min\{H(V_1), H(V_2)\} - H(V_1 \oplus_3 2V_2|Y)$. It can be verified that for pmf $p_{\underline{VX}Y}$ wherein V_1, V_2 are independently and uniformly distributed over $\mathcal{F}_3$, $X_1 = V_1$, $X_2 = V_2$, the achievable rate region is $\alpha(p_{\underline{VX}Y}) = \{R : R \leq 0.4790\}$. The computation rate achievable using SCC and separation technique are 0.194 and 0.168 respectively. The computation rate achievable using nested coset codes is $\frac{0.4790}{\log_2 3} = 0.3022$.*

Example 7.2.7 *Let S_1 and S_2 be independent and uniformly distributed binary sources and the decoder is interested in reconstructing the binary sum. The MAC is binary, i.e. $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{Y} = \mathbb{F}_2$ with transition probabilities $P(Y = 0|X_1 = x_1, X_2 = x_2) = 0.1$ if $x_1 \neq x_2$, $P(Y = 0|X_1 = X_2 = 0) = 0.8$ and $P(Y = 0|X_1 = X_2 = 1) = 0.9$. It can be easily verified that the channel is not linear, i.e., $\underline{X} - X_1 \oplus X_2 - Y$ is NOT a Markov chain. This restricts current known techniques to either separation based coding or SCC [16, Section V]. SCC yields a computation rate of 0.3291. The achievable rate region for the test channel $p_{\underline{VX}Y}$ where in V_1 and V_2 are independent and uniformly distributed binary sources, $X_1 = V_1, X_2 = V_2$ is given by $\{R : R \leq 0.4648\}$.*

We conclude by recognizing that example 7.2.7 is indeed a family of examples. As long as the MAC is close to additive but not additive, lending LCC inapplicable, we can expect nested coset codes to outperform separation and SCC. [77] presents more such examples.

7.3 General technique for computing sum of sources over a MAC

In this section, we propose a general technique for computing sum of sources over a MAC that subsumes separation and computation. The architecture of the code we propose is built on the principle that techniques based on structured coding are not in lieu of their counterparts based on unstructured coding. Indeed, the KM technique is outperformed by the Berger-Tung [64] strategy for a class of source distributions. A general strategy must therefore incorporate both.

We take the approach of Ahlswede and Han [48, Section VI], where in a two layer source code is proposed. Each source encoder j generates two message indices M_{j1}, M_{j2}. M_{j1} is an index to a Berger-Tung source code and M_{j2} is an index to a KM source code. The source decoder therefore needs M_{11}, M_{21} and $M_{12} \oplus M_{22}$ to reconstruct the quantizations and thus the sum of sources. We propose a two layer MAC channel code that is compatible with the above source code. The first layer of this code is a standard MAC channel code based on unstructured codes

[3, 4]. The messages input to this layer are communicated as is to the decoder. The second layer employs nested coset codes and is identical to the one proposed in theorem 7.2.2. A function of the codewords selected from each layer is input to the channel. The decoder decodes a triple - the pair of codewords selected from the first layer and a sum of codewords selected from the second layer - and thus reconstructs the required messages. The following characterization specifies rates of layers 1 and 2 separately and therefore differs slightly from [48, Theorem 10].

Definition 7.3.1 *Let $\mathbb{D}_{AH}(W_{\underline{S}})$ be collection of distributions $p_{T_1T_1S_1S_2}$ defined over $\mathcal{T}_1 \times \mathcal{T}_2 \times \mathcal{S}^2$ such that (a) $\mathcal{T}_1, \mathcal{T}_2$ are finite sets, (b) $p_{S_1S_2} = W_S$, (c) $T_1 - S_1 - S_2 - T_2$ is a Markov chain. For $p_{\underline{TS}} \in \mathbb{D}_{AH}(W_{\underline{S}})$, let*

$$\beta_S(p_{\underline{TS}}) := \left\{(R_{11}, R_{12}, R_2) \in \mathbb{R}^3 : R_{11} \geq I(T_1; S_1|T_2), R_{12} \geq I(T_2; S_2|T_1), R_2 \geq H(S_1 \oplus S_2|\underline{T}), R_{11} + R_{12} \geq I(\underline{T}; \underline{S})\right\}.$$

Let $\beta_S(W_{\underline{S}})$ denote convex closure of the union $\beta_S(p_{\underline{TS}})$ over $p_{\underline{TS}} \in \mathbb{D}_{AH}(W_{\underline{S}})$

We now characterize achievable rate region for communicating these indices over a MAC. We begin with a definition of test channels and the corresponding rate region.

Definition 7.3.2 *Let $\mathbb{D}_G$ be collection of distributions $p_{U_1U_2V_1V_2X_1X_2Y}$ defined on $\mathcal{U}_1 \times \mathcal{U}_2 \times \mathcal{S} \times \mathcal{S} \times \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}$ such that (i) $p_{\underline{UVX}} = p_{U_1V_1X_1}p_{U_2V_2X_2}$, (ii) $p_{Y|\underline{XUV}} = p_{Y|\underline{X}} = W_{Y|\underline{X}}$. For $p_{\underline{UVXY}} \in \mathbb{D}_G$, let $\beta_C(p_{\underline{UVXY}})$ be defined as the set*

$$\left\{\begin{array}{l} (R_{11}, R_{12}, R_2) \in \mathbb{R}^3 : 0 \leq R_{11} \leq I(U_1; Y, U_2, V_1 \oplus V_2), \quad 0 \leq R_{12} \leq I(U_2; Y, U_1, V_1 \oplus V_2), \\ R_{11} + R_{12} \leq I(\underline{U}; Y, V_1 \oplus V_2), \quad R_{11} + R_2 \leq \mathscr{H}_{\min}(V|U) + H(U_1) - H(V_1 \oplus V_2, U_1|Y, U_2) \\ R_2 \leq \mathscr{H}_{\min}(V|U) - H(V_1 \oplus V_2|Y, \underline{U}), \quad R_{12} + R_2 \leq \mathscr{H}_{\min}(V|U) + H(U_2) - H(V_1 \oplus V_2, U_2|Y, U_1) \\ R_{11} + R_{12} + R_2 \leq \mathscr{H}_{\min}(V|U) + H(U_1) + H(U_2) - H(V_1 \oplus V_2, \underline{U}|Y) \end{array}\right\}.$$

where $\mathscr{H}_{\min}(V|U) := \min\{H(V_1|U_1), H(V_2|U_2)\}$ and

$$\beta_C(W_{Y|\underline{X}}) cocl \left(\bigcup_{\substack{p_{\underline{UVXY}} \\ \in \mathbb{D}_G(W_{Y|\underline{X}})}} \beta_C(p_{\underline{UVXY}}) \right).$$

Theorem 7.3.3 *The sum of sources $(\mathcal{S}, W_S)$ is computable over MAC $(\underline{\mathcal{X}}, \mathcal{Y}, W_{Y|\underline{X}})$ if $\beta_S(W_{\underline{S}}) \cap \beta_C(W_{Y|\underline{X}}) \neq \phi$.* □

Remark 7.3.4 *It is immediate that the general strategy subsumes separation and computation based techniques. Indeed, substituting $\underline{T}, \underline{U}$ to be degenerate yields the conditions provided in theorem 7.2.2. Substituting $\underline{V}$ to be degenerate yields separation based technique.*

7.4 Concluding Remarks

Having decoded the sum of sources, we ask whether it would be possible to decode an arbitrary (non-additive) function of the sources using the above techniques? The answer is yes and the technique involves 'embedding'. Example 7.2.6 illustrates embedding and a framework is proposed in [77]. This leads us to the following fundamental question. The central element of the technique presented above was to decode the *sum* of transmitted codewords and use that to decode sum of KM message indices. If the MAC is 'far from additive', is it possible to decode a different bivariate function of transmitted codewords and use that to decode the desired function of the sources? The answer to the first question is yes Indeed, the elegance of joint typical encoding and decoding enables us reconstruct other 'well behaved' functions of transmitted codewords. We recognize that if codebooks take values over a finite field and were closed under addition, it was natural and more efficient to decode the sum. On the other hand, if the codebooks were taking values over an algebraic object, for example a group, and were closed with respect to group multiplication, it would be natural and efficient to decode the product of transmitted codewords. Since, we did not require the MAC to be linear in order to compute the sum of transmitted codewords, we will not require it to multiply in order for us to decode the product of transmitted codewords. We elaborate on this in [77].

Appendices

Appendix A

An upper bound on $P(\epsilon_1^c \cap \epsilon_2)$

Through out this appendix π denotes $\pi(\min\{(|\mathcal{X}| \cdot |\mathcal{S}|)^2, (|\mathcal{X}| + |\mathcal{S}| + |\mathcal{Y}| - 2) \cdot |\mathcal{X}| \cdot |\mathcal{S}|\})$ and $\mathcal{V} := \mathcal{F}_\pi$. We begin with a simple lemma. The following lemma holds for any $\mathcal{F}_q$ and we state it in this generality.

Lemma A.0.1 *Let $\mathcal{F}_q$ be the finite field of cardinality q. If generator matrices $G_I \in \mathcal{F}_q^{k\times n}$, $G_{O/I} \in \mathcal{F}_q^{l\times n}$ and bias vector $B^n \in \mathcal{F}_q^n$ of the random nested coset code $(n, k, l, G_I, G_{O/I}, B^n)$ are mutually independent and uniformly distributed on their respective range spaces, then codewords $V^n(a^k, m^l) := a^k G_I \oplus m^l G_{O/I} \oplus B^n$ are (i) uniformly distributed, and (ii) pairwise independent.* □

The proof follows form a simple counting argument and is omitted for the sake of brevity. The proof for the case $q = 2$ is provided in [14, Theorem 6.2.1] and the same argument holds for any field $\mathcal{F}_q$.

We derive an upper bound on $P(\epsilon_1^c \cap \epsilon_2)$ using a second moment method similar to that employed in [59].

$$
\begin{aligned}
P(\epsilon_1^c \cap \epsilon_2) &= \sum_{s^n \in T_{\frac{\delta}{4}}(p_S)} \sum_{m^l \in \mathcal{V}^l} P\left(\begin{smallmatrix} S^n = s^n, M^l = m^l \\ \phi_{\frac{\delta}{2}}(s^n, m^l) = 0 \end{smallmatrix}\right) = \sum_{s^n \in T_{\frac{\delta}{4}}(S)} \sum_{m^l \in \mathcal{V}^l} P\left(\begin{smallmatrix} S^n = s^n \\ M^l = m^l \end{smallmatrix}\right) P(\phi_{\frac{\delta}{2}}(s^n, m^l) = 0) \qquad \text{(A.1)} \\
&\leq \sum_{s^n \in T_{\frac{\delta}{4}}(S)} \sum_{m^l \in \mathcal{V}^l} P(S^n = s^n, M^l = m^l) P(|\phi_{\frac{\delta}{2}}(s^n, m^l) - \mathbb{E}\phi_{\frac{\delta}{2}}(s^n, m^l)| \geq \mathbb{E}\phi_{\frac{\delta}{2}}(s^n, m^l)) \\
&\leq \sum_{s^n \in T_{\frac{\delta}{4}}(S)} \sum_{m^l \in \mathcal{V}^l} P\left(S^n = s^n, M^l = m^l\right) \frac{\mathrm{Var}\left\{\phi_{\frac{\delta}{2}}(s^n, m^l)\right\}}{\left\{\mathbb{E}\left\{\phi_{\frac{\delta}{2}}(s^n, m^l)\right\}\right\}^2}, \qquad \text{(A.2)}
\end{aligned}
$$

where (A.1) is true since $\phi_{\frac{\delta}{2}}(s^n, m^l)$ is a function of random objects G_I, $G_{O/I}$ and B^n that are mutually independent of S^n, M^l, and (A.2) follows from Cheybyshev inequality.

We now evaluate first and second moments of $\phi_{\frac{\delta}{2}}(s^n, m^l)$. The expectation of $\phi_{\frac{\delta}{2}}(s^n, m^l)$ is

$$\mathbb{E}\phi_{\frac{\delta}{2}}(s^n, m^l) = \sum_{v^n \in T^n_{\frac{\delta}{2}}(V|s^n)} \sum_{a^k \in \mathcal{V}^k} P\left(V^n(a^k, M^l) = v^n\right) = \frac{|T^n_{\frac{\delta}{2}}(V|s^n)|}{\pi^{n-k}},$$

where the last equality follows from Lemma A.0.1(i). The second moment is

$$\begin{aligned}
\mathbb{E}\phi^2_{\frac{\delta}{2}}(s^n, m^l) &= \sum_{v^n, \tilde{v}^n \in T^n_{\frac{\delta}{2}}(V|s^n)} \sum_{a^k, \tilde{a}^k \in \mathcal{V}^k} P\left(V^n(a^k, M^l) = v^n, V^n(\tilde{a}^k, M^l) = \tilde{v}^n\right) \\
&= \sum_{\substack{v^n \in \\ T^n_{\frac{\delta}{2}}(V|s^n)}} \sum_{a^k \in \mathcal{V}^k} P\left(V^n(a^k, M^l) = v^n\right) + \sum_{\substack{v^n, \tilde{v}^n \in \\ T^n_{\frac{\delta}{2}}(V|s^n)}} \sum_{\substack{a^k, \tilde{a}^k \in \\ \mathcal{V}^k, a^k \neq \tilde{a}^k}} P\left(V^n(a^k, M^l) = v^n, V^n(\tilde{a}^k, M^l) = \tilde{v}^n\right) \\
&= \frac{\pi^k \left|T^n_{\frac{\delta}{4}}(V|s^n)\right|}{\pi^n} + \frac{\left|T^n_{\frac{\delta}{2}}(V|s^n)\right|^2 \pi^k \left(\pi^k - 1\right)}{\pi^{2n}}, \qquad \text{(A.3)}
\end{aligned}$$

where second term in (A.3) follows from Lemma A.0.1(ii). Substituting for first and second moments of $\phi_{\frac{\delta}{2}}(s^n, m^l)$, we have

$$\mathrm{Var}\left\{\phi_{\frac{\delta}{2}}(s^n, m^l)\right\} = \frac{\pi^k \left|T^n_{\frac{\delta}{2}}(V|s^n)\right|}{\pi^n}\left(1 - \frac{\left|T^n_{\frac{\delta}{2}}(V|s^n)\right|}{\pi^n}\right), \text{ thus } \frac{\mathrm{Var}\left\{\phi_{\frac{\delta}{2}}(s^n, m^l)\right\}}{\mathbb{E}\left\{\phi_{\frac{\delta}{2}}(s^n, m^l)\right\}^2} \leq \frac{\pi^{n-k}}{|T^n_{\frac{\delta}{2}}(V|s^n)|}. \qquad \text{(A.4)}$$

For $s^n \in T_{\frac{\delta}{4}}(S)$ lemma 2.4.2, guarantees existence of $N_3(\eta) \in \mathbb{N}$, such that for all $n \geq N_3(\eta)$, $|T_{\frac{\delta}{2}}(V|s^n)| \geq \exp\{n\left(H(V|S) - \delta\right)\}$. Substituting this lower bound in (A.4), we note,

$$\frac{\mathrm{Var}\left\{\phi_{\frac{\delta}{2}}(s^n, m^l)\right\}}{\mathbb{E}\left\{\phi_{\frac{\delta}{2}}(s^n, m^l)\right\}^2} \leq \frac{\pi^{n-k}}{|T^n_{\frac{\delta}{2}}(V|s^n)|} \leq \exp\left\{-n \log \pi \left(\frac{k}{n} - \left(1 - \frac{H(V|S)}{\log \pi} + \frac{\delta}{\log \pi}\right)\right)\right\}.$$

Substituting (A.5) in (A.2), we obtain

$$P(\epsilon_1^c \cap \epsilon_2) \quad \leq \quad \exp\left\{-n \log \pi \left(\frac{k}{n} - \left(1 - \frac{H(V|S)}{\log \pi} + \frac{\delta}{\log \pi}\right)\right)\right\}. \qquad \text{(A.5)}$$

From (3.3), we have

$$\frac{k}{n} - \left(1 - \frac{H(V|S)}{\log \pi} + \frac{\delta}{\log \pi}\right) \geq \frac{\frac{\eta}{8} - \delta}{\log \pi} \geq \frac{11\eta}{96 \log \pi} \qquad \text{(A.6)}$$

where the last inequality follows from choice of δ. Combining (A.5) and (A.6), we have $P(\epsilon_1^c \cap \epsilon_2) \leq \exp\left\{-\frac{11n\eta}{96}\right\} \leq \frac{\eta}{16}$ for all $n \geq N_4(\eta)$.

By choosing $\delta > 0$ sufficiently small, $\frac{k}{n}$ can be made arbitrarily close to $1 - \frac{H(V|S)}{\log \pi}$ and probability of encoding error can be made arbitrarily small by choosing a sufficiently large block length. The above findings are summarized in the following lemma.

Lemma A.0.2 *Let $\mathcal{S}$ be a finite set, $\mathcal{V} = \mathcal{F}_q$ a finite field and p_{SV}, a pmf on $\mathcal{S} \times \mathcal{V}$. Consider a random nested coset code $(n, k, l, G_I, G_{O/I}, B^n)$ denoted Λ_O/Λ_I, with bias vector $B^n \in \mathcal{V}^n$, generator matrices $G_I \in \mathcal{V}^{k\times n}$ and $G_{O/I} \in \mathcal{V}^{l\times n}$ mutually independent and uniformly distributed on their respective range spaces. Let $V^n(a^k, m^l) := a^k G_I \oplus m^l G_{O/I} \oplus B^n$ denote generic codeword in Λ_O/Λ_I. For $s^n \in \mathcal{S}^n$, $m^l \in \mathcal{V}^l$ and $\delta > 0$, let $\phi_\delta(s^n, m^l) := \sum_{a^k \in \mathcal{V}^k} 1_{\{(s^n, V^n(a^k, m^l)) \in T_\delta(S,V)\}}$. The following are true.*

(i) The codewords $V^n(a^k, m^l) : a^k \in \mathcal{V}^k$ are uniformly distributed and pairwise independent.

(ii) For any $\delta > 0$, $s^n \in T_{\frac{\delta}{2}}(S)$, $m^l \in \mathcal{V}^l$, there exists $N(\delta) \in \mathbb{N}$ such that for all $n \geq N(\delta)$,

$$P(\phi_\delta(s^n, m^l) = 0) \leq \exp\left\{-n \log q \left(\frac{k}{n} - \left(1 - \frac{H(V|S)}{\log q} - \frac{3\delta}{2 \log q}\right)\right)\right\}.$$

(iii) If $(S^n, M^l) \in \mathcal{S}^n \times \mathcal{V}^l$ are independent of $(G_I, G_{O/I}, B^n)$, then for all $n \geq N(\delta)$,

$$P(S^n \in T_{\frac{\delta}{2}}(S), \phi_\delta(S^n, M^l) = 0) \leq \exp\left\{-n \log q \left(\frac{k}{n} - \left(1 - \frac{H(V|S)}{\log q} - \frac{3\delta}{2 \log q}\right)\right)\right\}.$$

□

Appendix B

An upper bound on $P\big((\epsilon_1 \cup \epsilon_2 \cup \epsilon_3^c)^c \cap \epsilon_4\big)$

As is typical, our achievability proof hinges on independence of transmitted codeword (and hence received vector) and the contending codewords that are not transmitted. Towards this end, we begin with the following.

Lemma B.0.3 *Let $\mathcal{V}$ be the finite field of cardinality q. If generator matrices $G_I \in \mathcal{F}_q^{k\times n}$, $G_{O/I} \in \mathcal{V}^{l\times n}$ and bias vector $B^n \in \mathcal{F}_q^n$ of the random $(n,k,l,G_I,G_{O/I},B^n)$ nested coset code are mutually independent and uniformly distributed on their respective range spaces, then any coset is independent of any codeword in a different coset., i.e., the collection of codewords $(V^n(a^k,m^l) : a^k \in \mathcal{F}_q^k)$ and $V^n(\hat{a}^k,\hat{m}^l)$ are independent if $m^l \neq \hat{m}^l$.* □

Proof: Let $v_{a^k}^n \in \mathcal{F}_q^n$ for each $a^k \in \mathcal{F}_q^k$, and $\hat{v}^n \in \mathcal{F}_q^n$. We need to prove

$$P(V^n(a^k,m^l) = v_{a^k}^n : a^k \in \mathcal{F}_q^k, V^n(\hat{a}^k,m^l) = \hat{v}^n) = P(V^n(a^k,m^l) = v_{a^k}^n : a^k \in \mathcal{F}_q^k)P(V^n(\hat{a}^k,m^l) = \hat{v}^n).$$

If $(v_{a^k+\hat{a}^k}^n - v_{0^k}^n) \neq (v_{a^k}^n - v_{0^k}^n) + (v_{\hat{a}^k}^n - v_{0^k}^n)$ for some pair a^k, $\hat{a}^k \in \mathcal{F}_q^k$, the LHS and first term of RHS are zero and equality holds. Else,

$$\begin{aligned}
&P(V^n(a^k,m^l) = v_{a^k}^n : a^k \in \mathcal{F}_q^k, V^n(\hat{a}^k,m^l) = \hat{v}^n) \\
&= P(a^kG_I = v_{a^k}^n - v_{0^k}^n : a^k \in \mathcal{F}_q^k, V^n(0^k,m^l) = v_{0^k}^n, V^n(0^k,\hat{m}^l) = \hat{v}^n - v_{\hat{a}^k}^n) \\
&= P(a^kG_I = v_{a^k}^n - v_{0^k}^n : a^k \in \mathcal{F}_q^k)P(V^n(0^k,m^l) = v_{0^k}^n, V^n(0^k,\hat{m}^l) = \hat{v}^n - v_{\hat{a}^k}^n) \qquad \text{(B.1)} \\
&= P(a^kG_I = v_{a^k}^n - v_{0^k}^n : a^k \in \mathcal{F}_q^k)P(V^n(0^k,m^l) = v_{0^k}^n)P(V^n(0^k,\hat{m}^l) = \hat{v}^n - v_{\hat{a}^k}^n) \qquad \text{(B.2)} \\
&= P(a^kG_I = v_{a^k}^n - v_{0^k}^n : a^k \in \mathcal{F}_q^k, V^n(0^k,m^l) = v_{0^k}^n)P(\hat{m}^lG_{O/I} + B^n = \hat{v}^n - v_{\hat{a}^k}^n) \qquad \text{(B.3)} \\
&= P(V^n(a^k,m^l) = v_{a^k}^n : a^k \in \mathcal{F}_q^k)P(V^n(\hat{a}^k,m^l) = \hat{v}^n,)
\end{aligned}$$

where (B.1) and (B.3) follow from independence of $G_{O/I}$, B^n and G_I (B.2) follows from Lemma A.0.1(ii), and the last equality follows from invariance of the pmf of $V^n(a^k,m^l)$ with respect to a^k and m^l. ■

We emphasize the consequence of Lemma B.0.3 in the following remark.

Remark B.0.4 *If transmitted message $M^l \neq \hat{m}^l$, then Y^n is independent of $V^n(\hat{a}^k, \hat{m}^l)$. Indeed*

$$P(V^n(\hat{a}^k,\hat{m}^l)=\hat{v}^n, Y^n=y^n) = \sum_{(v^n_{a^k}\in\mathcal{V}^n: a^k\in\mathcal{V}^k)} \sum_{x^n\in\mathcal{X}^n} P\left(\begin{smallmatrix} C(M^l)=(v^n_{a^k}\in\mathcal{V}^n: a^k\in\mathcal{V}^k), \\ V^n(\hat{a}^k,\hat{m}^l)=\hat{v}^n, E(S^n,M^l)=x^n, Y^n=y^n \end{smallmatrix}\right)$$

$$= \sum_{(v^n_{a^k}\in\mathcal{V}^n: a^k\in\mathcal{V}^k)} \sum_{x^n\in\mathcal{X}^n} P\left(\begin{array}{l} C(M^l)=(v^n_{a^k}\in\mathcal{V}^n: a^k\in\mathcal{V}^k), \\ E(S^n,M^l)=x^n, Y^n=y^n \end{array}\right) P(V^n(\hat{a}^k,\hat{m}^l)=\hat{v}^n) \tag{B.4}$$

$$= P(V^n(\hat{a}^k,\hat{m}^l)=\hat{v}^n)P(Y^n=y^n) = \frac{P(Y^n=y^n)}{q^n}. \tag{B.5}$$

We have used (1) independence of $V^n(\hat{a}^k,\hat{m}^l)$ and $C(M^l)$ (lemma B.0.3), (2) $E(S^n,M^l)$ being a function of $C(M^l)$ and S^n is conditionally independent of $V^n(\hat{a}^k,\hat{m}^l)$ given $C(M^l)$, and (3) Y^n is conditionally independent of $V^n(\hat{a}^k,\hat{m}^l)$ given $E(S^n,M^l)$ in arriving at (B.4), and lemma A.0.1(i) in arriving at the last equality in (B.5).

We now provide an upper bound on $P((\epsilon_1\cup\epsilon_2\cup\epsilon_3)^c\cap\epsilon_4)$. Observe that

$$P((\epsilon_1\cup\epsilon_2\cup\epsilon_3)^c\cap\epsilon_4) \leq P\left(\bigcup_{\hat{a}^k\in\mathcal{V}^k}\bigcup_{\hat{m}^l\neq M^l}\{(V^n(\hat{a}^k,\hat{m}^l),Y^n)\in T_\delta(p_{VY})\}\right)$$

$$\leq \sum_{\substack{\hat{m}^l\in\mathcal{V}^l\\ \hat{m}^l\neq M^l}} \sum_{\hat{a}^k\in\mathcal{V}^k} \sum_{\substack{y^n\\ \in T_{\frac{\delta}{2}}}} \sum_{\substack{v^n\in\\ T_\delta(V|y^n)}} P(V^n(\hat{a}^k,\hat{m}^l)=v^n, Y^n=y^n)$$

$$= \sum_{\substack{\hat{m}^l\in\mathcal{V}^l\\ \hat{m}^l\neq M^l}} \sum_{\hat{a}^k\in\mathcal{V}^k} \sum_{\substack{y^n\\ \in T_{\frac{\delta}{2}}}} \sum_{\substack{v^n\in\\ T_\delta(V|y^n)}} P(V^n(\hat{a}^k,\hat{m}^l)=v^n)P(Y^n=y^n) = \sum_{\substack{\hat{m}^l\in\mathcal{V}^l\\ \hat{m}^l\neq M^l}} \sum_{\hat{a}^k\in\mathcal{V}^k} \sum_{\substack{y^n\\ \in T_{\frac{\delta}{2}}}} \sum_{\substack{v^n\in\\ T_\delta(V|y^n)}} \frac{P(Y^n=y^n)}{\pi^n} \tag{B.6}$$

$$\leq \sum_{y^n\in T_{\frac{\delta}{2}}} \frac{\pi^{k+l}|T_\delta(p_{V|Y}|y^n)|P(Y^n=y^n)}{\pi^n}, \tag{B.7}$$

where, the two equalities in (B.6) follow from (B.5). Lemma 2.4.2 guarantees existence of $N_5(\eta)\in\mathbb{N}$ such that for all $n\geq N_5(\eta)$ and $y^n\in T_{\frac{\delta}{2}}(p_Y)$, $|T_\delta(V|y^n)|\leq\exp\{n(H(V|Y)+\frac{3\delta}{2})\}$. Substituting this upper bound in (B.7), we conclude

$$P((\epsilon_1\cup\epsilon_2\cup\epsilon_3)^c\cap\epsilon_4) \leq \exp\left\{-n\log\pi\left(1-\frac{H(V|Y)}{\log\pi}-\frac{3\delta}{2\log\pi}-\frac{k+l}{n}\right)\right\} \tag{B.8}$$

for all $n\geq N_5(\eta)$.

Appendix C

Upper bound on $P(\epsilon_{l_j})$

Recall

$$\phi_j(q^n, M_j) := \sum_{a^{s_j} \in \mathcal{U}^{s_j}} \sum_{b_{jX} \in c_{jX}} 1_{\{I(a^{s_j})=M_{j1}, (q^n, U_j^n(a^{s_j}), X_j^n(M_{jX}, b_{jX})) \in T_{2\eta_2}(Q, U_j, X_j)\}}, \quad \mathcal{L}_j(n) := \frac{1}{2}\mathbb{E}\left\{\phi_j(q^n, M_j)\right\}$$

and $\epsilon_{l_j} = \{\phi_j(q^n, M_j) < \mathcal{L}_j(n)\}$. Employing Cheybyshev's inequality, we have

$$P(\epsilon_{l_j}) = P(\phi_j(q^n, M_j) < \mathcal{L}_j(n)) \leq P(|\phi_j(q^n, M_j) - \mathbb{E}\{\phi_j(q^n, M_j)\}| \geq \frac{1}{2}\mathbb{E}\{\phi_j(q^n, M_j)\}) \leq \frac{4\text{Var}\{\phi_j(q^n, M_j)\}}{\left(\mathbb{E}\{\phi_j(q^n, M_j)\}\right)^2}.$$

Note that $\text{Var}\left\{\phi_j(q^n, M_j)\right\} = \mathscr{T}_0 + \mathscr{T}_1 + \mathscr{T}_2 + \mathscr{T}_3 - \mathscr{T}_0^2$, where

$$\begin{aligned}
\mathscr{T}_0 &= \sum_{a^{s_j} \in \mathcal{U}^{s_j}} \sum_{b_{jX} \in c_{jX}} \sum_{\substack{(u_j^n, x_j^n) \in \\ T_{2\eta_2}(U_j, X_j|q^n)}} P\left(\begin{smallmatrix} I(a^{s_j})=M_{j1}, U_j(a^{s_j})=u_j^n \\ X_j^n(M_{jX}, b_{jX})=x_j^n \end{smallmatrix}\right) = \mathbb{E}\{\phi_j(q^n, M_j)\}, \qquad \text{(C.1)} \\
\mathscr{T}_1 &= \sum_{a^{s_j} \in \mathcal{U}^{s_j}} \sum_{\substack{b_{jX}, \tilde{b}_{jX} \in c_{jX} \\ b_{jX} \neq \tilde{b}_{jX}}} \sum_{\substack{(u_j^n, x_j^n), (u_j^n, \tilde{x}_j^n) \in \\ T_{2\eta_2}(U_j, X_j|q^n)}} P\left(\begin{smallmatrix} I(a^{s_j})=M_{j1}, X_j(M_{jX}, b_{jX})=x_j^n, \\ U_j(a^{s_j})=u_j^n, X_j(M_{jX}, \tilde{b}_{jX})=\tilde{x}_j^n \end{smallmatrix}\right), \\
\mathscr{T}_2 &= \sum_{\substack{a^{s_j}, \tilde{a}^{s_j} \in \mathcal{U}^{s_j} \\ a^{s_j} \neq \tilde{a}^{s_j}}} \sum_{b_{jX} \in c_{jX}} \sum_{\substack{(u_j^n, x_j^n), (\tilde{u}_j^n, x_j^n) \in \\ T_{2\eta_2}(U_j, X_j|q^n)}} P\left(\begin{smallmatrix} I(a^{s_j})=M_{j1}, I(\tilde{a}^{s_j})=M_{j1}, U_j(a^{s_j})=u_j^n, \\ X_j(M_{jX}, b_{jX})=x_j^n, U_j(\tilde{a}^{s_j})=\tilde{u}_j^n \end{smallmatrix}\right), \\
\mathscr{T}_3 &= \sum_{\substack{a^{s_j}, \tilde{a}^{s_j} \in \mathcal{U}^{s_j} \\ a^{s_j} \neq \tilde{a}^{s_j}}} \sum_{\substack{b_{jX}, \tilde{b}_{jX} \in c_{jX} \\ b_{jX} \neq \tilde{b}_{jX}}} \sum_{\substack{(u_j^n, x_j^n), (\tilde{u}_j^n, \tilde{x}_j^n) \in \\ T_{2\eta_2}(U_j, X_j|q^n)}} P\left(\begin{smallmatrix} I(a^{s_j})=M_{j1}, X_j(M_{jX}, b_{jX})=x_j^n, U_j(a^{s_j})=u_j^n, \\ I(\tilde{a}^{s_j})=M_{j1}, X_j(M_{jX}, \tilde{b}_{jX})=\tilde{x}_j^n, U_j(\tilde{a}^{s_j})=\tilde{u}_j^n \end{smallmatrix}\right).
\end{aligned}$$

Since

$$P\left(\begin{smallmatrix} I(a^{s_j})=M_{j1}, X_j(M_{jX}, b_{jX})=x_j^n, U_j(a^{s_j})=u_j^n, \\ I(\tilde{a}^{s_j})=M_{j1}, X_j(M_{jX}, \tilde{b}_{jX})=\tilde{x}_j^n, U_j(\tilde{a}^{s_j})=\tilde{u}_j^n \end{smallmatrix}\right) = P\left(\begin{smallmatrix} I(a^{s_j})=M_{j1}, U_j(a^{s_j})=u_j^n \\ X_j(M_{jX}, b_{jX})=x_j^n \end{smallmatrix}\right) P\left(\begin{smallmatrix} I(\tilde{a}^{s_j})=M_{j1}, U_j(\tilde{a}^{s_j})=\tilde{u}_j^n, \\ X_j(M_{jX}, \tilde{b}_{jX})=\tilde{x}_j^n \end{smallmatrix}\right),$$

we have $\mathscr{T}_3 \leq \mathscr{T}_0^2$, and therefore, we have $P(\epsilon_{1j}) \leq 4\frac{\mathscr{T}_0+\mathscr{T}_1+\mathscr{T}_2}{\mathscr{T}_0^2}$. Upper bound on conditional probability of jointly typical sequences (Lemma 2.2.3(iii)) and the number of conditionally typical sequences (Lemma 2.4.2), imply existence of $N_3(\eta_2) \in \mathbb{N}$, such that for all $n \geq N_3(\eta_2)$

$$\begin{aligned}
\mathscr{T}_0 &\geq \frac{\exp\left\{-nH(X_j|Q)+4n\eta_2\right\}|c_{jX}||T_{2\eta_2}(U_j,X_j|q^n)|}{\pi^{t_j+n-s_j}} \\
\mathscr{T}_1 &\leq \frac{\exp\left\{-2nH(X_j|Q)+8n\eta_2+nH(X_j|U_j,Q)+8n\eta_2\right\}|c_{jX}|(|c_{jX}|-1)|T_{2\eta_2}(U_j,X_j|q^n)|}{\pi^{t_j+n-s_j}} \\
\mathscr{T}_2 &\leq \frac{\exp\left\{-nH(X_j|Q)+4n\eta_2+nH(U_j|X_jQ)+8n\eta_2\right\}|c_{jX}||T_{2\eta_2}(U_j,X_j|q^n)|}{\pi^{2(t_j+n-s_j)}}.
\end{aligned} \tag{C.2}$$

Substituting upper and lower bounds for $|T_{2\eta_2}(U_j,X_j|q^n)|$ (lemma 2.4.2) guarantees existence of $N_4(\eta_2) \in \mathbb{N}$ such that for all $n \geq N_4(\eta_2)$, we have

$$\begin{aligned}
P(\epsilon_{1j}) \leq\ & 4\exp\left\{-n\log\pi\left[\frac{s_j}{n}-\frac{t_j}{n}+\frac{\log|c_{jX}|}{n\log\pi}-\left(1+\frac{H(X_j|Q)}{\log\pi}-\frac{H(U_j,X_j|Q)}{\log\pi}+\frac{8\eta_2}{\log\pi}\right)\right]\right\}+ \\
& 4\exp\left\{-n\log\pi\left[\frac{s_j}{n}-\frac{t_j}{n}-\left(1-\frac{H(U_j|Q)}{\log\pi}+\frac{35\eta_2}{\log\pi}\right)\right]\right\}+4\exp\left\{-n\left[\frac{\log|c_{jX}|}{n}-32\eta_2\right]\right\}.
\end{aligned}$$

Employing bounds on $\frac{s_j}{n}, \frac{t_j}{n}, \frac{\log|C_{jX}|}{n}$ in (4.11), (4.12) and the definition of δ, we have

$$\begin{aligned}
P(\epsilon_{1j}) &\leq 4\exp\left\{-n\left[\delta-\eta_3\left(1+\log\pi\right)-8\eta_2\right]\right\}+4\exp\left\{-n\left[\delta-36\eta_2\right]\right\}+4\exp\left\{-n\left[\delta-\eta_3-32\eta_2\right]\right\} \\
&\leq 12\exp\left\{-n\left(\delta-\frac{\eta\left[36+\log\pi\right]}{2^d}\right)\right\}
\end{aligned} \tag{C.3}$$

for $n \geq N_5(\eta) := \max\left\{N_3(\eta_2), N_4(\eta_2)\right\}$. Before, we conclude this appendix, let us confirm $\mathcal{L}_j(n)$ grows exponentially with n. This would imply $\epsilon_{1j} \subseteq \epsilon_{l_j}$ and therefore $\epsilon_{1j} \cap \epsilon_{l_j}^c = \phi$, the empty set. From (C.1), (C.2), we have

$$\begin{aligned}
\mathcal{L}_j(n) &= \frac{1}{2}\mathbb{E}\left\{\phi_j(q^n,M_j)\right\} = \frac{\mathscr{T}_0}{2} \geq \frac{\exp\left\{-nH(X_j|Q)+4n\eta_2\right\}|c_{jX}||T_{2\eta_2}(U_j,X_j|q^n)|}{2\pi^{t_j+n-s_j}} \\
&\geq \frac{1}{2}\exp\left\{n\log\pi\left[\frac{s_j}{n}-\frac{t_j}{n}+\frac{\log|c_{jX}|}{n\log\pi}-\left(1+\frac{H(X_j|Q)}{\log\pi}-\frac{H(U_j,X_j|Q)}{\log\pi}\right)\right]\right\} \\
&\geq \frac{1}{2}\exp\left\{n\log\pi\left[S_j-T_j-\eta_3+\frac{K_j-\eta_3}{n\log\pi}-\left(1+\frac{H(X_j|Q)}{\log\pi}-\frac{H(U_j,X_j|Q)}{\log\pi}\right)\right]\right\} \qquad \text{(C.4)} \\
&\geq \frac{1}{2}\exp\left\{n\left[\delta-\eta_3(1+\log\pi)\right]\right\}, \qquad \text{(C.5)}
\end{aligned}$$

where (C.4) follows from (4.11), (4.12) and the choice of $(S_j,T_j,K_j,L_j : j=2,3)$. With $\eta_3 = \frac{\eta}{2^d} \leq \frac{\delta}{2^d}$, $\mathcal{L}_j(n)$ grows exponentially with n if $2^d > 1+\log\pi$.

Appendix D

Upper bound on $P(\tilde{\epsilon}_1^c \cap \epsilon_3)$

In the first step, we derive an upper bound on $P(\tilde{\epsilon}_1^c \cap \epsilon_2)$, where $\tilde{\epsilon}_1 = \epsilon_1 \cup \epsilon_l$, and

$$\epsilon_2 = \{(q^n, U_2^n(A^{s_2}), U_3^n(A^{s_3}), X_1^n(M_1), X_2^n(M_{2X}, B_{2X}), X_3^n(M_{3X}, B_{3X})) \notin T_{\eta_4}(Q, U_2, U_3, \underline{X})\}. \tag{D.1}$$

was defined in (4.15). In the second step, we employ the result of conditional frequency typicality to provide an upper bound on $P((\epsilon_1 \cup \epsilon_{l_2} \cup \epsilon_{l_3} \cup \epsilon_2)^c \cap (\epsilon_{31} \cup \epsilon_{32} \cup \epsilon_{33}))$.

As an astute reader might have guessed, the proof of first step will employ conditional independence of the triple $X_1, (U_2, X_2), (U_3, X_3)$ given Q. The proof is non-trivial because of statistical dependence of the codebooks. We begin with the definition

$$\Theta(q^n) := \left\{ \begin{array}{l} (u_2^n, u_3^n, \underline{x}^n) \in \mathcal{U}_2^n \times \mathcal{U}_3^n \times \underline{\mathcal{X}}^n : (q^n, u_j^n, x_j^n) \in T_{2\eta_2}(Q, U_j, X_j) : j = 2, 3 \\ (q^n, x_1^n) \in T_{2\eta_2}(Q, X_1), (q^n, u_2^n, u_3^n, \underline{x}^n) \notin T_{\eta_4}(Q, U_2, U_3, \underline{X})) \end{array} \right\}.$$

Observe that

$$\begin{aligned} P(\tilde{\epsilon}_1^c \cap \epsilon_2) &= \sum_{\substack{(u_2^n, u_3^n, \underline{x}^n) \\ \in \Theta(q^n)}} P\left(\begin{smallmatrix} I_j(A^{s_j}) = M_{j1}, U_j^n(A^{s_j}) = u_j^n, X_j^n(M_{jX}, B_{jX}) = x_j^n \\ \phi_j(q^n, M_j) \geq \frac{1}{2}\mathbb{E}\{\phi_j(q^n, M_j)\} : j = 2,3, X_1^n(M_1) = x_1^n \end{smallmatrix} \right) \\ &= \sum_{\substack{(u_2^n, u_3^n, \underline{x}^n) \\ \in \Theta(q^n)}} P\left(\bigcup_{a^{s_2} \in \mathcal{U}_2^{s_2}} \bigcup_{a^{s_3} \in \mathcal{U}_3^{s_3}} \bigcup_{\substack{b_{2X} \in \\ c_{2X}}} \bigcup_{\substack{b_{3X} \in \\ c_{3X}}} \left\{ \begin{smallmatrix} I_j(a^{s_j}) = M_{j1}, U_j^n(a^{s_j}) = u_j^n, X_j^n(M_{jX}, b_{jX}) = x_j^n, A^{s_j} = a^{s_j} \\ \phi_j(q^n, M_j) \geq \frac{1}{2}\mathbb{E}\{\phi_j(q^n, M_j)\}, B_{jX} = b_{jX} : j = 2,3, X_1^n(M_1) = x_1^n \end{smallmatrix} \right\} \right) \\ &\leq \sum_{\substack{(u_2^n, u_3^n, \underline{x}^n) \\ \in \Theta(q^n)}} \sum_{\substack{a^{s_2} \in \\ \mathcal{U}_2^{s_2}}} \sum_{\substack{a^{s_3} \in \\ \mathcal{U}_3^{s_3}}} \sum_{\substack{b_{2X} \in \\ c_{2X}}} \sum_{\substack{b_{3X} \in \\ c_{3X}}} P\left(\begin{smallmatrix} I_j(a^{s_j}) = M_{j1}, U_j^n(a^{s_j}) = u_j^n \\ X_j^n(M_{jX}, b_{jX}) = x_j^n, 2\phi_j(q^n, M_j) \geq \\ \mathbb{E}\{\phi_j(q^n, M_j)\} : j = 2,3, X_1^n(M_1) = x_1^n \end{smallmatrix} \right) P\left(\begin{smallmatrix} A^{s_j} = a^{s_j} \\ B_{jX} = b_{jX} \\ : j = 2,3 \end{smallmatrix} \middle| \begin{smallmatrix} I_j(a^{s_j}) = M_{j1}, U_j^n(a^{s_j}) = u_j^n \\ X_j^n(M_{jX}, b_{jX}) = x_j^n, 2\phi_j(q^n, M_j) \geq \\ \mathbb{E}\{\phi_j(q^n, M_j)\} : j = 2,3, X_1^n(M_1) = x_1^n \end{smallmatrix} \right) \end{aligned}$$

$$\leq \sum_{\substack{(u_2^n,u_3^n,\underline{x}^n)\\ \in\Theta(q^n)}} \sum_{\substack{a^{s_2}\in\\ \mathcal{U}_2^{s_2}}} \sum_{\substack{a^{s_3}\in\\ \mathcal{U}_3^{s_3}}} \sum_{\substack{b_{2X}\in\\ c_{2X}}} \sum_{\substack{b_{3X}\in\\ c_{3X}}} P\left(\begin{smallmatrix} I_j(a^{s_j})=M_{j1}, U_j^n(a^{s_j})=u_j^n \\ X_j^n(M_{jX},b_{jX})=x_j^n,:j=2,3 \\ X_1^n(M_1)=x_1^n \end{smallmatrix}\right) \prod_{j=2}^{3} P\left(\begin{smallmatrix} A^{s_j}=a^{s_j} \\ B_{jX}=b_{jX} \end{smallmatrix}\middle|\begin{smallmatrix} I_j(a^{s_j})=M_{j1} \\ \phi_j(q^n,M_j)\geq \frac{1}{2}\mathbb{E}\{\phi_j(q^n,M_j)\} \end{smallmatrix}\right). \quad \text{(D.2)}$$

Let us now evaluate a generic term in the above sum (D.2). Since the codebooks $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3, \Lambda_2, \Lambda_2$ are mutually independent, the probability of the event in question factors as

$$P\left(\begin{smallmatrix} U_j^n(a^{s_j})=u_j^n, X_j^n(M_{jX},b_{jX})=x_j^n, \\ I_j(a^{s_j})=M_{j1}:j=2,3, X_1^n(M_1)=x_1^n \end{smallmatrix}\right) = P(X_1^n(M_1)=x_1^n) P\left(\begin{smallmatrix} U_j^n(a^{s_j})=u_j^n, \\ I_j(a^{s_j})=M_{j1} \end{smallmatrix} : j=2,3\right) \prod_{j=2}^{3} P(X_j^n(M_{jX},b_{jX})=x_j^n)$$

Furthermore, (i) mutual independence of $I_j(a^{s_j}) : a^{s_j} \in \mathcal{U}_j^{s_j} : j=2,3, G_3, B_2^n, B_3^n$, (ii) uniform distribution of the indices $I_j(a^{s_j}) : a^{s_j} \in \mathcal{U}_j^{s_j} : j=2,3$ and (iii) distribution of codewords in $\mathcal{C}_j : j=1,2,3$ imply

$$P\left(\begin{smallmatrix} U_j^n(a^{s_j})=u_j^n, X_j^n(M_{jX},b_{jX})=x_j^n, \\ I_j(a^{s_j})=M_{j1}:j=2,3, X_1^n(M_1)=x_1^n \end{smallmatrix}\right) = P(U_j^n(a^{s_j})=u_j^n : j=2,3) \frac{\prod_{j=1}^{3}\prod_{t=1}^{n} p_{X_j|Q}(x_{jt}|q_t)}{\pi^{t_2+t_3}} \quad \text{(D.3)}$$

The following simple lemma enables us characterize $P(U_j^n(a^{s_j})=u_j^n : j=2,3)$.

Lemma D.0.5 *Let $s_2, s_3, n \in \mathbb{N}$ be such that $s_2 \leq s_3$. Let $G_3^T := [G_2^T \;\; G_{3/2}^T] \in \mathcal{F}_\pi^{s_3\times n}$ be a random matrix such that $G_2 \in \mathcal{F}_\pi^{s_2\times n}$ and $B_2^n, B_3^n \in \mathcal{F}_\pi^n$ be random vectors such that G_3, B_2^n, B_3^n be mutually independent and uniformly distributed over their respective range spaces. For $j=2,3$ and any $a^{s_j} \in \mathcal{F}_\pi^{s_j}$, let $U(a^{s_j}) := a^{s_j}G_j \oplus B_j^n$ be a random vector in the corresponding coset. Then $P(U_j^n(a^{s_j})=u_j^n : j=2,3) = \frac{1}{\pi^{2n}}$.* □

Proof: The proof follows from a simple counting argument. It maybe verified that for every $g_3 \in \mathcal{F}_\pi^{s_3\times n}$, there exists a unique pair of vectors $b_2^n, b_3^n \in \mathcal{F}_\pi^n$ such that $a^{s_j}g_j \oplus b_j^n = u_j^n$ for $j=2,3$. Therefore

$$|\left\{(g_3,b_2^n,b_3^n) \in \mathcal{F}_\pi^{s_3\times n} \times \mathcal{F}_\pi^n \times \mathcal{F}_\pi^n : a^{s_j}g_j \oplus b_j^n = u_j^n \text{ for } j=2,3\right\}| = \pi^{ns_3}.$$

Now employing the mutually independence and uniformly distribution of G_3, B_2^n, B_3^n, we have the probability of the event in question to be

$$\frac{|\left\{(g_3,b_2^n,b_3^n) \in \mathcal{F}_\pi^{s_3\times n} \times \mathcal{F}_\pi^n \times \mathcal{F}_\pi^n : a^{s_j}g_j \oplus b_j^n = u_j^n \text{ for } j=2,3\right\}|}{|\left\{(g_3,b_2^n,b_3^n) \in \mathcal{F}_\pi^{s_3\times n} \times \mathcal{F}_\pi^n \times \mathcal{F}_\pi^n\right\}|} = \frac{\pi^{ns_3}}{\pi^{ns_3+2n}} = \frac{1}{\pi^{2n}}.$$

■

We therefore have

$$P\left(\begin{smallmatrix} U_j^n(a^{s_j})=u_j^n, X_j^n(M_{jX},b_{jX})=x_j^n, \\ I_j(a^{s_j})=M_{j1}:j=2,3, X_1^n(M_1)=x_1^n \end{smallmatrix}\right) = \frac{\prod_{j=1}^{3}\prod_{t=1}^{n} p_{X_j|Q}(x_{jt}|q_t)}{\pi^{2n+t_2+t_3}} \leq \frac{\prod_{t=1}^{n} p_{X_1|Q}(x_{1t}|q_t)\exp\{-nH(X_2|Q)\}}{\exp\{-8n\eta_2 + nH(X_3|Q)\}\pi^{2n+t_2+t_3}} \quad \text{(D.4)}$$

Encoders 2 and 3 choose one among the jointly typical pairs uniformly at random. Hence,

$$\prod_{j=2}^{3} P\left(\begin{smallmatrix} A^{s_j}=a^{s_j} \\ B_{jX}=b_{jX} \end{smallmatrix}\Big|_{\phi_j(q^n,M_j)\geq\frac{1}{2}\mathbb{E}\{\phi_j(q^n,M_j)\}}^{I_j(a^{s_j})=M_{j1}}\right) = \frac{4}{\mathbb{E}\left\{\phi_2(q^n,M_2)\right\}\mathbb{E}\left\{\phi_3(q^n,M_3)\right\}}. \tag{D.5}$$

It maybe verified from (C.1) that

$$2\mathcal{L}_j(n) = \mathbb{E}\left\{\phi_j(q^n,M_j)\right\} \geq \pi^{s_j-t_j-n}|c_{jX}|\exp\left\{-n(H(X_j|Q)+4\eta_2)\right\}|T_{2\eta_2}(U_j,X_j|q^n)|. \tag{D.6}$$

Substituting (D.6), (D.5) and (D.4) in (D.2), we have

$$\begin{aligned} P(\tilde{\epsilon}_1^c\cap\epsilon_2) &\leq \sum_{\substack{(u_2^n,u_3^n,\underline{x}^n)\\ \in\Theta(q^n)}} \frac{\exp\{n8\eta_2\}\prod_{t=1}^n p_{X_1|Q}(x_{1t}|q_t)}{|T_{2\eta_2}(U_2,X_2|q^n)||T_{2\eta_2}(U_3,X_3|q^n)|} \\ &\leq \sum_{\substack{(u_2^n,u_3^n,\underline{x}^n)\\ \in\Theta(q^n)}} \prod_{t=1}^n p_{X_1|Q}(x_{1t}|q_t)\frac{\exp\left\{16n\eta_2 - nH(U_3,X_3|Q)\right\}}{\exp\left\{nH(U_2,X_2|Q)\right\}} \end{aligned} \tag{D.7}$$

where the last inequality follows from lower bound on size of the conditional typical set (lemma 2.4.2). We now employ the lower bound for conditional probability of jointly typical vectors. In particular,

$$\exp\left\{-nH(U_j,X_j|Q)-4n\eta_2\right\} \leq \prod_{t=1}^n p_{U_j,X_j|Q}(u_{jt},x_{jt}|q_t) \leq \exp\left\{-nH(U_j,X_j|Q)+4n\eta_2\right\} \tag{D.8}$$

for any $(u_2^n,u_3^n,\underline{x}^n)\in\Theta(q^n)$. Substituting lower bound (D.8) in (D.7), for $n\geq N_1(\eta_3)$, we have

$$\begin{aligned} P(\tilde{\epsilon}_1^c\cap\epsilon_2) &\leq \left[\sum_{\substack{(u_2^n,u_3^n,\underline{x}^n)\\ \in\Theta(q^n)}} \prod_{t=1}^n p_{X_1|Q}(x_{1t}|q_t)\prod_{j=2}^3\prod_{t=1}^n p_{U_jX_j|Q}(u_{jt}x_{jt}|q_t)\right]\exp\left\{24n\eta_2\right\} \\ &\leq \left[\sum_{\substack{(u_2^n,u_3^n,\underline{x}^n)\\ \in\Theta(q^n)}} \prod_{t=1}^n p_{X_1U_2X_2U_3X_3|Q}(x_{1t},u_{2t},x_{2t},u_{3t},x_{3t}|q_t)\right]\exp\left\{24n\eta_2\right\}, \end{aligned} \tag{D.9}$$

where (D.9) follows from conditional mutual independence of the triple $X_1,(U_2,X_2)$ and (U_3,X_3) given Q. We now employ the exponential upper bound claimed in (2.1) of lemma 2.3.1.[1] Under the condition $\eta_4\geq 4\eta_2$, a 'conditional

[1]In reality, we need a 'conditional version' of (2.1) of lemma 2.3.1. Establishing this only involves substituting the exponential upper bound stated in lemma 2.3.2 in place of the Cheybyshev inequality in lemma 2.4.1.

version' of lemma 2.3.1 guarantees existence of $N_6(\eta_4) \in \mathbb{N}$ and $\mu > 0$, such that for all $n \geq N_6(\eta_4)$,

$$\sum_{\substack{(u_2^n, u_3^n, \underline{x}^n) \\ \in \Theta(q^n)}} \prod_{t=1}^{n} p_{X_1U_2X_2U_3X_3|Q}(x_{1t}, u_{2t}, x_{2t}, u_{3t}, x_{3t}|q_t) \leq 2\exp\{-n^3\mu\eta_4^2\} \tag{D.10}$$

to enable us conclude

$$P(\tilde{\epsilon}_1^c \cap \epsilon_2) \leq 2\exp\{-n(n^2\mu\eta_4^2 - 2\eta_1 - 2\eta_3\log\pi - 16\eta_2)\} = 2\exp\{-n\left(n^2\mu\eta_4^2 - \frac{\eta}{2^{d-5}}\right)\} \tag{D.11}$$

for all $n \geq N_7(\eta) := \max\{N_6(\eta_4), N_1(\eta_3)\}$.

This gets us to the second step where we seek an upper bound on $P((\tilde{\epsilon}_1 \cup \epsilon_2)^c \cap \epsilon_3)$, where

$$\epsilon_3 = \{(q^n, U_2^n(A^{s_2}), U_3^n(A^{s_3}), X_1^n(M_1), X_2^n(M_{2X}, B_{2X}), X_3^n(M_{3X}, B_{3X}), \underline{Y}^n) \notin T_{2\eta_4}(Q, X_1, U_2, U_3, \underline{X}, \underline{Y})\} \tag{D.12}$$

was defined in (4.16). Deriving an upper bound on $P((\tilde{\epsilon}_1 \cup \epsilon_2)^c \cap \epsilon_3)$ employs conditional frequency typicality and the Markov chain $(Q, U_2, U_3) - \underline{X} - \underline{Y}$. In the sequel, we prove existence of $N_{12}(\eta_4) \in \mathbb{N}$ such that for all $n \geq N_8(\eta_4)$, $P(\epsilon_2^c \cap \epsilon_3) \leq \frac{\eta_4}{32}$.

If

$$\overline{\Theta}(q^n) := \left\{ \begin{array}{ll} (u_2^n, u_3^n, \underline{x}^n, \underline{y}^n) \in \mathcal{U}_2^n \times \mathcal{U}_3^n \times \underline{\mathcal{X}}^n \times \underline{\mathcal{Y}}^n : & (u_2^n, u_3^n, \underline{x}^n) \in T_{\eta_4}(U_2, U_3, \underline{X}|q^n), \\ & (u_2^n, u_3^n, \underline{x}^n, \underline{y}^n) \notin T_{2\eta_4}(U_2, U_3, \underline{X}, \underline{Y}|q^n) \end{array} \right\},$$

then

$$\begin{aligned}
P(\epsilon_2^c \cap \epsilon_3) &= \sum_{\substack{(u_2^n, u_3^n, \underline{x}^n, \underline{y}^n) \\ \in \overline{\Theta}(q^n)}} P\left(U_j^n(A^{s_j}) = u_j^n, X_j^n(M_{jX}, B_{jX}) = x_j^n : j = 2,3, X_1^n(M_1) = x_1^n, \underline{Y}^n = \underline{y}^n\right) \\
&= \sum_{\substack{(u_2^n, u_3^n, \underline{x}^n, \underline{y}^n) \\ \in \overline{\Theta}(q^n)}} P\left(\begin{smallmatrix} U_j^n(A^{s_j}) = u_j^n, X_1^n(M_1) = x_1^n, \\ X_j^n(M_{jX}, B_{jX}) = x_j^n : j = 2,3, \end{smallmatrix}\right) P\left(\underline{Y}^n = \underline{y}^n \left| \begin{smallmatrix} U_j^n(A^{s_j}) = u_j^n, X_1^n(M_1) = x_1^n, \\ X_j^n(M_{jX}, B_{jX}) = x_j^n : j = 2,3, \end{smallmatrix}\right.\right) \\
&= \sum_{\substack{(u_2^n, u_3^n, \underline{x}^n, \underline{y}^n) \\ \in \overline{\Theta}(q^n)}} P\left(\begin{smallmatrix} U_j^n(A^{s_j}) = u_j^n, X_1^n(M_1) = x_1^n, \\ X_j^n(M_{jX}, B_{jX}) = x_j^n : j = 2,3, \end{smallmatrix}\right) \prod_{t=1}^{n} W_{\underline{Y}|\underline{X}}(\underline{y}_t|\underline{x}_t) \\
&= \sum_{\substack{(u_2^n, u_3^n, \underline{x}^n, \underline{y}^n) \\ \in \overline{\Theta}(q^n)}} P\left(\begin{smallmatrix} U_j^n(A^{s_j}) = u_j^n, X_1^n(M_1) = x_1^n, \\ X_j^n(M_{jX}, B_{jX}) = x_j^n : j = 2,3, \end{smallmatrix}\right) \prod_{t=1}^{n} p_{\underline{Y}|\underline{X}U_2U_3}(\underline{y}_t|\underline{x}_t, u_{2t}, u_{3t}) \qquad \text{(D.13)} \\
&\leq \sum_{\substack{(u_2^n, u_3^n, \underline{x}^n) \in \\ T_{\eta_4}(U_2, U_3, \underline{X}|q^n)}} P\left(\begin{smallmatrix} U_j^n(A^{s_j}) = u_j^n, X_1^n(M_1) = x_1^n, \\ X_j^n(M_{jX}, B_{jX}) = x_j^n : j = 2,3, \end{smallmatrix}\right) \sum_{\substack{y^n : y^n \notin \\ T_{2\eta_4}(Y|u_2^n, u_3^n, \underline{x}^n)}} \prod_{t=1}^{n} p_{\underline{Y}|\underline{X}U_2U_3}(\underline{y}_t|\underline{x}_t, u_{2t}, u_{3t}), \qquad \text{(D.14)}
\end{aligned}$$

where (D.13) follows from the Markov chain $(Q, U_2, U_3) - \underline{X} - \underline{Y}$. Once again, the lower bound on the probability of conditional typical set (lemma 2.4.1) enables us conclude the existence $N_8(\eta_4) \in \mathbb{N}$ such that for all $n \geq N_8(\eta_4)$,

$$\sum_{\substack{y^n \in \\ T_{2\eta_4}(\underline{Y}|u_2^n, u_3^n, \underline{x}^n)}} \prod_{t=1}^{n} p_{\underline{Y}|\underline{X}U_2U_3}(\underline{y}_t|\underline{x}_t, u_{2t}, u_{3t}) \leq \frac{\eta_4}{32}$$

and therefore $P(\epsilon_2^c \cap \epsilon_3) \leq \frac{\eta_4}{32}$.

Appendix E

Upper bound on $P\big((\tilde{\epsilon}_1 \cup \epsilon_2 \cup \epsilon_3)^c \cap \epsilon_{41}\big)$

In this appendix, our objective is to derive an upper bound on $P((\tilde{\epsilon}_1 \cup \epsilon_2 \cup \epsilon_3)^c \cap \epsilon_{41})$. Recall that $\tilde{\epsilon}_1 = \epsilon_1 \cup \epsilon_l$,

$$(\epsilon_1 \cup \epsilon_2 \cup \epsilon_3)^c \cap \epsilon_{41} = \bigcup_{a^{s_3} \in \mathcal{U}_3^{s_3}} \bigcup_{\hat{m}_1 \neq M_1} \left\{ \begin{pmatrix} U_j^n(A^{s_j}) : j=2,3, X_1^n(M_1), \\ X_j^n(M_{jX}, B_{jX}), : j=2,3, Y_1^n \end{pmatrix} \in \hat{T}(q^n), \begin{pmatrix} U_{\oplus}^n(a^{s_3}), Y_1^n \\ X_1^n(\hat{m}_1) \end{pmatrix} \in T_{4\eta_4}(U_2 \oplus U_3, Y_1, X_1 | q^n) \right\}.$$

where

$$\hat{T}(q^n) := \left\{ \begin{matrix} (u_2^n, u_3^n, \underline{x}^n, y_1^n) \in \\ \mathcal{U}_2^n \times \mathcal{U}_3^n \times \underline{\mathcal{X}}^n \times \mathcal{Y}_1^n \end{matrix} : \begin{matrix} (u_2^n, u_3^n, \underline{x}^n, y_1^n) \in T_{2\eta_4}(U_2, U_3, \underline{X}, Y_1 | q^n), (u_2^n, u_3^n, \underline{x}^n) \in T_{\eta_4}(U_2, U_3, \underline{X} | q^n) \\ (u_j^n, x_j^n) \in T_{2\eta_2}(U_j, X_j | q^n) : j=2,3, x_1^n \in T_{2\eta_2}(X_1 | q^n) \end{matrix} \right\}.$$

Employing the union bound, we have

$$P((\tilde{\epsilon}_1 \cup \epsilon_2 \cup \epsilon_3)^c \cap \epsilon_{41}) \leq \sum_{\substack{\hat{a}^{s_3} \in \\ \mathcal{U}_3^{s_3}}} \sum_{\substack{m_1, \hat{m}_1 \\ \hat{m}_1 \neq m_1}} \sum_{\substack{(u_2^n, u_3^n, \underline{x}^n, y_1^n) \in \\ \hat{T}(q^n)}} \sum_{\substack{(\hat{u}^n, \hat{x}_1^n) \in \\ T_{4\eta_4}(U_2 \oplus U_3, X_1 | y_1^n, q^n)}} P\left(\left\{ \begin{matrix} X_j^n(M_{jX}, B_{jX}) = x_j^n, U_j^n(A^{s_j}) = u_j^n \\ I(A^{s_j}) = M_{j1}, X_1^n(M_1) = x_1^n, U_{\oplus}(\hat{a}^{s_3}) = \hat{u}^n \\ X_1^n(\hat{m}_1) = \hat{x}_1^n, Y_1^n = y_1^n, M_1 = m_1 : j=2,3 \end{matrix} \right\} \cap \epsilon_l^c \right) \quad \text{(E.1)}$$

We evaluate a generic term in the above sum. Defining $\mathscr{S}(\hat{a}^{s_3}) := \{(a^{s_2}, a^{s_3}) \in \mathcal{U}_2^{s_2} \times \mathcal{U}_3^{s_3} : a^{s_2} 0^{s_+} \oplus a^{s_3} \neq \hat{a}^{s_3}\}$, where $s_+ := s_3 - s_2$, $\mathscr{S}^c(\hat{a}^{s_3}) := (\mathcal{U}_2^{s_2} \times \mathcal{U}_3^{s_3}) \setminus \mathscr{S}(\hat{a}^{s_3})$, and

$$E := \left\{ \begin{matrix} X_j^n(m_{jX}, b_{jX}) = x_j^n, U_j^n(a^{s_j}) = u_j^n, M_j = m_j \\ I(a^{s_j}) = m_{j1} X_1^n(m_1) = x_1^n, U_{\oplus}(\hat{a}^{s_3}) = \hat{u}^n, \\ X_1^n(\hat{m}_1) = \hat{x}_1^n, M_1 = m_1 : j=2,3, \end{matrix} \right\}$$

we have

$$P\left(\left\{\begin{smallmatrix} X_j^n(M_{jX},B_{jX})=x_j^n, U_j^n(A^{s_j})=u_j^n \\ I(A^{s_j})=M_{j1}, X_1^n(M_1)=x_1^n, U_\oplus(\hat{a}^{s_3})=\hat{u}^n \\ X_1^n(\hat{m}_1)=\hat{x}_1^n, Y_1^n=y_1^n, M_1=m_1 : j=2,3 \end{smallmatrix}\right\} \cap \epsilon_l^c\right) = \sum_{m_2,m_3}\sum_{b_{2X},b_{3X}}\sum_{\substack{(a^{s_2},a^{s_3}) \\ \in \mathscr{S}(\hat{a}^{s_3})}} P\left(E \cap \epsilon_l^c \cap \left\{\begin{smallmatrix} Y_1^n=y_1^n, A^{s_j}=a^{s_j} \\ B_{jX}=b_{jX} : j=2,3 \end{smallmatrix}\right\}\right)$$
$$+ \sum_{m_2,m_3}\sum_{b_{2X},b_{3X}}\sum_{\substack{(a^{s_2},a^{s_3}) \\ \in \mathscr{S}^c(\hat{a}^{s_3})}} P\left(E \cap \epsilon_l^c \cap \left\{\begin{smallmatrix} Y_1^n=y_1^n, A^{s_j}=a^{s_j} \\ B_{jX}=b_{jX} : j=2,3 \end{smallmatrix}\right\}\right) \tag{E.2}$$

Note that

$$P\left(Y_1^n = y_1^n \middle| E \cap \epsilon_l^c \cap \left\{\begin{smallmatrix} A^{s_j}=a^{s_j} \\ B_{jX}=b_{jX} : j=2,3 \end{smallmatrix}\right\}\right) = W_{Y_1|\underline{X}}^n(y_1^n|\underline{x}^n), \tag{E.3}$$
$$P\left(E \cap \epsilon_l^c \cap \left\{\begin{smallmatrix} A^{s_j}=a^{s_j} \\ B_{jX}=b_{jX} : j=2,3 \end{smallmatrix}\right\}\right) = P(E)P\left(\begin{smallmatrix} A^{s_j}=a^{s_j} \\ B_{jX}=b_{jX} : j=2,3 \end{smallmatrix}\middle| E \cap \epsilon_l^c\right) = P(E)\tfrac{1}{\mathcal{L}_2(n)\mathcal{L}_3(n)} \tag{E.4}$$

Moreover, for $(u_2^n, u_3^n, x_1^n, x_2^n, x_3^n, y_1^n) \in \hat{T}(q^n)$, $(\hat{u}^n, \hat{x}_1^n) \in T_{4\eta_4}(U_2 \oplus U_3, X_1|y_1^n, q^n)$, we have

$$P(E) \leq \begin{cases} \frac{P(M_j=m_j : j=2,3, M_1=m_1)}{\pi^{3n+t_2+t_3}\exp\left\{n(H(X_1|Q)+\sum_{j=1}^3 H(X_j|Q)-20\eta_4)\right\}} & \text{if } (a^{s_2},a^{s_3}) \in \mathscr{S}(\hat{a}^{s_3}), \\ \frac{P(M_{jX}=m_{jX} : j=2,3, M_1=m_1) W_{Y_1|\underline{X}}^n(y_1^n|\underline{x}^n) 1_{\{\hat{u}^n = u_2^n \oplus u_3^n\}}}{\pi^{2n+t_2+t_3}\exp\left\{n(H(X_1|Q)+\sum_{j=1}^3 H(X_j|Q)-20\eta_4)\right\}} & \text{if } (a^{s_2},a^{s_3}) \in \mathscr{S}^c(\hat{a}^{s_3}) \end{cases} \tag{E.5}$$

In deriving the above upper bounds, we have used the upper bound on conditional probability of jointly typical sequences proved in lemma 2.2.3(iii). We have also employed independence of (triple in the former and pair in the latter) codewords in the coset code. Substituting (E.3), (E.4) and (E.5), in (E.2), we have

$$P\left(\left\{\begin{smallmatrix} X_j^n(M_{jX},B_{jX})=x_j^n, U_j^n(A^{s_j})=u_j^n \\ I(A^{s_j})=M_{j1}, X_1^n(M_1)=x_1^n, U_\oplus(\hat{a}^{s_3})=\hat{u}^n \\ X_1^n(\hat{m}_1)=\hat{x}_1^n, Y_1^n=y_1^n, M_1=m_1 : j=2,3 \end{smallmatrix}\right\} \cap \epsilon_l^c\right) \leq \frac{\pi^{s_2-t_2} P(M_1=m_1) W_{Y_1|\underline{X}}^n(y_1^n|\underline{x}^n)|c_{2X}||c_{3X}| \quad \left[\frac{\pi^{s_3}}{\pi^n} + 1_{\{\hat{u}^n = u_2^n \oplus u_3^n\}}\right]}{\pi^{2n+t_3}\exp\left\{n(H(X_1|Q)+\sum_{j=1}^3 H(X_j|Q)-20\eta_4)\right\} \qquad \mathcal{L}_2(n)\mathcal{L}_3(n)}. \tag{E.6}$$

Our next step is to substitute (E.6) in (E.1). Let us restate (E.1) below as (E.7) for ease of reference.

$$P((\tilde{\epsilon}_1 \cup \epsilon_2 \cup \epsilon_3)^c \cap \epsilon_{41}) \leq \sum_{\substack{\hat{a}^{s_3} \in \\ \mathcal{U}_3^{s_3}}} \sum_{\substack{m_1,\hat{m}_1 \\ \hat{m}_1 \neq m_1}} \sum_{\substack{(u_2^n,u_3^n,\underline{x}^n,y_1^n) \in \\ \hat{T}(q^n)}} \sum_{\substack{(\hat{u}^n,\hat{x}_1^n) \in \\ T_{4\eta_4}(U_2 \oplus U_3, X_1|y_1^n,q^n)}} P\left(\left\{\begin{smallmatrix} X_j^n(M_{jX},B_{jX})=x_j^n, U_j^n(A^{s_j})=u_j^n \\ I(A^{s_j})=M_{j1}, X_1^n(M_1)=x_1^n, U_\oplus(\hat{a}^{s_3})=\hat{u}^n \\ X_1^n(\hat{m}_1)=\hat{x}_1^n, Y_1^n=y_1^n, M_1=m_1 : j=2,3 \end{smallmatrix}\right\} \cap \epsilon_l^c\right) \tag{E.7}$$

We do some spade work before we substitute (E.6) in (E.7). (E.6) is a sum of two terms. The first term is not dependent on the arguments of the outermost summation in (E.7). Moreover, lemma 2.4.2 guarantees existence of $N_9(\eta_4) \in \mathbb{N}$ such that for all $n \geq N_9(\eta_4)$, we have $|T_{4\eta_4}(U_2 \oplus U_3, X_1|y_1^n, q^n)| \leq \exp\{n(H(U_2 \oplus U_3, X_1|Y_1, Q)) + 8\eta_4\}$. Substituting this upper bound, the summation in (E.7) corresponding to the first term in (E.6) is upper bounded by

$$\mathscr{T}_1 := \sum_{\hat{a}^{s_3}} \sum_{\substack{m_1,\hat{m}_1 \\ \hat{m}_1 \neq m_1}} \sum_{\substack{(u_2^n,u_3^n,\underline{x}^n,y_1^n) \in \\ \hat{T}(q^n)}} \frac{W_{Y_1|\underline{X}}^n(y_1^n|\underline{x}^n)}{\mathcal{L}_2(n)\mathcal{L}_3(n)} \frac{\pi^{s_2+s_3}|c_{2X}||c_{3X}|P(M_1=m_1)\exp\{n(H(U_2 \oplus U_3, X_1|Y_1, Q))\}}{\pi^{3n+t_2+t_3}\exp\left\{n(H(X_1|Q)+\sum_{j=1}^3 H(X_j|Q)-28\eta_4)\right\}}.$$

The indicator in the second term of (E.6) restricts the outermost summation in (E.7) to $\hat{x}_1^n \in T_{4\eta_4}(X_1|u_2^n \oplus u_3^n, y_1^n, q^n)$. As earlier, note that the second term is independent of $\hat{x}_1^n$. Once again, employing the lemma 2.4.2, there exists $N_{10}(\eta_4) \in \mathbb{N}$, such that for all $n \geq N_{10}(\eta_4)$, $|T_{4\eta_4}(X_1|u_2^n \oplus u_3^n, y_1^n, q^n)| \leq \exp\{n(H(X_1|U_2 \oplus U_3, Y_1, Q) + 8\eta_4)\}$. Substituting this upper bound, the summation in (E.7) corresponding to the second term in (E.6) is upper bounded by

$$\mathscr{T}_2 := \sum_{\hat{a}^{s_3}} \sum_{\substack{m_1, \hat{m}_1 \\ \hat{m}_1 \neq m_1}} \sum_{\substack{(u_2^n, u_3^n, \underline{x}^n, y_1^n) \in \\ \hat{T}(q^n)}} \frac{W_{Y_1|\underline{X}}^n(y_1^n|\underline{x}^n)}{\mathcal{L}_2(n)\mathcal{L}_3(n)} \frac{\pi^{s_2}|c_{2X}||c_{3X}|P(M_1 = m_1)\exp\{n(H(X_1|U_2 \oplus U_3, Y_1, Q))\}}{\pi^{2n+t_2+t_3}\exp\left\{n(H(X_1|Q) + \sum_{j=1}^3 H(X_j|Q) - 28\eta_4)\right\}}.$$

It can be verified that

$$\sum_{\substack{(u_2^n, u_3^n, \underline{x}^n, y_1^n) \in \\ \hat{T}(q^n)}} W_{Y_1|\underline{X}}^n(y_1^n|\underline{x}^n) \leq \min\{|T_{2\eta_2}(U_2, X_2|q^n)||T_{2\eta_2}(U_2, X_2|q^n)||T_{2\eta_2}(X_1|q^n)|, |T_{\eta_4}(U_2, U_3, \underline{X}|q^n)|\}. \tag{E.8}$$

Using (E.8) and lower bounds $\mathcal{L}_j(n) : j = 2, 3$ from (D.6), we have

$$\mathscr{T}_1 \leq 2\frac{\pi^{s_3}\exp\{-n(2H(X_1|Q) - 8\eta_2 - R_1 - \eta_3)\}|T_{2\eta_2}(X_1|q^n)|}{\pi^n\exp\{-n(H(U_2 \oplus U_3, X_1|Y_1, Q) + 28\eta_4)\}} \leq 2\frac{\pi^{s_3}\exp\{-n(H(X_1|Q) - 12\eta_2 - R_1 - \eta_3)\}}{\pi^n\exp\{-n(H(U_2 \oplus U_3, X_1|Y_1, Q) + 28\eta_4)\}},$$

where the last inequality above follows from upper bound on $|T_{2\eta_2}(X_1|q^n)|$ (Lemma 2.4.2). An identical sequence of steps yields

$$\mathscr{T}_2 \leq 2\frac{\exp\{-n(H(X_1|Q) - 28\eta_4 - R_1 - \eta_3)\}}{\exp\{-n(H(X_1|U_2 \oplus U_3, Y_1, Q) + 12\eta_2)\}}.$$

for sufficiently large n. Employing the upper bound $\frac{s_3}{n} \leq S_3 + \eta_3$, and the choice $\eta_1 = \eta_3 = \frac{\eta}{2^d}$, for sufficiently large n, we have

$$\begin{aligned} P((\tilde{\epsilon}_1 \cup \epsilon_2 \cup \epsilon_3)^c \cap \epsilon_{41}) \leq 2\exp\{n(28\eta_4 + (13 + \log\pi)\eta_3 + S_3\log\pi + R_1 - \log\pi - H(X_1|Q) + H(X_1, U_2 \oplus U_3|Y_1, Q))\} \\ +2\exp\{n(28\eta_4 + (13 + \log\pi)\eta_3 + R_1 - I(X_1; U_2 \oplus U_3, Y_1|Q))\}. \end{aligned}$$

Employing the definition of δ and $\eta_3 = \frac{\eta}{2^d}$

$$P((\tilde{\epsilon}_1 \cup \epsilon_2 \cup \epsilon_3)^c \cap \epsilon_{41}) \leq 4\exp\left\{-n\left[\delta - 28\eta_4 - \frac{\eta(13 + \log\pi)}{2^d}\right]\right\}.$$

for all $n \geq \max\{N_1(\eta_3), N_9(\eta_4), N_{10}(\eta_4)\}$.

Appendix F

Upper bound on $P\big((\tilde{\epsilon}_1 \cup \epsilon_3)^c \cap \epsilon_{4j}\big)$

While it seems that analysis of this event is similar to the error event over a point-to-point channel, and is therefore straight forward, the structure of the code lends this considerable complexity. A few remarks are in order. Firstly, the distribution induced on the codebooks does not lend the bins $C_{j1}(m_{j1}) : m_{j1} \in \mathcal{M}_{j1}$ to be statistically independent. Secondly, since the cloud center and satellite codebooks are binned, the error event needs to be carefully partitioned and analyzed separately.

In this appendix, we seek an upper bound on $P((\tilde{\epsilon}_1 \cup \epsilon_3)^c \cap \epsilon_{4j})$ for $j = 2, 3$. Let $(\epsilon_1 \cup \epsilon_3)^c \cap \epsilon_{4j} = \epsilon_{4j}^1 \cup \epsilon_{4j}^2 \cup \epsilon_{4j}^3$, where

$$\epsilon_{4j}^1 := \bigcup_{\hat{m}_{j1} \neq M_{j1}} \bigcup_{\hat{a}^{s_j} \in \mathcal{U}_j^{s_j}} \bigcup_{\hat{b}_{jX} \in c_{jX}} \left\{ \begin{smallmatrix} (q^n, U_j(\hat{a}^{s_j}), X_j(M_{jX}, \hat{b}_{jX}), Y_j^n) \in T_{4\eta_4}(Q, U_j, V_j, Y_j), \ (q^n, U_j(A^{s_j}), X_j^n(M_{jX}, B_{jX})) \in \\ T_{2\eta_2}(Q, U_j, X_j), \ I_j(\hat{a}^{s_j}) = \hat{m}_{j1}, \ (q^n, U_j^n(A^{s_j}), X_j^n(M_{jX}, B_{jX}), Y_j^n) \in T_{2\eta_4}(Q, U_j, X_j, Y_j) \end{smallmatrix} \right\},$$

$$\epsilon_{4j}^2 := \bigcup_{\hat{m}_{jX} \neq M_{jX}} \bigcup_{a^{s_j} \in \mathcal{U}_j^{s_j}} \bigcup_{b_{jX} \in c_{jX}} \left\{ \begin{smallmatrix} (q^n, U_j(a^{s_j}), X_j(\hat{m}_{jX}, b_{jX}), Y_j^n) \in T_{4\eta_4}(Q, U_j, V_j, Y_j), \ (q^n, U_j(A^{s_j}), X_j^n(M_{jX}, B_{jX})) \in \\ T_{2\eta_2}(Q, U_j, X_j), I_j(a^{s_j}) = M_{j1}, \ (q^n, U_j^n(A^{s_j}), X_j^n(M_{jX}, B_{jX}), Y_j^n) \in T_{2\eta_4}(Q, U_j, X_j, Y_j) \end{smallmatrix} \right\},$$

$$\epsilon_{4j}^3 := \bigcup_{\substack{\hat{m}_{j1} \neq \\ M_{j1}}} \bigcup_{\substack{\hat{m}_{jX} \neq \\ M_{jX}}} \bigcup_{a^{s_j} \in \mathcal{U}_j^{s_j}} \bigcup_{b_{jX} \in c_{jX}} \left\{ \begin{smallmatrix} (q^n, U_j(a^{s_j}), X_j(\hat{m}_{jX}, b_{jX}), Y_j^n) \in T_{4\eta_4}(Q, U_j, V_j, Y_j), \ (q^n, U_j(A^{s_j}), X_j^n(M_{jX}, B_{jX})) \in \\ T_{2\eta_2}(Q, U_j, X_j), \ I_j(a^{s_j}) = \hat{m}_{j1}, \ (q^n, U_j^n(A^{s_j}), X_j^n(M_{jX}, B_{jX}), Y_j^n) \in T_{2\eta_4}(Q, U_j, X_j, Y_j) \end{smallmatrix} \right\}.$$

The event of interest is $\epsilon_l^c \cap (\epsilon_{4j}^1 \cup \epsilon_{4j}^2 \cup \epsilon_{4j}^3)$. Since $\epsilon_{l_j}^c \cap (\epsilon_{4j}^1 \cup \epsilon_{4j}^2 \cup \epsilon_{4j}^3)$ contains the above error event, it suffices to

derive upper bounds on $P(\epsilon_{l_j}^c \cap \epsilon_{4j}^1), P(\epsilon_{l_j}^c \cap \epsilon_{4j}^2), P(\epsilon_{l_j}^c \cap \epsilon_{4j}^3)$. We begin by studying $P(\epsilon_{l_j}^c \cap \epsilon_{4j}^1)$. Defining,

$$\tilde{T}(q^n) := \left\{(u_j^n, x_j^n, y_j^n) \in T_{2\eta_4}(U_j, X_j, Y_j|q^n) : (u_j^n, x_j^n) \in T_{2\eta_2}(U_j, X_j|q^n)\right\}, \text{ we have} \quad \text{(F.1)}$$

$$P(\epsilon_{l_j}^c \cap \epsilon_{4j}^1) = P\left(\bigcup_{\substack{m_{j1}, \hat{m}_{j1} \in \mathcal{M}_{j1} \\ m_{j1} \neq \hat{m}_{j1}}} \bigcup_{\substack{\hat{a}^{s_j} \\ \in \mathcal{U}_j^{s_j}}} \bigcup_{\substack{\hat{b}_{jX} \\ \in c_{jX}}} \bigcup_{\substack{(u_j^n, x_j^n, y_j^n) \in \\ \tilde{T}(q^n)}} \bigcup_{\substack{(\hat{u}_j^n, \hat{x}_j^n) \in \\ T_{4\eta_4}(U_j, X_j|y_j^n, q^n)}} \left\{\begin{array}{c} U_j(A^{s_j}) = u_j^n, U_j(\hat{a}^{s_j}) = \hat{u}_j^n, M_{j1} = m_{j1} \\ I_j(A^{s_j}) = m_{j1}, Y_j^n = y_j^n, I_j(\hat{a}^{s_j}) = \hat{m}_{j1}, \\ X_j^n(M_{jX}, B_{jX}) = x_j^n, X_j^n(M_{jX}, \hat{b}_{jX}) = \hat{x}_j^n \end{array}\right\} \cap \epsilon_{l_j}^c\right)$$

$$\leq \sum_{\substack{m_{j1}, \hat{m}_{j1} \in \mathcal{M}_{j1} \\ m_{j1} \neq \hat{m}_{j1}}} \sum_{\substack{\hat{a}^{s_j} \\ \in \mathcal{U}_j^{s_j}}} \sum_{\substack{\hat{b}_{jX} \\ \in c_{jX}}} \sum_{\substack{(u_j^n, x_j^n, y_j^n) \in \\ \tilde{T}(q^n)}} \sum_{\substack{(\hat{u}_j^n, \hat{x}_j^n) \in \\ T_{4\eta_4}(U_j, X_j|y_j^n, q^n)}} P\left(\left\{\begin{array}{c} U_j(A^{s_j}) = u_j^n, U_j(\hat{a}^{s_j}) = \hat{u}_j^n, M_{j1} = m_{j1} \\ I_j(A^{s_j}) = m_{j1}, Y_j^n = y_j^n, I_j(\hat{a}^{s_j}) = \hat{m}_{j1}, \\ X_j^n(M_{jX}, B_{jX}) = x_j^n, X_j^n(M_{jX}, \hat{b}_{jX}) = \hat{x}_j^n \end{array}\right\} \cap \epsilon_{l_j}^c\right). \quad \text{(F.2)}$$

We now consider two factors of generic term in the above summation. Since $X_1^n(M_1), X_{\tilde{\jmath}}^n(M_{\tilde{\jmath}X}, B_{\tilde{\jmath}X})$ is independent of the collection $U_j(A^{s_j}), U_j(\hat{a}^{s_j}), M_{j1}, I_j(A^{s_j}), I_j(\hat{a}^{s_j}), X_j^n(M_{jX}, B_{jX}), X_j^n(M_{jX}, \hat{b}_{jX})$ for any $(\hat{a}^{s_j}, \hat{b}_{jX})$, and $Y_1^n - (X_1^n(M_1), X_j^n(M_{jX}, B_{jX}) : j = 2, 3) - (U_j(A^{s_j}), U_j(\hat{a}^{s_j}), M_{j1}, I_j(A^{s_j}), I_j(\hat{a}^{s_j}), X_j^n(M_{jX}, \hat{b}_{jX}))$ is a Markov chain, we have

$$P\left(Y_j^n = y_j^n \middle| \begin{array}{c} U_j(A^{s_j}) = u_j^n, U_j(\hat{a}^{s_j}) = \hat{u}_j^n, M_{j1} = m_{j1} \\ \phi_j(q^n, M_j) \geq \mathcal{L}_j(n), I_j(A^{s_j}) = m_{j1}, I_j(\hat{a}^{s_j}) = \hat{m}_{j1}, \\ X_j^n(M_{jX}, B_{jX}) = x_j^n, X_j^n(M_{jX}, \hat{b}_{jX}) = \hat{x}_j^n \end{array}\right) = P\left(Y_j^n = y_j^n | X_j^n(M_{jX}, B_{jX}) = x_j^n\right) =: \hat{\theta}\left(y_j^n | x_j^n\right).$$

By the law of total probability, we have

$$P\left(\begin{array}{c} U_j(A^{s_j}) = u_j^n, U_j(\hat{a}^{s_j}) = \hat{u}_j^n, M_{j1} = m_{j1} \\ \phi_j(q^n, M_j) \geq \mathcal{L}_j(n), I_j(A^{s_j}) = m_{j1}, I_j(\hat{a}^{s_j}) = \hat{m}_{j1}, \\ X_j^n(M_{jX}, B_{jX}) = x_j^n, X_j^n(M_{jX}, \hat{b}_{jX}) = \hat{x}_j^n \end{array}\right) = \sum_{m_{jX} \in \mathcal{M}_{jX}} \sum_{a^{s_j} \in \mathcal{U}_j^{s_j}} P\left(\left\{\begin{array}{c} U_j(a^{s_j}) = u_j^n, U_j(\hat{a}^{s_j}) = \hat{u}_j^n, M_j = m_j, B_{jX} = \hat{b}_{jX} \\ A^{s_j} = a^{s_j}, I_j(a^{s_j}) = m_{j1}, I_j(\hat{a}^{s_j}) = \hat{m}_{j1}, \\ X_j^n(m_{jX}, \hat{b}_{jX}) = x_j^n, X_j^n(m_{jX}, \hat{b}_{jX}) = \hat{x}_j^n \end{array}\right\} \cap \epsilon_{l_j}^c\right) +$$

$$+ \sum_{m_{jX} \in \mathcal{M}_{jX}} \sum_{a^{s_j} \in \mathcal{U}_j^{s_j}} \sum_{\substack{b_{jX} \in c_{jX} \\ b_{jX} \neq \hat{b}_{jX}}} P\left(\left\{\begin{array}{c} U_j(a^{s_j}) = u_j^n, U_j(\hat{a}^{s_j}) = \hat{u}_j^n, M_j = m_j, B_{jX} = b_{jX} \\ A^{s_j} = a^{s_j}, I_j(a^{s_j}) = m_{j1}, I_j(\hat{a}^{s_j}) = \hat{m}_{j1}, \\ X_j^n(m_{jX}, b_{jX}) = x_j^n, X_j^n(m_{jX}, \hat{b}_{jX}) = \hat{x}_j^n \end{array}\right\} \cap \epsilon_{l_j}^c\right).$$

Now recognize that a generic term of the sum in (F.2) is a product of the left hand sides of the above two identities. Before we substitute the right hand sides of the above two identities in (F.2), we simplify the terms involved in the second identity (involving the two sums). Denoting

$$E^1 := \left\{\begin{array}{c} U_j(a^{s_j}) = u_j^n, U_j(\hat{a}^{s_j}) = \hat{u}_j^n, M_j = m_j \\ I_j(a^{s_j}) = m_{j1}, I_j(\hat{a}^{s_j}) = \hat{m}_{j1}, \\ X_j^n(m_{jX}, b_{jX}) = x_j^n, X_j^n(m_{jX}, \hat{b}_{jX}) = \hat{x}_j^n \end{array}\right\}, \text{ we have,}$$

$$P\left(\left\{\begin{array}{c} U_j(a^{s_j}) = u_j^n, U_j(\hat{a}^{s_j}) = \hat{u}_j^n, M_j = m_j, B_{jX} = b_{jX} \\ A^{s_j} = a^{s_j}, I_j(a^{s_j}) = m_{j1}, I_j(\hat{a}^{s_j}) = \hat{m}_{j1}, \\ X_j^n(m_{jX}, b_{jX}) = x_j^n, X_j^n(m_{jX}, \hat{b}_{jX}) = \hat{x}_j^n \end{array}\right\} \cap \epsilon_{l_j}^c\right) \leq P\left(E^1\right) P\left(\begin{array}{c} A^{s_j} = a^{s_j} \\ B_{jX} = b_{jX} \end{array} \middle| E^1 \cap \epsilon_{l_j}^c\right) \text{ where,}$$

$$P(E^1) = P\left(\begin{array}{c} M_j = m_j, I_j(a^{s_j}) = m_{j1}, I_j(\hat{a}^{s_j}) = \hat{m}_{j1}, \\ X_j^n(m_{jX}, b_{jX}) = x_j^n, X_j^n(m_{jX}, \hat{b}_{jX}) = \hat{x}^n \end{array}\right) P\left(\begin{array}{c} U_j^n(\hat{a}^{s_j}) = \hat{u}_j^n \\ U_j(a^{s_j}) = u_j^n \end{array}\right), \quad P\left(\begin{array}{c} A^{s_j} = a^{s_j} \\ B_{jX} = b_{jX} \end{array} \middle| E^1 \cap \epsilon_{l_j}^c\right) = \frac{1}{\mathcal{L}_j(n)} = \frac{2}{\mathbb{E}\{\phi_j(q^n, M_j)\}} \quad \text{(F.3)}$$

Let us work with $P(E^1)$. If $\hat{m}_{j1} \neq m_{j1}$ and $\hat{a}^{s_j} \neq a^{s_j}$, then lemma 2.2.3(iii) guarantees

$$P\left(\begin{smallmatrix} M_j=m_j, I_j(a^{s_j})=m_{j1}, I_j(\hat{a}^{s_j})=\hat{m}_{j1}, \\ X_j^n(m_{jX},b_{jX})=x_j^n, X_j^n(m_{jX},\hat{b}_{jX})=\hat{x}^n \end{smallmatrix}\right) P\left(\begin{smallmatrix} U_j^n(\hat{a}^{s_j})=\hat{u}_j^n \\ U_j(a^{s_j})=u_j^n \end{smallmatrix}\right) \leq \begin{cases} \frac{P(M_j=m_j)\exp\{-n(2H(X_j|Q))\}}{\pi^{2n+2t_j}\exp\{-n4\eta_2-n8\eta_4\}} & \text{if } \hat{b}_{jX} \neq b_{jX} \\ \frac{P(M_j=m_j)\exp\{-n(H(X_j|Q))\}}{\pi^{2n+2t_j}\exp\{-n4\eta_2\}} & \text{otherwise.} \end{cases} \tag{F.4}$$

Substituting the above observations in (F.2), we have

$$\begin{aligned} P(\epsilon_{l_j}^c \cap \epsilon_{4j}^1) \leq & \sum_{m_j \in \mathcal{M}_j} \sum_{\hat{m}_{j1} \neq m_{j1}} \sum_{\substack{a^{s_j}, \hat{a}^{s_j} \\ a^{s_j} \neq \hat{a}^{s_j}}} \sum_{\substack{b_{jX}, \hat{b}_{jX} \\ \hat{b}_{jX} \neq b_{jX}}} \sum_{\substack{(u_j^n, x_j^n, y_j^n) \in \\ \tilde{T}(q^n)}} \hat{\theta}\left(y_j^n | x_j^n\right) \sum_{\substack{(\hat{u}_j^n, \hat{x}_j^n) \in \\ T_{4\eta_4}(U_j, X_j | y_j^n, q^n)}} \frac{P(M_j=m_j)\exp\{-2nH(X_j|Q)\}}{\pi^{2n+2t_j}\exp\{-n4\eta_2-n8\eta_4\}\mathcal{L}_j(n)} + \\ & + \sum_{m_j \in \mathcal{M}_j} \sum_{\hat{m}_{j1} \neq m_{j1}} \sum_{\substack{a^{s_j}, \hat{a}^{s_j} \\ a^{s_j} \neq \hat{a}^{s_j}}} \sum_{b_{jX} \in c_{jX}} \sum_{\substack{(u_j^n, x_j^n, y_j^n) \in \\ \tilde{T}(q^n)}} \hat{\theta}\left(y_j^n | x_j^n\right) \sum_{\substack{\hat{u}_j^n \in \\ T_{4\eta_4}(U_j | x_j^n, y_j^n, q^n)}} \frac{P(M_j=m_j)\exp\{-nH(X_j|Q)\}}{\pi^{2n+2t_j}\exp\{-n4\eta_2\}\mathcal{L}_j(n)}. \end{aligned}$$

We now employ the upper bound on cardinality of the conditional frequency typical sets $T_{4\eta_4}(U_j, X_j | y_j^n, q^n)$ and $T_{4\eta_4}(U_j | x_j^n, y_j^n, q^n)$. There exists $N_{11}(\eta_4) \in \mathbb{N}$ such that for every $n \geq N_{11}(\eta_4)$,

$$|T_{4\eta_4}(U_j, X_j | y_j^n, q^n)| \leq \exp\left\{n(H(U_j, X_j | Y_j, Q) + 8\eta_4)\right\}, \quad |T_{4\eta_4}(U_j | x_j^n, y_j^n, q^n)| \leq \exp\left\{n(H(U_j | X_j, Y_j, Q) + 8\eta_4)\right\},$$

for any $(x_j^n, y_j^n, q^n) \in T_{2\eta_4}(X_j, Y_j, Q)$. Therefore, for $n \geq N_{11}(\eta_4)$, we have

$$\begin{aligned} P(\epsilon_{l_j}^c \cap \epsilon_{4j}^1) \leq & \sum_{m_j \in \mathcal{M}_j} \sum_{\hat{m}_{j1} \neq m_{j1}} \sum_{\substack{a^{s_j}, \hat{a}^{s_j} \\ a^{s_j} \neq \hat{a}^{s_j}}} \sum_{\substack{b_{jX}, \hat{b}_{jX} \\ \hat{b}_{jX} \neq b_{jX}}} \sum_{\substack{(u_j^n, x_j^n, y_j^n) \in \\ \tilde{T}(q^n)}} \hat{\theta}\left(y_j^n | x_j^n\right) \frac{P(M_j=m_j)\exp\{-2nH(X_j|Q)+n16\eta_4\}}{\pi^{2n+2t_j}\exp\{-n4\eta_2-nH(U_j,X_j|Y_j,Q)\}\mathcal{L}_j(n)} + \\ & + \sum_{m_j \in \mathcal{M}_j} \sum_{\hat{m}_{j1} \neq m_{j1}} \sum_{\substack{a^{s_j}, \hat{a}^{s_j} \\ a^{s_j} \neq \hat{a}^{s_j}}} \sum_{b_{jX} \in c_{jX}} \sum_{\substack{(u_j^n, x_j^n, y_j^n) \in \\ \tilde{T}(q^n)}} \hat{\theta}\left(y_j^n | x_j^n\right) \frac{P(M_j=m_j)\exp\{-nH(X_j|Q)+8n\eta_4\}}{\pi^{2n+2t_j}\exp\{-n4\eta_2-nH(U_j|X_j,Y_j,Q)\}\mathcal{L}_j(n)} \\ & \leq \sum_{m_j \in \mathcal{M}_j} \sum_{\hat{m}_{j1} \neq m_{j1}} \sum_{\substack{a^{s_j}, \hat{a}^{s_j} \\ a^{s_j} \neq \hat{a}^{s_j}}} \sum_{\substack{b_{jX}, \hat{b}_{jX} \\ \hat{b}_{jX} \neq b_{jX}}} \sum_{\substack{(u_j^n, x_j^n) \in \\ T_{2\eta_2}(U_j, X_j | q^n)}} \frac{P(M_j=m_j)\exp\{-2nH(X_j|Q)+n16\eta_4\}}{\pi^{2n+2t_j}\exp\{-n4\eta_2-nH(U_j,X_j|Y_j,Q)\}\mathcal{L}_j(n)} + \\ & + \sum_{m_j \in \mathcal{M}_j} \sum_{\hat{m}_{j1} \neq m_{j1}} \sum_{\substack{a^{s_j}, \hat{a}^{s_j} \\ a^{s_j} \neq \hat{a}^{s_j}}} \sum_{b_{jX} \in c_{jX}} \sum_{\substack{(u_j^n, x_j^n) \in \\ T_{2\eta_2}(U_j, X_j | q^n)}} \frac{P(M_j=m_j)\exp\{-nH(X_j|Q)+8n\eta_4\}}{\pi^{2n+2t_j}\exp\{-n4\eta_2-nH(U_j|X_j,Y_j,Q)\}\mathcal{L}_j(n)}. \end{aligned}$$

Substituting the lower bound for $\mathcal{L}_j(n)$ from (D.6) and noting that the terms in the summation do not depend on the arguments of the sum, for $n \geq N_{11}(\eta_4)$, it can be verified that

$$P(\epsilon_{l_j}^c \cap \epsilon_{4j}^1) \leq 2\frac{\pi^{s_j}\exp\{-nH(X_j|Q)+8n\eta_4+4n\eta_2\}}{\pi^n\exp\{-nH(U_j|X_j,Y_j,Q)\}}\left(\frac{\exp\{-nH(X_j|Q)+8n\eta_4\}}{\exp\{-nH(X_j|Y_j,Q)-nK_j\}}+1\right).$$

Finally, substituting the upper bound on $\frac{s_j}{n}$ in (4.11), δ and the choice $\eta_1 = \eta_2 = \eta_3 = \frac{\eta}{2^d}$, we have

$$\begin{aligned} P(\epsilon_{l_j}^c \cap \epsilon_{4j}^1) &\leq 2\exp\{-n\left[(\log\pi - H(U_j|X_j,Y_j,Q)) - S_j\log\pi - (\eta_3\log\pi + 8\eta_4 + 4\eta_2)\right]\} + \\ &+ \quad 2\exp\{-n\left[(\log\pi + H(X_j|Q) - H(U_j,X_j|Y_j,Q)) - (S_j\log\pi + K_j) - (\eta_3\log\pi + 16\eta_4 + 4\eta_2)\right]\} \\ &\leq \quad 4\exp\{-n\left[\delta - (\eta_3\log\pi + 16\eta_4 + 8\eta_2)\right]\} \leq 4\exp\left\{-n\left(\delta - \left(\frac{\eta(8+\log\pi)}{2^d} + 16\eta_4\right)\right)\right\} \end{aligned} \tag{F.5}$$

for $n \geq N_{11}(\eta_4)$.

We follow a similar sequence of steps to derive an upper bound on $P(\epsilon_{4j}^2)$. Defining $\tilde{T}(q^n)$ as in (F.1), we have

$$\begin{aligned} P(\epsilon_{l_j}^c \cap \epsilon_{4j}^2) &= P\left(\bigcup_{\substack{m_{jX},\hat{m}_{jX}\in\mathcal{M}_{jX}\\ \hat{m}_{jX}\neq m_{jX}}} \bigcup_{\substack{\hat{a}^{s_j}\\ \in U_j^{s_j}}} \bigcup_{\substack{\hat{b}_{jX}\\ \in c_{jX}}} \bigcup_{\substack{(u_j^n,x_j^n,y_j^n)\in\\ \tilde{T}(q^n)}} \bigcup_{\substack{(\hat{u}_j^n,\hat{x}_j^n)\in\\ T_{4\eta_4}(U_j,X_j|y_j^n,q^n)}} \left\{\begin{smallmatrix} X_j^n(\hat{m}_{jX},\hat{b}_{jX})=\hat{x}_j^n, U_j(\hat{a}^{s_j})=\hat{u}_j^n, Y_j^n=y_j^n \\ I_j(A^{s_j})=I_j(\hat{a}^{s_j})=M_{j1}, M_{jX}=m_{jX}, \\ X_j^n(M_{jX},B_{jX})=x_j^n, U_j(A^{s_j})=u_j^n \end{smallmatrix}\right\} \cap \epsilon_{l_j}^c\right) \\ &\leq \sum_{\substack{m_{jX},\hat{m}_{jX}\in\mathcal{M}_{jX}\\ \hat{m}_{jX}\neq m_{jX}}} \sum_{\substack{\hat{a}^{s_j}\\ \in U_j^{s_j}}} \sum_{\substack{\hat{b}_{jX}\\ \in c_{jX}}} \sum_{\substack{(u_j^n,x_j^n,y_j^n)\in\\ \tilde{T}(q^n)}} \sum_{\substack{(\hat{u}_j^n,\hat{x}_j^n)\in\\ T_{4\eta_4}(U_j,X_j|y_j^n,q^n)}} P\left(\left\{\begin{smallmatrix} X_j^n(\hat{m}_{jX},\hat{b}_{jX})=\hat{x}_j^n, U_j(\hat{a}^{s_j})=\hat{u}_j^n, Y_j^n=y_j^n \\ I_j(A^{s_j})=I_j(\hat{a}^{s_j})=M_{j1}, M_{jX}=m_{jX}, \\ X_j^n(M_{jX},B_{jX})=x_j^n, U_j(A^{s_j})=u_j^n \end{smallmatrix}\right\} \cap \epsilon_{l_j}^c\right) \end{aligned} \tag{F.6}$$

We now consider two factors of a generic term in the above sum. Since $X_1^n(M_1), X_{\hat{\jmath}}^n(M_{\hat{\jmath}X}, B_{\hat{\jmath}X})$ is independent of the collection $X_j^n(\hat{m}_{jX},\hat{b}_{jX}), U_j(\hat{a}^{s_j}), I_j(A^{s_j}), I_j(\hat{a}^{s_j}), M_{jX}, X_j^n(M_{jX},B_{jX}), U_j(A^{s_j})$ for any $(\hat{a}^{s_j},\hat{b}_{jX})$ as long as $\hat{m}_{jX} \neq M_{jX}$, and $Y_1^n - (X_1^n(M_1), X_j^n(M_{jX},B_{jX}) : j = 2,3) - (X_j^n(\hat{m}_{jX},\hat{b}_{jX}), U_j(\hat{a}^{s_j}), I_j(A^{s_j}), I_j(\hat{a}^{s_j}), M_{jX}, X_j^n(M_{jX},B_{jX}), U_j(A^{s_j}))$ is a Markov chain, we have

$$P\left(Y_j^n = y_j^n \middle| \left\{\begin{smallmatrix} X_j^n(\hat{m}_{jX},\hat{b}_{jX})=\hat{x}_j^n, U_j(\hat{a}^{s_j})=\hat{u}_j^n \\ I_j(A^{s_j})=I_j(\hat{a}^{s_j})=M_{j1}, M_{jX}=m_{jX}, \\ X_j^n(M_{jX},B_{jX})=x_j^n, U_j(A^{s_j})=u_j^n \end{smallmatrix}\right\} \cap \epsilon_{l_j}^c\right) = P\left(Y_j^n = y_j^n | X_j^n(M_{jX},B_{jX}) = x_j^n\right) =: \hat{\theta}\left(y_j^n | x_j^n\right).$$

By the law of total probability, we have

$$\begin{aligned} P\left(\left\{\begin{smallmatrix} X_j^n(\hat{m}_{jX},\hat{b}_{jX})=\hat{x}_j^n, U_j(\hat{a}^{s_j})=\hat{u}_j^n \\ I_j(A^{s_j})=I_j(\hat{a}^{s_j})=M_{j1}, M_{jX}=m_{jX}, \\ X_j^n(M_{jX},B_{jX})=x_j^n, U_j(A^{s_j})=u_j^n \end{smallmatrix}\right\} \cap \epsilon_{l_j}^c\right) &= \sum_{m_{j1}\in\mathcal{M}_{j1}} \sum_{b_{jX}\in c_{jX}} P\left(\left\{\begin{smallmatrix} X_j^n(\hat{m}_{jX},\hat{b}_{jX})=\hat{x}_j^n, U_j(\hat{a}^{s_j})=\hat{u}_j^n, A^{s_j}=\hat{a}^{s_j} \\ I_j(\hat{a}^{s_j})=M_{j1}, M_j=m_j, B_{jX}=b_{jX} \\ X_j^n(m_{jX},b_{jX})=x_j^n, U_j(\hat{a}^{s_j})=u_j^n \end{smallmatrix}\right\} \cap \epsilon_{l_j}^c\right) \\ &+ \sum_{m_{j1}\in\mathcal{M}_{j1}} \sum_{b_{jX}\in c_{jX}} \sum_{\substack{a^{s_j}\in\mathcal{U}_j^{s_j}\\ a^{s_j}\neq\hat{a}^{s_j}}} P\left(\left\{\begin{smallmatrix} X_j^n(\hat{m}_{jX},\hat{b}_{jX})=\hat{x}_j^n, U_j(\hat{a}^{s_j})=\hat{u}_j^n, A^{s_j}=a^{s_j} \\ I_j(a^{s_j})=I_j(\hat{a}^{s_j})=M_{j1}, M_j=m_j, B_{jX}=b_{jX} \\ X_j^n(m_{jX},b_{jX})=x_j^n, U_j(a^{s_j})=u_j^n \end{smallmatrix}\right\} \cap \epsilon_{l_j}^c\right). \end{aligned}$$

Now recognize that a generic term of the sum in (F.6) is a product of the left hand sides of the above two identities. Before we substitute the right hand sides of the above two identities in (F.6), we simplify the terms involved in the

second identity (involving the two sums). Denoting

$$E^2 := \left\{ \begin{smallmatrix} X_j^n(\hat{m}_{jX},\hat{b}_{jX})=\hat{x}_j^n, U_j(\hat{a}^{s_j})=\hat{u}_j^n, \\ I_j(a^{s_j})=I_j(\hat{a}^{s_j})=m_{j1}, M_j=m_j \\ X_j^n(m_{jX},b_{jX})=x_j^n, U_j(a^{s_j})=u_j^n \end{smallmatrix} \right\}, \text{ we have,}$$

$$P\left(\left\{ \begin{smallmatrix} X_j^n(\hat{m}_{jX},\hat{b}_{jX})=\hat{x}_j^n, U_j(\hat{a}^{s_j})=\hat{u}_j^n, A^{s_j}=a^{s_j} \\ I_j(a^{s_j})=I_j(\hat{a}^{s_j})=m_{j1}, M_j=m_j, B_{jX}=b_{jX} \\ X_j^n(m_{jX},b_{jX})=x_j^n, U_j(a^{s_j})=u_j^n \end{smallmatrix} \right\} \cap \epsilon_{l_j}^c \right) \leq P(E^2) P\left(\begin{smallmatrix} A^{s_j}=a^{s_j} \\ B_{jX}=b_{jX} \end{smallmatrix} \middle| E^2 \cap \epsilon_{l_j}^c \right),$$

where $P\left(A^{s_j}=a^{s_j}, B_{jX}=b_{jX} | E^2 \cap \epsilon_{l_j}^c\right) = \frac{1}{\mathcal{L}_j(n)}$. Let us now evaluate $P(E^2)$. For $\hat{m}_{jX} \neq m_{jX}$, lemma 2.2.3(iii) guarantees

$$\begin{aligned} P\left(\begin{smallmatrix} X_j^n(\hat{m}_{jX},\hat{b}_{jX})=\hat{x}_j^n, U_j(\hat{a}^{s_j})=\hat{u}_j^n, \\ I_j(a^{s_j})=I_j(\hat{a}^{s_j})=m_{j1}, M_j=m_j \\ X_j^n(m_{jX},b_{jX})=x_j^n, U_j(a^{s_j})=u_j^n \end{smallmatrix} \right) &= P(M_j=m_j) P\left(\begin{smallmatrix} X_j^n(\hat{m}_{jX},\hat{b}_{jX})=\hat{x}_j^n \\ X_j^n(m_{jX},b_{jX})=x_j^n \end{smallmatrix} \right) P\left(\begin{smallmatrix} U_j(a^{s_j})=u_j^n \\ U_j(\hat{a}^{s_j})=\hat{u}_j^n \end{smallmatrix} \right) P\left(\begin{smallmatrix} I_j(a^{s_j})=m_{j1} \\ I_j(\hat{a}^{s_j})=m_{j1} \end{smallmatrix} \right) \\ &= \begin{cases} \frac{P(M_j=m_j)\exp\{-2nH(X_j|Q)\}}{\pi^{n+t_j}\exp\{-4n\eta_2-8n\eta_4\}} & \text{if } a^{s_j}=\hat{a}^{s_j}, u_j^n=\hat{u}_j^n \\ \frac{P(M_j=m_j)\exp\{-2nH(X_j|Q)\}}{\pi^{2n+2t_j}\exp\{-4n\eta_2-8n\eta_4\}} & \text{if } a^{s_j} \neq \hat{a}^{s_j} \end{cases}. \end{aligned}$$

Substituting the above observations in (F.6), we have

$$\begin{aligned} P(\epsilon_{l_j}^c \cap \epsilon_{4j}^2) \leq & \sum_{m_j \in \mathcal{M}_j} \sum_{\hat{m}_{jX} \neq m_{jX}} \sum_{\substack{b_{jX},\hat{b}_{jX} \\ \in c_{jX}}} \sum_{\substack{a^{s_j} \in \\ \mathcal{U}_j^{s_j}}} \sum_{\substack{(u_j^n,x_j^n,y_j^n) \in \\ \tilde{T}(q^n)}} \hat{\theta}\left(y_j^n|x_j^n\right) \sum_{\substack{\hat{x}_j^n \in \\ T_{4\eta_4}(X_j|u_j^n,y_j^n,q^n)}} \frac{P(M_j=m_j)\exp\{-2nH(X_j|Q)\}}{\pi^{n+t_j}\exp\{-n4\eta_2-n8\eta_4\}\mathcal{L}_j(n)} + \\ & + \sum_{m_j \in \mathcal{M}_j} \sum_{\hat{m}_{jX} \neq m_{jX}} \sum_{\substack{b_{jX},\hat{b}_{jX} \\ \in c_{jX}}} \sum_{\substack{a^{s_j},\hat{a}^{s_j} \in \mathcal{U}_j^{s_j} \\ a^{s_j} \neq \hat{a}^{s_j}}} \sum_{\substack{(u_j^n,x_j^n,y_j^n) \in \\ \tilde{T}(q^n)}} \hat{\theta}\left(y_j^n|x_j^n\right) \sum_{\substack{(\hat{u}_j^n,\hat{x}_j^n) \in \\ T_{4\eta_4}(U_j,X_j|y_j^n,q^n)}} \frac{P(M_j=m_j)\exp\{-2nH(X_j|Q)\}}{\pi^{2n+2t_j}\exp\{-n4\eta_2-n8\eta_4\}\mathcal{L}_j(n)}. \end{aligned}$$

We now employ the upper bounds on $|T_{4\eta_4}(X_j|u_j^n,y_j^n,q^n)|$ and $|T_{4\eta_4}(U_j,X_j|y_j^n,q^n)|$. There exists $N_{15}(\eta_4) \in \mathbb{N}$ such that for all $n \geq N_{15}(\eta_4)$, $|T_{4\eta_4}(X_j|u_j^n,y_j^n,q^n)| \leq \exp\{n(H(X_j|U_j,Y_j,Q)+8\eta_4)\}$ and $|T_{4\eta_4}(U_j,X_j|y_j^n,q^n)| \leq \exp\{n(H(U_j,X_j|Y_j,Q)+8\eta_4)\}$ for all $(u_j^n,y_j^n,q^n) \in T_{2\eta_4}(U_j,Y_j,Q)$. For $n \geq N_{15}(\eta_4)$, we have

$$\begin{aligned} P(\epsilon_{l_j}^c \cap \epsilon_{4j}^2) \leq & \sum_{m_j \in \mathcal{M}_j} \sum_{\hat{m}_{jX} \neq m_{jX}} \sum_{\substack{b_{jX},\hat{b}_{jX} \\ \in c_{jX}}} \sum_{\substack{a^{s_j} \in \\ \mathcal{U}_j^{s_j}}} \sum_{\substack{(u_j^n,x_j^n,y_j^n) \in \\ \tilde{T}(q^n)}} \hat{\theta}\left(y_j^n|x_j^n\right) \frac{\pi^{-n-t_j}P(M_j=m_j)\exp\{-2nH(X_j|Q)\}}{\exp\{-n4\eta_2-n16\eta_4-nH(X_j|U_j,Y_j,Q)\}\mathcal{L}_j(n)} + \\ & + \sum_{m_j \in \mathcal{M}_j} \sum_{\hat{m}_{jX} \neq m_{jX}} \sum_{\substack{b_{jX},\hat{b}_{jX} \\ \in c_{jX}}} \sum_{\substack{a^{s_j},\hat{a}^{s_j} \in \mathcal{U}_j^{s_j} \\ a^{s_j} \neq \hat{a}^{s_j}}} \sum_{\substack{(u_j^n,x_j^n,y_j^n) \in \\ \tilde{T}(q^n)}} \hat{\theta}\left(y_j^n|x_j^n\right) \frac{\pi^{-2n-2t_j}P(M_j=m_j)\exp\{-2nH(X_j|Q)\}}{\exp\{-n4\eta_2-n16\eta_4-nH(X_j,U_j|Y_j,Q)\}\mathcal{L}_j(n)} \\ \leq & \sum_{m_j \in \mathcal{M}_j} \sum_{\hat{m}_{jX} \neq m_{jX}} \sum_{\substack{b_{jX},\hat{b}_{jX} \\ \in c_{jX}}} \sum_{\substack{a^{s_j} \in \\ \mathcal{U}_j^{s_j}}} \sum_{\substack{(u_j^n,x_j^n) \in \\ T_{2\eta_2}(U_j,X_j|q^n)}} \frac{\pi^{-n-t_j}P(M_j=m_j)\exp\{-2nH(X_j|Q)\}}{\exp\{-n4\eta_2-n16\eta_4-nH(X_j|U_j,Y_j,Q)\}\mathcal{L}_j(n)} + \\ & + \sum_{m_j \in \mathcal{M}_j} \sum_{\hat{m}_{jX} \neq m_{jX}} \sum_{\substack{b_{jX},\hat{b}_{jX} \\ \in c_{jX}}} \sum_{\substack{a^{s_j},\hat{a}^{s_j} \in \mathcal{U}_j^{s_j} \\ a^{s_j} \neq \hat{a}^{s_j}}} \sum_{\substack{(u_j^n,x_j^n) \in \\ T_{2\eta_2}(U_j,X_j|q^n)}} \frac{\pi^{-2n-2t_j}P(M_j=m_j)\exp\{-2nH(X_j|Q)\}}{\exp\{-n4\eta_2-n16\eta_4-nH(X_j,U_j|Y_j,Q)\}\mathcal{L}_j(n)}. \end{aligned}$$

Substituting the lower bound for $\mathcal{L}_j(n)$ from (D.6), we have

$$\begin{aligned}
P(\epsilon_{l_j}^c \cap \epsilon_{4j}^2) &\leq 2 \sum_{m_j \in \mathcal{M}_j} \sum_{\hat{m}_{jX} \neq m_{jX}} \sum_{\substack{b_{jX}, \hat{b}_{jX} \\ \in c_{jX}}} \sum_{\substack{a^{s_j} \in \\ \mathcal{U}_j^{s_j}}} \frac{P(M_j = m_j) \exp\{-nH(X_j|Q) + n16\eta_4\}}{\pi^{s_j} \exp\{-n8\eta_2 - nH(X_j|U_j, Y_j, Q)\}|c_{jX}|} + \\
&\quad +2 \sum_{m_j \in \mathcal{M}_j} \sum_{\hat{m}_{jX} \neq m_{jX}} \sum_{\substack{b_{jX}, \hat{b}_{jX} \\ \in c_{jX}}} \sum_{\substack{a^{s_j}, \hat{a}^{s_j} \in \mathcal{U}_j^{s_j} \\ a^{s_j} \neq \hat{a}^{s_j}}} \frac{P(M_j = m_j)\pi^{-s_j} \exp\{-nH(X_j|Q) + n16\eta_4\}}{\pi^{n+t_j} \exp\{-n8\eta_2 - nH(X_j, U_j|Y_j, Q)\}|c_{jX}|} \\
&\leq 2 \sum_{m_j \in \mathcal{M}_j} \sum_{\hat{m}_{jX} \neq m_{jX}} \sum_{\substack{b_{jX}, \hat{b}_{jX} \\ \in c_{jX}}} \frac{P(M_j = m_j) \exp\{-nH(X_j|Q) + n16\eta_4\}}{\exp\{-n8\eta_2 - nH(X_j|U_j, Y_j, Q)\}|c_{jX}|} + \\
&\quad +2 \sum_{m_j \in \mathcal{M}_j} \sum_{\hat{m}_{jX} \neq m_{jX}} \sum_{\substack{b_{jX}, \hat{b}_{jX} \\ \in c_{jX}}} \frac{P(M_j = m_j)\pi^{s_j} \exp\{-nH(X_j|Q) + n16\eta_4\}}{\pi^{n+t_j} \exp\{-n8\eta_2 - nH(X_j, U_j|Y_j, Q)\}|c_{jX}|} \\
&\leq 2 \sum_{m_j \in \mathcal{M}_j} \sum_{\hat{m}_{jX} \neq m_{jX}} \frac{P(M_j = m_j) \exp\{-nH(X_j|Q) + n16\eta_4\}}{\exp\{-n8\eta_2 - nH(X_j|U_j, Y_j, Q) - nK_j\}} + \\
&\quad +2 \sum_{m_j \in \mathcal{M}_j} \sum_{\hat{m}_{jX} \neq m_{jX}} \frac{P(M_j = m_j)\pi^{s_j} \exp\{-nH(X_j|Q) + n16\eta_4\}}{\pi^{n+t_j} \exp\{-n8\eta_2 - nH(X_j, U_j|Y_j, Q) - nK_j\}} \\
&\leq 2 \frac{\exp\{-nH(X_j|Q) + nL_j + n\eta_3 + n16\eta_4\}}{\exp\{-n8\eta_2 - nH(X_j|U_j, Y_j, Q) - nK_j\}} \left[1 + \frac{\exp\{nH(U_j|Y_j, Q)\}}{\pi^{n+t_j-s_j}}\right]
\end{aligned}$$

For $n \geq N_{16}(\eta) := \max\{N_{15}(\eta_4), N_{12}(\eta_2), N_1(\eta_3)\}$, substituting (i) the upper bound on $\frac{s_j}{n}$ in (4.11), (ii) δ, and the choice $\eta_2 = \eta_3 = \frac{\eta}{2^d}$, we have

$$\begin{aligned}
P(\epsilon_{l_j}^c \cap \epsilon_{4j}^2) &\leq 2\exp\{-n(I(X_j; U_j, Y_j|Q) - K_j - L_j - [9\eta_3 + 16\eta_4])\} \\
&\quad +2\exp\left\{-n\left[\left(\begin{smallmatrix}\log\pi + H(X_j|Q) - \\ H(X_j, U_j|Y_j, Q)\end{smallmatrix}\right) - \left(\begin{smallmatrix}K_j + L_j + \\ (S_j - T_j)\log\pi\end{smallmatrix}\right) - [(9 + \log\pi)\eta_3 + 16\eta_4]\right]\right\} \\
&\leq 4\exp\left\{-n\left(\delta - \left(\frac{\eta(9 + \log\pi)}{2^d} + 16\eta_4\right)\right)\right\}. \qquad \text{(F.7)}
\end{aligned}$$

We are left to study $P(\epsilon_{4j}^3)$. Defining $\tilde{T}(q^n)$ as in (F.1), and

$$E^3 := \left\{ \begin{smallmatrix} X_j^n(\hat{m}_{jX}, \hat{b}_{jX}) = \hat{x}_j^n, U_j(\hat{a}^{s_j}) = \hat{u}_j^n \\ I_j(a^{s_j}) = m_{j1}, I_j(\hat{a}^{s_j}) = \hat{m}_{j1} \\ X_j^n(m_{jX}, b_{jX}) = x_j^n, U_j(a^{s_j}) = u_j^n, M_j = m_j \end{smallmatrix} \right\} \qquad \text{(F.8)}$$

the union bound yields

$$P(\epsilon_{l_j}^c \cap \epsilon_{4j}^3) \leq \sum_{\substack{m_{j1}, \hat{m}_{j1} \\ m_{j1} \neq \hat{m}_{j1}}} \sum_{\substack{m_{jX}, \hat{m}_{jX} \\ m_{jX} \neq \hat{m}_{jX}}} \sum_{\substack{a^{s_j}, \hat{a}^{s_j} \\ \hat{a}^{s_j} \neq a^{s_j}}} \sum_{b_{jX}, \hat{b}_{jX}} \sum_{\substack{(u_j^n, x_j^n, y_j^n) \in \\ \tilde{T}(q^n)}} \sum_{\substack{(\hat{u}_j^n, \hat{x}_j^n) \in \\ T_{4\eta_4}(U_j, X_j|y_j^n, q^n)}} P\left(\left\{Y_j^n = {\scriptstyle A^{s_j} = a^{s_j}, B_{jX} = b_{jX}}\right\} \cap E^3 \cap \epsilon_{l_j}^c\right) \qquad \text{(F.9)}$$

As earlier, we consider a generic term in the above sum and simplify the same. Observe that

$$\begin{aligned} P\left(Y_j^n = y_j^n \middle| \left\{\begin{smallmatrix} A^{s_j}=a^{s_j} \\ B_{jX}=b_{jX}\end{smallmatrix}\right\} \cap E^3 \cap \epsilon_{l_j}^c\right) &= P\left(Y_j^n = y_j^n | X_j^n(M_{jX}, B_{jX}) = x_j^n\right) =: \hat{\theta}\left(y_j^n | x_j^n\right), \\ P\left(\left\{\begin{smallmatrix} A^{s_j}=a^{s_j} \\ B_{jX}=b_{jX}\end{smallmatrix}\right\} \cap E^3 \cap \epsilon_{l_j}^c\right) &\leq P(E^3) P\left(\left\{\begin{smallmatrix} A^{s_j}=a^{s_j} \\ B_{jX}=b_{jX}\end{smallmatrix}\right\} \middle| E^3 \cap \epsilon_{l_j}^c\right) \\ &\leq \frac{P(M_j = m_j)\exp\{-2nH(X_j|Q)\}}{\pi^{2n+2t_j}\exp\{-4n\eta_2 - 8n\eta_4\}} \frac{1}{\mathcal{L}_j(n)}. \end{aligned}$$

Substituting the above observations in (F.9), we have

$$P(\epsilon_{l_j}^c \cap \epsilon_{4j}^3) \leq \sum_{\substack{m_{j1}, \hat{m}_{j1} \\ m_{j1} \neq \hat{m}_{j1}}} \sum_{\substack{m_{jX}, \hat{m}_{jX} \\ m_{jX} \neq \hat{m}_{jX}}} \sum_{\substack{a^{s_j}, \hat{a}^{s_j} \\ \hat{a}^{s_j} \neq a^{s_j}}} \sum_{b_{jX}, \hat{b}_{jX}} \sum_{\substack{(u_j^n, x_j^n, y_j^n) \in \\ \tilde{T}(q^n)}} \hat{\theta}\left(y_j^n | x_j^n\right) \sum_{\substack{(\hat{u}_j^n, \hat{x}_j^n) \in \\ T_{4\eta_4}(U_j, X_j | y_j^n, q^n)}} \frac{P(M_j = m_j)\exp\{-2nH(X_j|Q)\}}{\pi^{2n+2t_j}\exp\{-4n\eta_2 - 8n\eta_4\}\mathcal{L}_j(n)}.$$

There exists $N_{15}(\eta_4) \in \mathbb{N}$ such that for all $n \geq \max\{N_{12}(\eta_2), N_{15}(\eta_4)\}$, we have

$$|T_{4\eta_4}(U_j, X_j | y_j^n, q^n)| \leq \exp\{n(H(U_j, X_j | Y_j, Q) + 8\eta_4)\} \text{ for all } (y_j^n, q^n) \in T_{2\eta_4}(Y_j, Q)$$

and hence

$$\begin{aligned} P(\epsilon_{l_j}^c \cap \epsilon_{4j}^3) &\leq \sum_{\substack{m_{j1}, \hat{m}_{j1} \\ m_{j1} \neq \hat{m}_{j1}}} \sum_{\substack{m_{jX}, \hat{m}_{jX} \\ m_{jX} \neq \hat{m}_{jX}}} \sum_{\substack{a^{s_j}, \hat{a}^{s_j} \\ \hat{a}^{s_j} \neq a^{s_j}}} \sum_{b_{jX}, \hat{b}_{jX}} \sum_{\substack{(u_j^n, x_j^n) \in \\ T_{2\eta_2}(U_j, X_j | q^n)}} \frac{\pi^{-2n-2t_j} P(M_j = m_j)\exp\{-2nH(X_j|Q)\}}{\exp\{-n4\eta_2 - n16\eta_4 - nH(X_j, U_j|Y_j, Q)\}\mathcal{L}_j(n)} \\ &\leq 2 \sum_{\substack{m_{j1}, \hat{m}_{j1} \\ m_{j1} \neq \hat{m}_{j1}}} \sum_{\substack{m_{jX}, \hat{m}_{jX} \\ m_{jX} \neq \hat{m}_{jX}}} \frac{\pi^{s_j} P(M_j = m_j)\exp\{-nH(X_j|Q) + n16\eta_4\}}{\pi^{n+t_j}\exp\{-n8\eta_2 - nH(X_j, U_j|Y_j, Q) - nK_j\}} \\ &\leq 2\frac{\pi^{s_j} P(M_j = m_j)\exp\{-nH(X_j|Q) + n16\eta_4 + nL_j\}}{\pi^n \exp\{-n8\eta_2 - nH(X_j, U_j|Y_j, Q) - nK_j - n\eta_3\}} \\ &\leq 2\exp\left\{-n\left[\begin{pmatrix}\log\pi + H(X_j|Q) - \\ H(X_j, U_j|Y_j, Q)\end{pmatrix} - \begin{pmatrix}K_j + L_j + \\ S_j \log\pi\end{pmatrix} - \begin{pmatrix}9\eta_3 + 16\eta_4 \\ + \log\pi\eta_3\end{pmatrix}\right]\right\} \\ &\leq 2\exp\left\{-n\left(\delta - \left(\frac{\eta(9 + \log\pi)}{2^d} + 16\eta_4\right)\right)\right\}. \end{aligned} \tag{F.10}$$

We now collect all the upper bounds derived in (F.5), (F.7) and (F.10). For $n \geq \max\{N_{14}(\eta), N_{16}(\eta)\}$, we have

$$P((\tilde{\epsilon}_1 \cup \epsilon_3)^c \cap \epsilon_{4j}) \leq 10\exp\left\{-n\left(\delta - \left(\frac{\eta(9 + \log\pi)}{2^d} + 16\eta_4\right)\right)\right\} \tag{F.11}$$

Appendix G

Characterization for no rate loss in PTP-STx

We now develop the connection between upper bound (5.54) and the capacity of a point to point channel with non-causal state [7]. We only describe the relevant additive channel herein and refer the interested reader to either to [7] or [26, Chapter 7] for a detailed study. The notation employed in this section and appendix H is specific to these sections.

Consider a point to point channel with binary input and output alphabets $\mathcal{X} = \mathcal{Y} = \{0, 1\}$. The channel transition probabilities depend on a random parameter, called state that takes values in the binary alphabet $\mathcal{S} = \{0, 1\}$. The discrete time channel is time-invariant, memoryless and used without feedback. The channel is additive, i.e., if S, X and Y denote channel state, input and output respectively, then $P(Y = x \oplus s | X = x, S = s) = 1 - \delta$, where $\oplus$ denotes addition in binary field and $\delta \in (0, \frac{1}{2})$. The state is independent and identically distributed across time with $P(S = 1) = \epsilon \in (0, 1)$.[1] The input is constrained by an additive Hamming cost, i.e., the cost of transmitting $x^n \in \mathcal{X}^n$ is $\sum_{t=1}^{n} 1_{\{x_t=1\}}$ and average cost of input per symbol is constrained to be $\tau \in (0, \frac{1}{2})$.

The quantities of interest - left and right hand sides of (5.63) - are related to two scenarios with regard to knowledge of state for the above channel. In the first scenario we assume the state sequence is available to the encoder non-causally and the decoder has no knowledge of the same. In the second scenario, we assume knowledge of state is available to both the encoder and decoder non-causally. Let $\mathcal{C}_T(\tau, \delta, \epsilon), \mathcal{C}_{TR}(\tau, \delta, \epsilon)$ denote the capacity of the channel in the first and second scenarios respectively. It turns out, the left hand side of (5.63) is upper bounded by $\mathcal{C}(\tau, \delta, \epsilon)$ and the right hand side of (5.63) is $\mathcal{C}_{TR}(\tau, \delta, \epsilon)$. A necessary condition for (5.63) to hold, is therefore $\mathcal{C}_T(\tau, \delta, \epsilon) = \mathcal{C}_{TR}(\tau, \delta, \epsilon)$. For the point to point channel with non-causal state, this equality is popularly referred to

[1] Through appendices G,H we prove if $\delta, \tau \in (0, \frac{1}{2})$ and $\epsilon \in (0, 1)$, then $\alpha_T(\tau, \eta, \epsilon) < h_b(\tau * \eta) - h_b(\eta)$. This implies statement of lemma G.0.13.

as *no rate loss*. We therefore seek the condition for no rate loss.

The objective of this section and appendix H is to study the condition under which $\mathcal{C}_T(\tau,\delta,\epsilon) = \mathcal{C}_{TR}(\tau,\delta,\epsilon)$. In this section, we characterize each of these quantities, in the standard information theoretic way, in terms of a maximization of an objective function over a particular collection of probability mass functions.

We begin with a characterization of $\mathcal{C}_T(\tau,\delta,\epsilon)$ and $\mathcal{C}_{TR}(\tau,\delta,\epsilon)$.

Definition G.0.6 *Let $\mathbb{D}_T(\tau,\delta,\epsilon)$ denote the set of all probability mass functions p_{USXY} defined on $\mathcal{U}\times\mathcal{S}\times\mathcal{X}\times\mathcal{Y}$ that satisfy (i) $p_S(1) = \epsilon$, (ii) $p_{Y|XSU}(x\oplus s|x,s,u) = p_{Y|XS}(x\oplus s|x,s) = 1-\delta$, (iii) $P(X=1)\leq\tau$. For $p_{USXY}\in\mathbb{D}_T(\tau,\delta,\epsilon)$, let $\alpha_T(p_{USXY}) = I(U;Y) - I(U;S)$ and $\alpha_T(\tau,\delta,\epsilon) = \sup_{p_{USXY}\in\mathbb{D}_T(\tau,\delta,\epsilon)} \alpha_T(p_{USXY})$.*

Theorem G.0.7 $\mathcal{C}_T(\tau,\delta,\epsilon) = \alpha_T(\tau,\delta,\epsilon)$ □

This is a well known result in information theory and we refer the reader to [7] or [26, Section 7.6, Theorem 7.3] for a proof.

Definition G.0.8 *Let $\mathbb{D}_{TR}(\tau,\delta,\epsilon)$ denote the set of all probability mass functions p_{SXY} defined on $\mathcal{S}\times\mathcal{X}\times\mathcal{Y}$ that satisfy (i) $p_S(1) = \epsilon$, (ii) $p_{Y|XS}(x\oplus s|x,s) = 1-\delta$, (iii) $P(X=1)\leq\tau$. For $p_{SXY}\in\mathbb{D}_{TR}(\tau,\delta,\epsilon)$, let $\alpha_{TR}(p_{SXY}) = I(X;Y|S)$ and $\alpha_{TR}(\tau,\delta,\epsilon) = \sup_{p_{SXY}\in\mathbb{D}_{TR}(\tau,\delta,\epsilon)} \alpha_{TR}(p_{SXY})$.*

Theorem G.0.9 $\mathcal{C}_{TR}(\tau,\delta,\epsilon) = \alpha_{TR}(\tau,\delta,\epsilon)$ □

This can be argued using Shannon's characterization of point to point channel capacity [1] and we refer the reader to [26, Section 7.4.1] for a proof.

Remark G.0.10 *From the definition of $\mathcal{C}_T(\tau,\delta,\epsilon)$ and $\mathcal{C}_{TR}(\tau,\delta,\epsilon)$, it is obvious that $\mathcal{C}_T(\tau,\delta,\epsilon)\leq\mathcal{C}_{TR}(\tau,\delta,\epsilon)$, we provide an alternative argument based on theorems G.0.7, G.0.9. For any $p_{USXY}\in\mathbb{D}_T(\tau,\delta,\epsilon)$, it is easy to verify the corresponding marginal $p_{SXY}\in\mathbb{D}_{TR}(\tau,\delta,\epsilon)$ and moreover $\alpha_T(p_{USXY}) = I(U;Y) - I(U;S) \leq I(U;YS) - I(U;S) = I(U;Y|S) = H(Y|S) - H(Y|US) \leq H(Y|S) - H(Y|USX) \overset{(a)}{=} H(Y|S) - H(Y|SX) = I(X;Y|S) = \alpha_{TR}(p_{SXY}) \leq \mathcal{C}_{TR}(\tau,\delta,\epsilon)$, where (a) follows from Markov chain $U-(S,X)-Y$ ((ii) of definition G.0.6). Since this this true for every $p_{USXY}\in\mathbb{D}_T(\tau,\delta,\epsilon)$, we have $\mathcal{C}_T(\tau,\delta,\epsilon)\leq\mathcal{C}_{TR}(\tau,\delta,\epsilon)$.*

We provide an alternate characterization for $\mathcal{C}_{TR}(\tau,\delta,\epsilon)$.

Lemma G.0.11 *For $p_{USXY}\in\mathbb{D}_T(\tau,\delta,\epsilon)$, let $\beta_{TR}(p_{USXY}) = I(U;Y|S)$ and $\beta_{TR}(\tau,\delta,\epsilon) = \sup_{p_{USXY}\in\mathbb{D}_T(\tau,\delta,\epsilon)} \beta_{TR}(p_{USXY})$. Then $\beta_{TR}(\tau,\delta,\epsilon) = \alpha_{TR}(\tau,\delta,\epsilon) = \mathcal{C}_{TR}(\tau,\delta,\epsilon)$.* □

Proof: We first prove $\beta_{TR}(\tau,\delta,\epsilon)\leq\alpha_{TR}(\tau,\delta,\epsilon)$. Note that for any $p_{USXY}\in\mathbb{D}_T(\tau,\delta,\epsilon)$, the corresponding marginal $p_{SXY}\in\mathbb{D}_{TR}(\tau,\delta,\epsilon)$. Moreover, $\beta_{TR}(p_{USXY}) = I(U;Y|S) = H(Y|S) - H(Y|US) \leq H(Y|S) - H(Y|USX) \overset{(a)}{=} H(Y|S) - H(Y|SX) = I(X;Y|S) = \alpha_{TR}(p_{SXY})$, where (a) follows from Markov chain $U-(S,X)-Y$ ((ii) of definition G.0.6). Therefore, $\beta_{TR}(\tau,\delta,\epsilon)\leq\alpha_{TR}(\tau,\delta,\epsilon)$. Conversely, given $p_{SXY}\in\mathbb{D}_{TR}(\tau,\delta,\epsilon)$, define $\mathcal{U} = \{0,1\}$ and a probability mass function q_{USXY} defined on $\mathcal{U}\times\mathcal{S}\times\mathcal{X}\times\mathcal{Y}$ as $q_{USXY}(u,s,x,y) = p_{SXY}(s,x,y)1_{\{u=x\}}$.

Clearly $q_{SXY} = p_{SXY}$ and hence (i) and (iii) of definition G.0.6 are satisfied. Note that $q_{USX}(x,s,x) = p_{SX}(s,x)$, and hence $q_{Y|XSU}(y|x,s,x) = p_{Y|XS}(y|x,s) = W_{Y|XS}(y|x,s)$. Hence $q_{USXY} \in \mathbb{D}_{TR}(\tau,\delta,\epsilon)$. It is easy to verify $\beta_{TR}(q_{USXY}) = \alpha_{TR}(p_{SXY})$ and therefore $\beta_{TR}(\tau,\delta,\epsilon) \geq \alpha_{TR}(\tau,\delta,\epsilon)$. ■

We now derive a characterization of the condition under which $\mathcal{C}_{TR}(\tau,\delta,\epsilon) = \mathcal{C}_T(\tau,\delta,\epsilon)$. Towards that end, we first prove uniqueness of the pmf that achieves $\mathcal{C}_{TR}(\tau,\delta,\epsilon)$.

Lemma G.0.12 *Suppose $p_{SXY}, q_{SXY} \in \mathbb{D}_{TR}(\tau,\delta,\epsilon)$ are such that $\alpha_{TR}(p_{SXY}) = \alpha_{TR}(q_{SXY}) = \mathcal{C}_{TR}(\tau,\delta,\epsilon)$, then $p_{SXY} = q_{SXY}$. Moreover, if $\alpha_{TR}(p_{SXY}) = \mathcal{C}_{TR}(\tau,\delta,\epsilon)$, then $p_{SX} = p_S p_X$, i.e., S and X are independent.* □

Proof: Clearly, if $q_{SXY} \in \mathbb{D}_{TR}(\tau,\delta,\epsilon)$ satisfies $q_{SX} = q_S q_X$ with $q_X(1) = \tau$, then $\alpha_{TR}(q_{SXY}) = h_b(\tau * \delta) - h_b(\delta)$ and since $\mathcal{C}_{TR}(\tau,\delta,\epsilon) \leq h_b(\tau * \delta) - h_b(\delta)$,[2] we have $\mathcal{C}_{TR}(\tau,\delta,\epsilon) = h_b(\tau * \delta) - h_b(\delta)$. Let $p_{SXY} \in \mathbb{D}_{TR}(\tau,\delta,\epsilon)$ be another pmf for which $\alpha_{TR}(p_{SXY}) = h_b(\tau * \delta) - h_b(\delta)$. Let $\chi_0 := p_{X|S}(1|0)$ and $\chi_1 := p_{X|S}(1|1)$. $\alpha_{TR}(p_{SXY}) = I(X;Y|S) = H(Y|S) - H(Y|X,S) = H(X \oplus S \oplus N|S) - h_b(\delta)$. We focus on the first term

$$
\begin{aligned}
H(X \oplus S \oplus N|S) &= (1-\epsilon)H(X \oplus 0 \oplus N|S=0) + \epsilon H(X \oplus 1 \oplus N|S=1) \\
&= (1-\epsilon)h_b(\chi_0(1-\delta) + (1-\chi_0)\delta) + \epsilon h_b(\chi_1(1-\delta) + (1-\chi_1)\delta) \\
&\leq h_b((1-\epsilon)\chi_0(1-\delta) + (1-\epsilon)(1-\chi_0)\delta + \epsilon\chi_1(1-\delta) + \epsilon(1-\chi_1)\delta) \qquad \text{(G.1)} \\
&= h_b(p_X(1)(1-\delta) + (1-p_X(1))\delta) = h_b(\delta + p_X(1)(1-2\delta)) \leq h_b(\delta + \tau(1-2\delta)) = h_b(\tau * \delta) \qquad \text{(G.2)}
\end{aligned}
$$

where (G.1) follows from concavity of binary entropy function $h_b(\cdot)$ and inequality in (G.2) follows from $\delta \in (0, \frac{1}{2})$. We therefore have $\alpha_{TR}(p_{SXY}) = h_b(\tau * \delta) - h_b(\delta)$ if and only if equality holds in (G.1), (G.2). $h_b(\cdot)$ being strictly concave, equality holds in (G.1) if and only if $\epsilon \in \{0,1\}$ or $\chi_0 = \chi_1$. The range of ϵ precludes the former and therefore $\chi_0 = \chi_1$. This proves $p_{SX} = p_S p_X$ and $p_X(1) = \tau$. Given $p_{SXY} \in \mathbb{D}_{TR}(\tau,\delta,\epsilon)$, these constrains completely determine p_{SXY} and we have $p_{SXY} = q_{SXY}$. ■

Following is the main result of this section.

Lemma G.0.13 *$\mathcal{C}_{TR}(\tau,\delta,\epsilon) = \mathcal{C}_T(\tau,\delta,\epsilon)$ if and only if there exists a pmf $p_{USXY} \in \mathbb{D}_T(\tau,\delta,\epsilon)$ such that*

(i) the corresponding marginal achieves $\mathcal{C}_{TR}(\tau,\delta,\epsilon)$, i.e., $\alpha_{TR}(p_{SXY}) = \mathcal{C}_{TR}(\tau,\delta,\epsilon)$,

(ii) $S - Y - U$ is a Markov chain.

(iii) $X - (U,S) - Y$ is a Markov chain.

□

Proof: We first prove the reverse implication, i.e., the if statement. Note that $\mathcal{C}_{TR}(\tau,\delta,\epsilon) = \alpha_{TR}(p_{SXY}) = I(X;Y|S) = H(Y|S) - H(Y|XS) \overset{(a)}{=} H(Y|S) - H(Y|XSU) \overset{(b)}{=} H(Y|S) - H(Y|US) = I(U;Y|S) = I(U;YS) -$

[2] This can be easily verified using standard information theoretic arguments.

$I(U;S) \overset{(c)}{=} I(U;Y) - I(U;S) \leq \mathcal{C}_T(\tau,\delta,\epsilon)$, where (a) follows from (ii) of definition G.0.6, (b) follows from hypothesis 3) and (c) follows from hypothesis 2). We therefore have $\mathcal{C}_{TR}(\tau,\delta,\epsilon) \leq \mathcal{C}_T(\tau,\delta,\epsilon)$, and the reverse inequality follows from remark G.0.10.

Conversely, let $p_{USXY} \in \mathbb{D}_T(\tau,\delta,\epsilon)$ achieve $\mathcal{C}_T(\tau,\delta,\epsilon)$, i.e., $\alpha_T(p_{USXY}) = \mathcal{C}_T(\tau,\delta,\epsilon)$. We have $\mathcal{C}_T(\tau,\delta,\epsilon) = \alpha_T(p_{USXY}) = I(U;Y) - I(U;S) \overset{(b)}{\leq} I(U;YS) - I(U;S) = I(U;Y|S) = H(Y|S) - H(Y|US) \overset{(c)}{\leq} H(Y|S) - H(Y|USX) \overset{(a)}{=} H(Y|S) - H(Y|SX) = I(X;Y|S) = \alpha_{TR}(p_{SXY}) \leq \mathcal{C}_{TR}(\tau,\delta,\epsilon)$, where (a) follows from Markov chain $U - (S,X) - Y$ ((ii) of definition G.0.6). Equality of $\mathcal{C}_{TR}(\tau,\delta,\epsilon), \mathcal{C}_T(\tau,\delta,\epsilon)$ implies equality in (b), (c) and thus $I(U;S|Y) = 0$ and $H(Y|US) = H(Y|USX)$ and moreover $\alpha_{TR}(p_{SXY}) = \mathcal{C}_{TR}(\tau,\delta,\epsilon)$. ■

For the particular binary additive point to point channel with state, we strengthen the condition for no rate loss in the following lemma.

Lemma G.0.14 *If $p_{USXY} \in \mathbb{D}_T(\tau,\delta,\epsilon)$ satisfies*

(i) $S - Y - U$ is a Markov chain.

(ii) $X - (U,S) - Y$ is a Markov chain.

then $H(X|U,S) = 0$, or in other words, there exists a function $f : \mathcal{U} \times \mathcal{S} \to \mathcal{X}$ such that $P(X = f(U,S)) = 1$. □

Proof: We prove this by contradiction. In particular, we prove $H(X|U,S) > 0$ violates Markov chain $X - (U,S) - Y$.

If $H(X|U,S) > 0$, then $H(X \oplus S|U,S) > 0$. Indeed, $0 < H(X|U,S) \leq H(X,S|U,S) = H(X \oplus S, S|U,S) = H(S|U,S) + H(X \oplus S|U,S) = H(X \oplus S|U,S)$. Since (U,S,X) is independent of $X \oplus S \oplus Y$ and in particular, $(U,S,S \oplus X)$ is independent of $X \oplus S \oplus Y$, we have $H((X \oplus S) \oplus (X \oplus S \oplus Y)|U,S) > H(X \oplus S \oplus Y|U,S) = h_b(\delta) = H(Y|U,S,X)$, where the first inequality follows from concavity of binary entropy function. But $(X \oplus S) \oplus (X \oplus S \oplus Y) = Y$ and we have therefore proved $H(Y|U,S) > H(Y|U,S,X)$ contradicting Markov chain $X - (U,S) - Y$. ■

We summarize the conditions for no rate loss below.

Theorem G.0.15 *$\mathcal{C}_{TR}(\tau,\delta,\epsilon) = \mathcal{C}_T(\tau,\delta,\epsilon)$ if and only if there exists a pmf $p_{USXY} \in \mathbb{D}_T(\tau,\delta,\epsilon)$ such that*

(i) the corresponding marginal achieves $\mathcal{C}_{TR}(\tau,\delta,\epsilon)$, i.e., $\alpha_{TR}(p_{SXY}) = \mathcal{C}_{TR}(\tau,\delta,\epsilon)$, and in particular S and X are independent,

(ii) $S - Y - U$ is a Markov chain.

(iii) $X - (U,S) - Y$ is a Markov chain,

(iv) $H(X|U,S) = 0$, or in other words, there exists a function $f : \mathcal{U} \times \mathcal{S} \to \mathcal{X}$ such that $P(X = f(U,S)) = 1$.

□

Appendix H

The binary additive dirty PTP-STx suffers a rate loss

This section is dedicated to proving proposition 1. We begin with an upper bound on cardinality of auxiliary set involved in characterization of $\mathcal{C}_T(\tau, \delta, \epsilon)$.

Lemma H.0.16 *Consider a point to point channel with state information available at transmitter. Let $\mathcal{S}, \mathcal{X}$ and $\mathcal{Y}$ denote state, input and output alphabets respectively. Let $W_S, W_{Y|XS}$ denote pmf of state, channel transition probabilities respectively. The input is constrained with respect to a cost function $\kappa : \mathcal{X} \times \mathcal{S} \to [0, \infty)$. Let $\mathbb{D}_T(\tau)$ denote the collection of all probability mass functions p_{UXSY} defined on $\mathcal{U} \times \mathcal{X} \times \mathcal{S} \times \mathcal{Y}$, where $\mathcal{U}$ is an arbitrary set, such that (i) $p_S = W_S$, (ii) $p_{Y|XSU} = p_{Y|XS} = W_{Y|XS}$ and (iii) $\mathbb{E}\{\kappa(X, S)\} \leq \tau$. Moreover, let*

$$\overline{\mathbb{D}_T}(\tau) = \left\{ p_{UXSY} \in \mathbb{D}_T(\tau) : |\mathcal{U}| \leq \min \left\{ {}^{|\mathcal{X}|\cdot|\mathcal{S}|,}_{|\mathcal{X}|+|\mathcal{S}|+|\mathcal{Y}|-2} \right\} \right\}.$$

For $p_{UXSY} \in \mathbb{D}_T(\tau)$, let $\alpha(p_{UXSY}) = I(U;Y) - I(U;S)$. Let

$$\alpha_T(\tau) = \sup_{p_{UXSY} \in \mathbb{D}_T(\tau)} \alpha(p_{UXSY}), \qquad \overline{\alpha_T}(\tau) = \sup_{p_{UXSY} \in \overline{\mathbb{D}_T}(\tau)} \alpha(p_{UXSY}).$$

Then $\alpha_T(\tau) = \overline{\alpha_T}(\tau)$. □

Proof: The proof is based on Fenchel-Eggleston-Carathéodory [78], [26, Appendix C] theorem which is stated here for ease of reference.

Lemma H.0.17 *let $\mathcal{A}$ be a finite set and $\mathcal{Q}$ be an arbitrary set. Let $\mathcal{P}$ be a connected compact subset of pmfs on $\mathcal{A}$ and $p_{A|Q}(\cdot|q) \in \mathcal{P}$ for each $q \in \mathcal{Q}$. For $j = 1, 2, \cdots, d$ let $g_j : \mathcal{P} \to \mathbb{R}$ be continuous functions. Then for every $Q \sim F_Q$ defined on $\mathcal{Q}$, there exist a random variable $\overline{Q} \sim p_{\overline{Q}}$ with $|\overline{\mathcal{Q}}| \leq d$ and a collection of pmfs $p_{A|\overline{Q}}(\cdot|\overline{q}) \in \mathcal{P}$,*

one for each $\overline{q} \in \overline{\mathcal{Q}}$*, such that*

$$\int_{\mathcal{Q}} g_j(p_{A|Q}(a|q))dF_Q(q) = \sum_{\overline{q}\in\overline{\mathcal{Q}}} g_j(p_{A|\overline{Q}}(a|\overline{q}))p_{\overline{Q}}(\overline{q}).$$

□

The proof involves identifying $g_j : j = 1, 2\cdots, d$ such that rate achievable and cost expended are preserved. We first prove the bound $|\mathcal{U}| \leq |\mathcal{X}| \cdot |\mathcal{S}|$.

Set $\mathcal{Q} = \mathcal{U}$ and $\mathcal{A} = \mathcal{X} \times \mathcal{S}$ and $\mathcal{P}$ denote the connected compact subset of pmfs on $\mathcal{X} \times \mathcal{S}$. Without loss of generality, let $\mathcal{X} = \{1, 2, \cdots, |\mathcal{X}|\}$ and $\mathcal{S} = \{1, 2, \cdots, |\mathcal{S}|\}$. For $i = 1, 2, \cdots, |\mathcal{X}|$ and $k = 1, 2, \cdots, |\mathcal{S}| - 1$, let $g_{i,k}(\pi_{X,S}) = \pi_{X,S}(i, k)$ and $g_{l,|\mathcal{S}|}(\pi_{X,S}) = \pi_{X,S}(l, |\mathcal{S}|)$ for $l = 1, 2, \cdots, |\mathcal{X}| - 1$. Let $g_{|\mathcal{X}|\cdot|\mathcal{S}|}(\pi_{X,S}) = H(S) - H(Y)$. It can be verified that

$$\begin{aligned} g_{|\mathcal{X}|\cdot|\mathcal{S}|}(\pi_{X,S}) &= -\sum_{s\in\mathcal{S}}(\sum_{x\in\mathcal{X}} \pi_{X,S}(x, s)) \log_2(\sum_{x\in\mathcal{X}} \pi_{X,S}(x, s)) + \sum_{y\in\mathcal{Y}} \theta(y) \log_2(\theta(y)), \text{ where} \\ \theta(y) &= \sum_{(x,s)\in\mathcal{X}\times\mathcal{S}} \pi_{X,S}(x, s)W_{Y|XS}(y|x, s) \end{aligned} \tag{H.1}$$

where, is continuous. An application of lemma H.0.17 using the above set of functions, the upper bound $|\mathcal{X}| \cdot |\mathcal{S}|$ on $|\mathcal{U}|$ can be verified.

We now outline proof of upper bound $|\mathcal{X}| + |\mathcal{S}| + |\mathcal{Y}| - 2$ on $|\mathcal{U}|$. Without loss of generality, we assume $\mathcal{X} = \{1, \cdots, |\mathcal{X}|\}$, $\mathcal{S} = \{1, \cdots, |\mathcal{S}|\}$ and $\mathcal{Y} = \{1, \cdots, |\mathcal{Y}|\}$. As earlier, set $\mathcal{Q} = \mathcal{U}$ and $\mathcal{A} = \mathcal{X}\times\mathcal{S}$ and $\mathcal{P}$ denote the connected compact subset of pmfs on $\mathcal{X}\times\mathcal{S}$. For $j = 1, \cdots, |\mathcal{S}|-1$, let $g_j(\pi_{X,S}) = \sum_{x\in\mathcal{X}} \pi_{X,S}(x, j)$. For $j = |\mathcal{S}|, \cdots, |\mathcal{S}|+|\mathcal{Y}|-2$, let $g_j(\pi_{X,S}) = \sum_{(x,s)\in\mathcal{X}\times\mathcal{S}} \pi_{X,S}(x, s)W_{Y|X,S}(j - |\mathcal{S}| + 1|x, s)$. For $j = |\mathcal{S}| + |\mathcal{Y}| - 1, \cdots, |\mathcal{S}| + |\mathcal{Y}| + |\mathcal{X}| - 3$, let $g_j(\pi_{X,S}) = \sum_{s\in\mathcal{S}} \pi_{X,S}(j - |\mathcal{S}| - |\mathcal{Y}| + 2, s)$. Let $g_t(\pi_{X,S}) = H(S) - H(Y)$, i.e.,

$$g_t(\pi_{X,S}) = -\sum_{s\in\mathcal{S}}(\sum_{x\in\mathcal{X}} \pi_{X,S}(x, s)) \log_2(\sum_{x\in\mathcal{X}} \pi_{X,S}(x, s)) + \sum_{y\in\mathcal{Y}} \theta(y) \log_2(\theta(y)),$$

where $t = |\mathcal{S}| + |\mathcal{Y}| + |\mathcal{X}| - 2$, and $\theta(y)$ as is in (H.1). The rest of the proof follows by simple verification. ■

Proposition 1 *There exists no probability mass function* p_{UXSY} *defined on* $\mathcal{U}\times\mathcal{S}\times\mathcal{X}\times\mathcal{Y}$ *where* $\mathcal{U} = \{0, 1, 2, 3\}, \mathcal{X} = \mathcal{S} = \mathcal{Y} = \{0, 1\}$*, such that*

(i) X *and* S *are independent with* $P(S = 1) = \epsilon$*,* $P(X = 1) = \tau$*, where* $\epsilon \in (0, 1)$*,* $\tau \in (0, \frac{1}{2})$*,*

(ii) $p_{Y|X,S,U}(x \oplus s|x, s, u) = p_{Y|X,S}(x \oplus s|x, s) = 1 - \delta$ *for every* $(u, x, s, y) \in \mathcal{U} \times \mathcal{S} \times \mathcal{X} \times \mathcal{Y}$*, where* $\delta \in (0, \frac{1}{2})$*,*

(iii) $U - Y - S$ *and* $X - (U, S) - Y$ *are Markov chains, and*

(iv) $p_{X|US}(x|u, s) \in \{0, 1\}$ *for each* $(u, s, x) \in \mathcal{U} \times \mathcal{S} \times \mathcal{X}$*.*

USY	p_{USY}	USY	p_{USY}
000	$(1-\epsilon)(1-\theta)\beta_0$	200	$(1-\epsilon)(1-\theta)\beta_2$
001	$(1-\epsilon)\theta\gamma_0$	201	$(1-\epsilon)\theta\gamma_2$
010	$\epsilon\theta\beta_0$	210	$\epsilon\theta\beta_2$
011	$\epsilon(1-\theta)\gamma_0$	211	$\epsilon(1-\theta)\gamma_2$
100	$(1-\epsilon)(1-\theta)\beta_1$	300	$(1-\epsilon)(1-\theta)\beta_3$
101	$(1-\epsilon)\theta\gamma_1$	301	$(1-\epsilon)\theta\gamma_3$
110	$\epsilon\theta\beta_1$	310	$\epsilon\theta\beta_3$
111	$\epsilon(1-\theta)\gamma_1$	311	$\epsilon(1-\theta)\gamma_3$

Table H.1: p_{USY}

$p_{USX}(0,0,0) = p_{US}(0,0)z_0$	$p_{USX}(0,1,0) = p_{US}(0,1)z_4$
$p_{USX}(1,0,0) = p_{US}(1,0)z_1$	$p_{USX}(1,1,0) = p_{US}(1,1)z_5$
$p_{USX}(2,0,0) = p_{US}(2,0)z_2$	$p_{USX}(2,1,0) = p_{US}(2,1)z_6$
$p_{USX}(3,0,0) = p_{US}(3,0)z_3$	$p_{USX}(3,1,0) = p_{US}(3,1)z_7$

Table H.2: p_{USX}

Proof: The proof is by contradiction. If there exists such a pmf p_{USXY} then conditions 1) and 2) completely specify it's marginal on $\mathcal{S}\times\mathcal{X}\times\mathcal{Y}$ and it maybe verified that $p_{SY}(0,0) = (1-\epsilon)(1-\theta), p_{SY}(0,1) = (1-\epsilon)\theta, p_{SY}(1,0) = \epsilon\theta, p_{SY}(1,1) = \epsilon(1-\theta)$, where $\theta := \delta(1-\tau) + (1-\delta)\tau$ takes a value in $(0,1)$. Since $\epsilon \in (0,1)$, $p_{SY}(s,y) \in (0,1)$ for each $(s,y) \in \mathcal{S}\times\mathcal{Y}$. If we let $\beta_i := p_{U|Y}(i|0) : i = 0,1,2,3$ and $\gamma_j := p_{U|Y}(j|1) : j = 0,1,2,3$, then Markov chain $U - Y - S$ implies p_{USY} is as in table H.1. Since X is a function of (U,S)[1], there exist $z_i \in \{0,1\} : i = 0,1,\cdots,7$ such that entries of table H.2 hold true. Moreover, condition 4) and Markov chain $X - (U,S) - Y$ implies p_{USXY} is completely determined in terms of entries of table H.1 and $z_i : i = 0,1,\cdots,7$. For example $p_{USXY}(3,0,1,1) = p_{USY}(3,0,1)(1-z_3)$. This enables us compute marginal p_{SXY} in terms of entries of table H.1 and $z_i : i = 0,1,\cdots,7$. This marginal must satisfy conditions 1) and 2) which implies/is equivalent to the last two columns of table H.3

[1]With probability 1

being equal.

$$
\begin{aligned}
p_{SYX}(0,0,0) &= (1-\epsilon)(1-\theta)\left[\beta_0 z_0+\beta_1 z_1+\beta_2 z_2+\beta_3 z_3\right]=(1-\tau)(1-\epsilon)(1-\delta) && \text{(H.2)}\\
p_{SYX}(0,0,1) &= (1-\epsilon)(1-\theta)\left[1-\beta_0 z_0-\beta_1 z_1-\beta_2 z_2-\beta_3 z_3\right]=\tau(1-\epsilon)\delta\\
p_{SYX}(0,1,0) &= (1-\epsilon)\theta\left[\gamma_0 z_0+\gamma_1 z_1+\gamma_2 z_2+\gamma_3 z_3\right]=(1-\tau)(1-\epsilon)\delta && \text{(H.3)}\\
p_{SYX}(0,1,1) &= (1-\epsilon)\theta\left[1-\gamma_0 z_0-\gamma_1 z_1-\gamma_2 z_2-\gamma_3 z_3\right]=\tau(1-\epsilon)(1-\delta)\\
p_{SYX}(1,0,0) &= \epsilon\theta\left[\beta_0 z_4+\beta_1 z_5+\beta_2 z_6+\beta_3 z_7\right]=(1-\tau)\epsilon\delta && \text{(H.4)}\\
p_{SYX}(1,0,1) &= \epsilon\theta\left[1-\beta_0 z_4-\beta_1 z_5-\beta_2 z_6-\beta_3 z_7\right]=\tau\epsilon(1-\delta)\\
p_{SYX}(1,1,0) &= \epsilon(1-\theta)\left[\gamma_0 z_4+\gamma_1 z_5+\gamma_2 z_6+\gamma_3 z_7\right]=(1-\tau)\epsilon(1-\delta) && \text{(H.5)}\\
p_{SYX}(1,1,1) &= \epsilon(1-\theta)\left[1-\gamma_0 z_4-\gamma_1 z_5-\gamma_2 z_6-\gamma_3 z_7\right]=\tau\epsilon\delta
\end{aligned}
$$

Since $\epsilon \notin \{0,1\}$, (H.2),(H.5) imply

$$(1-\theta)\left[\beta_0 z_0+\beta_1 z_1+\beta_2 z_2+\beta_3 z_3\right]=(1-\theta)\left[\gamma_0 z_4+\gamma_1 z_5+\gamma_2 z_6+\gamma_3 z_7\right]$$

which further implies

$$\beta_0 z_0+\beta_1 z_1+\beta_2 z_2+\beta_3 z_3=\gamma_0 z_4+\gamma_1 z_5+\gamma_2 z_6+\gamma_3 z_7=:\psi_1$$

Similarly (H.3),(H.4) imply

$$\gamma_0 z_0+\gamma_1 z_1+\gamma_2 z_2+\gamma_3 z_3=\beta_0 z_4+\beta_1 z_5+\beta_2 z_6+\beta_3 z_7=:\psi_2$$

We now argue there exists no choice of values for $z_i : i=0,1\cdots,7$. Towards that end, we make a couple of observations. Firstly, we argue $\psi_1 \neq \psi_2$. Since $\epsilon \neq 1$ and $\theta \in (0,1)$, we have $\psi_1=\frac{(1-\tau)(1-\delta)}{(1-\theta)}$ and $\psi_2=\frac{(1-\tau)\delta}{\theta}$ from (H.2) and (H.3) respectively. Equating ψ_1 and ψ_2, we obtain either $\tau=1$ or $\tau=0$ or $\delta=\frac{1}{2}$. Since none of the latter conditions hold, we conclude $\psi_1 \neq \psi_2$. Secondly, one can verify $\psi_1+\psi_2-1=\frac{\delta(1-\delta)(1-2\tau)}{\theta(1-\theta)}$. Since $\delta \in (0,\frac{1}{2}), \theta \in (0,1)$ and $\tau \in (0,\frac{1}{2})$, $\psi_1+\psi_2>1$. We now eliminate the possible choices for $z_i : i=0,1\cdots,7$ through the following cases. let $m:=|\{i \in \{0,1,2,3\} : z_i=1\}|$ and $l:=|\{i \in \{4,5,6,7\} : z_i=1\}|$.

Case 1: All of z_0,z_1,z_2,z_3 or all of z_4,z_5,z_6,z_7 are equal to 0, i.e., $m=0$ or $l=0$. This implies $\psi_1=\psi_2=0$ contradicting $\psi_1 \neq \psi_2$.

Case 2: All of z_0,z_1,z_2,z_3 or all of z_4,z_5,z_6,z_7 are equal to 1, i.e., $m=4$ or $l=4$. This implies $\psi_1=\psi_2=1$ contradicting $\psi_1 \neq \psi_2$.

Cases 1 and 2 imply $m,l \in \{1,2,3\}$.

SYX	p_{SYX}	
000	$(1-\epsilon)(1-\theta)\left[\beta_0 z_0+\beta_1 z_1+\beta_2 z_2+\beta_3 z_3\right]$	$(1-\tau)(1-\epsilon)(1-\delta)$
001	$(1-\epsilon)(1-\theta)\left[1-\beta_0 z_0-\beta_1 z_1-\beta_2 z_2-\beta_3 z_3\right]$	$\tau(1-\epsilon)\delta$
010	$(1-\epsilon)\theta\left[\gamma_0 z_0+\gamma_1 z_1+\gamma_2 z_2+\gamma_3 z_3\right]$	$(1-\tau)(1-\epsilon)\delta$
011	$(1-\epsilon)\theta\left[1-\gamma_0 z_0-\gamma_1 z_1-\gamma_2 z_2-\gamma_3 z_3\right]$	$\tau(1-\epsilon)(1-\delta)$
100	$\epsilon\theta\left[\beta_0 z_4+\beta_1 z_5+\beta_2 z_6+\beta_3 z_7\right]$	$(1-\tau)\epsilon\delta$
101	$\epsilon\theta\left[1-\beta_0 z_4-\beta_1 z_5-\beta_2 z_6-\beta_3 z_7\right]$	$\tau\epsilon(1-\delta)$
110	$\epsilon(1-\theta)\left[\gamma_0 z_4+\gamma_1 z_5+\gamma_2 z_6+\gamma_3 z_7\right]$	$(1-\tau)\epsilon(1-\delta)$
111	$\epsilon(1-\theta)\left[1-\gamma_0 z_4-\gamma_1 z_5-\gamma_2 z_6-\gamma_3 z_7\right]$	$\tau\epsilon\delta$

Table H.3: Enforcing conditions 1) and 2) for p_{SXY}

UXSY	p_{UXSY}	UXSY	p_{UXSY}
0000	$(1-\epsilon)(1-\theta)\beta_0$	2000	$(1-\epsilon)(1-\theta)\beta_2$
0001	$(1-\epsilon)\theta\beta_3$	2001	$(1-\epsilon)\theta\gamma_2$
0110	$\epsilon\theta\beta_0$	2010	$\epsilon\theta\beta_2$
0111	$\epsilon(1-\theta)\beta_3$	2011	$\epsilon(1-\theta)\gamma_2$
1000	$(1-\epsilon)(1-\theta)\beta_1$	3100	$(1-\epsilon)(1-\theta)\beta_3$
1001	$(1-\epsilon)\theta\gamma_1$	3101	$(1-\epsilon)\theta\beta_0$
1010	$\epsilon\theta\beta_1$	3010	$\epsilon\theta\beta_3$
1011	$\epsilon(1-\theta)\gamma_1$	3011	$\epsilon(1-\theta)\beta_0$

Table H.4: p_{UXSY}

Case 3: $m = l = 3$. If i_1, i_2, i_3 are distinct indices in $\{0,1,2,3\}$ such that $z_{i_1} = z_{i_2} = z_{i_3} = 1$, then one among $z_{i_1+4}, z_{i_2+4}, z_{i_3+4}$ has to be 0. Else $\psi_1 = \beta_{i_1} + \beta_{i_2} + \beta_{i_3}$ and $\psi_2 = \beta_{i_1} z_{i_1+4} + \beta_{i_2} z_{i_2+4} + \beta_{i_3} z_{i_3+4} = \beta_{i_1} + \beta_{i_2} + \beta_{i_3} = \psi_1$ contradicting $\psi_1 \neq \psi_2$. Let us consider the case $z_0 = z_1 = z_2 = 1$, $z_3 = z_4 = 0$ and $z_5 = z_6 = z_7 = 1$. Table H.4 tabulates p_{USXY} for this case. We have $\psi_1 = \beta_0 + \beta_1 + \beta_2 = \gamma_1 + \gamma_2 + \gamma_3$ or equivalently $\psi_1 = 1 - \beta_3 = 1 - \gamma_0$ and $\psi_2 = \gamma_0 + \gamma_1 + \gamma_3 = \beta_1 + \beta_2 + \beta_3$ or equivalently $\psi_2 = 1 - \gamma_3 = 1 - \beta_0$. These imply $\gamma_3 = \beta_0$, $\gamma_0 = \beta_3$ which further imply $\gamma_1 + \gamma_2 = \beta_1 + \beta_2$ (since $1 = \gamma_0 + \gamma_1 + \gamma_2 + \gamma_3 = \beta_0 + \beta_1 + \beta_{2+}\beta_3$). From table H.4, one can verify

$$p_{U|XSY}(0|0,0,1) = \frac{\beta_3(1-\epsilon)\theta}{(1-\epsilon)\theta(\beta_3+\gamma_1+\gamma_2)} = \frac{\beta_3}{\beta_1+\beta_2+\beta_3},$$
$$p_{U|XS}(0|0,0) = \frac{(1-\theta)\beta_0 + \theta\beta_3}{(1-\theta)(\beta_0+\beta_1+\beta_2) + \theta(\beta_3+\gamma_1+\gamma_2)}$$

UXSY	p_{UXSY}	UXSY	p_{UXSY}
0000	$(1-\epsilon)(1-\theta)\beta_0$	2000	$(1-\epsilon)(1-\theta)\beta_2$
0001	$(1-\epsilon)\theta\gamma_0$	2001	$(1-\epsilon)\theta\beta_2$
0110	$\epsilon\theta\beta_0$	2010	$\epsilon\theta\beta_2$
0111	$\epsilon(1-\theta)\gamma_0$	2011	$\epsilon(1-\theta)\beta_2$
1000	$(1-\epsilon)(1-\theta)\beta_1$	3100	$(1-\epsilon)(1-\theta)\beta_3$
1001	$(1-\epsilon)\theta\gamma_1$	3101	$(1-\epsilon)\theta\gamma_3$
1110	$\epsilon\theta\beta_1$	3010	$\epsilon\theta\beta_3$
1111	$\epsilon(1-\theta)\gamma_1$	3011	$\epsilon(1-\theta)\gamma_3$

Table H.5: p_{UXSY}

The Markov chain $U-(X,S)-Y$ implies $p_{U|XSY}(0|0,0,1) = p_{U|XS}(0|0,0)$. Equating the right hand sides of the above equations, we obtain $(1-\theta)(\beta_0-\beta_3)(\beta_1+\beta_2)=0$. Since $\theta \neq 0$, $\beta_1+\beta_2=0$ or $\beta_0=\beta_3$. If $\beta_0=\beta_3$, then $1-\beta_3=\psi_1=\psi_2=1-\beta_0$ thus contradicting $\psi_1 \neq \psi_2$. If $\beta_1+\beta_2=0$, then $\beta_0+\beta_3=1$ implying $\psi_1+\psi_2=1$ contradicting $\psi_1+\psi_2>1$.

Case 4: $m=3, l=2$. Let us assume $z_0=z_1=z_2=z_6=z_7=1, z_3=z_4=z_5=0$. We then have $\psi_1=\beta_0+\beta_1+\beta_2=\gamma_2+\gamma_3$ and $\psi_2=\gamma_0+\gamma_1+\gamma_2=\beta_2+\beta_3$. Since $\beta_0+\beta_1+\beta_2=1-\beta_3$ and $\gamma_0+\gamma_1+\gamma_2=1-\gamma_3$, we have $\gamma_2+\gamma_3=1-\beta_3$ and $\beta_2+\beta_3=1-\gamma_3$ and therefore $\gamma_2=\beta_2$.Table H.5 tabulates p_{USXY} for this case. From table H.5, one can verify

$$p_{U|XSY}(2|0,0,1) = \tfrac{\beta_2(1-\epsilon)\theta}{(1-\epsilon)\theta(\beta_2+\gamma_0+\gamma_1)} = \tfrac{\beta_2}{\beta_2+\gamma_0+\gamma_1},$$
$$p_{U|XS}(2|0,0) = \frac{\beta_2}{(1-\theta)(\beta_0+\beta_1)+\theta(\gamma_0+\gamma_1)+\beta_2}$$

The Markov chain $U-(X,S)-Y$ implies $p_{U|XSY}(2|0,0,1) = p_{U|XS}(2|0,0)$. Equating the RHS of the above equations, we obtain $\beta_0+\beta_1=\gamma_0+\gamma_1$. This implies $\beta_2+\beta_3=\gamma_2+\gamma_3$. However $\psi_1=\beta_2+\beta_3$ and $\psi_2=\gamma_2+\gamma_3$, this contradicting $\psi \neq \psi_2$.

Let us assume $z_0=z_1=z_2=z_5=z_6=1$ and $z_3=z_4=z_7=0$. It can be verified that $\psi_1=\beta_0+\beta_1+\beta_2=\gamma_1+\gamma_2$ and $\psi_2=\gamma_0+\gamma_1+\gamma_2=\beta_1+\beta_2$. This implies $\psi_1-\psi_2=\beta_0=-\gamma_0$. Since β_0 and γ_0 are non-negative, $\beta_0=\gamma_0=0$ implying $\psi_1-\psi_2=0$, contradicting $\psi_1 \neq \psi_2$.

Case 5: $m=3, l=1$. Assume $z_0=z_1=z_2=z_4=1$, $z_3=z_5=z_6=z_7=0$. It can be verified that $\psi_1=\beta_0+\beta_1+\beta_2=\gamma_0$ and $\psi_2=\gamma_0+\gamma_1+\gamma_2=\beta_0$. Therefore $\psi_1-\psi_2=\beta_1+\beta_2$ and $\psi_2-\psi_1=\gamma_1+\gamma_2$. Since $\beta_i,\gamma_i : i \in \{0,1,2,3\}$ are non-negative, $\psi_1-\psi_2 \geq 0$ and $\psi_2-\psi_1 \geq 0$ contradicting $\psi_1 \neq \psi_2$.

Assume $z_0=z_1=z_2=z_7=1$ and $z_3=z_4=z_5=z_6=0$. In this case, $\psi_1=\beta_0+\beta_1+\beta_2=\gamma_3$, $\psi_2=\gamma_0+\gamma_1+\gamma_2=1-\gamma_3$. We have $\psi_1+\psi_2=1$ contradicting $\psi_1+\psi_2>1$.

UXSY	p_{UXSY}	UXSY	p_{UXSY}
0000	$(1-\epsilon)(1-\theta)\beta_0$	2100	$(1-\epsilon)(1-\theta)\beta_2$
0001	$(1-\epsilon)\theta\gamma_0$	2101	$(1-\epsilon)\theta\gamma_2$
0110	$\epsilon\theta\beta_0$	2010	$\epsilon\theta\beta_2$
0111	$\epsilon(1-\theta)\gamma_0$	2011	$\epsilon(1-\theta)\gamma_2$
1000	$(1-\epsilon)(1-\theta)\beta_1$	3100	$(1-\epsilon)(1-\theta)\beta_3$
1001	$(1-\epsilon)\theta\gamma_1$	3101	$(1-\epsilon)\theta\gamma_3$
1010	$\epsilon\theta\beta_1$	3110	$\epsilon\theta\beta_3$
1011	$\epsilon(1-\theta)\gamma_1$	3111	$\epsilon(1-\theta)\gamma_3$

Table H.6: p_{UXSY}

Case 6: $m = 2, l = 2$. Assume $z_0 = z_1 = z_4 = z_5 = 1$, $z_2 = z_3 = z_6 = z_7 = 0$. Note that $\psi_1 = \beta_0 + \beta_1 = \gamma_0 = \gamma_1$, $\psi_2 = \gamma_0 + \gamma_1 = \beta_0 + \beta_1$ contradicting $\psi_1 \neq \psi_2$.

Assume $z_0 = z_1 = z_6 = z_7 = 1$, $z_2 = z_3 = z_4 = z_5 = 0$. Note that $\psi_1 = \beta_0 + \beta_1 = \gamma_2 + \gamma_3$, $\psi_2 = \gamma_0 + \gamma_1 = \beta_2 + \beta_3$ contradicting $\psi_1 + \psi_2 > 1$.

Assume $z_0 = z_1 = z_5 = z_6 = 1$, $z_2 = z_3 = z_4 = z_7 = 0$. Note that $\psi_1 = \beta_0 + \beta_1 = \gamma_1 + \gamma_2$, $\psi_2 = \gamma_0 + \gamma_1 = \beta_1 + \beta_2$ and therefore $\beta_2 + \beta_3 = \gamma_0 + \gamma_3$ and $\beta_0 + \beta_3 = \gamma_2 + \gamma_3$. We observe

$$\psi_1 - \psi_2 = \beta_0 - \beta_2 = \gamma_2 - \gamma_0 \tag{H.6}$$

PMF p_{UXSY} is tabulated in H.6 for this case. Table H.6 enables us compute conditional pmf $p_{U|XSY}$ which is tabulated in table H.7. Markov chain $U - (X, S) - Y$ implies columns 2 and 4 of table H.7 are identical. This implies

$$\frac{\beta_0}{\gamma_0} \overset{(a)}{=} \frac{\beta_0+\beta_1}{\gamma_0+\gamma_1} \overset{(b)}{=} \frac{\beta_1}{\gamma_1}, \frac{\beta_2}{\gamma_2} \overset{(c)}{=} \frac{\beta_2+\beta_3}{\gamma_2+\gamma_3} \overset{(d)}{=} \frac{\beta_3}{\gamma_3}, \quad \text{and} \quad \frac{\beta_0}{\gamma_0} \overset{(e)}{=} \frac{\beta_0+\beta_3}{\gamma_0+\gamma_3} \overset{(f)}{=} \frac{\beta_3}{\gamma_3}, \tag{H.7}$$

where (a),(b),(c),(d) in (H.7) is obtained by equating rows 1, 3, 5, 7 of columns 2 and 4 respectively and (e) and (f) in (H.7) are obtained by equating rows 2 and 8 of columns 2 and 4 respectively. (H.7), enables us conclude

$$\frac{\beta_0}{\gamma_0} = \frac{\beta_1}{\gamma_1} = \frac{\beta_2}{\gamma_2} = \frac{\beta_3}{\gamma_3}.$$

Since $\beta_0 + \beta_1 + \beta_2 + \beta_3 = \gamma_0 + \gamma_1 + \gamma_2 + \gamma_3 = 1$, we have $\beta_i = \gamma_i$ for each $i \in \{0, 1, 2, 3\}$ which yields $\psi_1 = \psi_2$ in (H.6) contradicting $\psi_1 \neq \psi_2$.

UXSY	$p_{U\|XSY}$	UXSY	$p_{U\|XSY}$
0000	$\frac{\beta_0}{\beta_0+\beta_1}$	0001	$\frac{\gamma_0}{\gamma_0+\gamma_1}$
0110	$\frac{\beta_0}{\beta_0+\beta_3}$	0111	$\frac{\gamma_0}{\gamma_0+\gamma_3}$
1000	$\frac{\beta_1}{\beta_0+\beta_1}$	1001	$\frac{\gamma_1}{\gamma_0+\gamma_1}$
1010	$\frac{\beta_1}{\beta_1+\beta_2}$	1011	$\frac{\gamma_1}{\gamma_1+\gamma_2}$
2100	$\frac{\beta_2}{\beta_2+\beta_3}$	2101	$\frac{\gamma_2}{\gamma_2+\gamma_3}$
2010	$\frac{\beta_2}{\beta_1+\beta_2}$	2011	$\frac{\gamma_2}{\gamma_1+\gamma_2}$
3100	$\frac{\beta_3}{\beta_2+\beta_3}$	3101	$\frac{\gamma_3}{\gamma_2+\gamma_3}$
3110	$\frac{\beta_3}{\beta_0+\beta_3}$	3111	$\frac{\gamma_3}{\gamma_0+\gamma_3}$

Table H.7: $p_{U|XSY}$

Case 7: $m = 2, l = 1$. Assume $z_0 = z_1 = z_4 = 1, z_2 = z_3 = z_5 = z_6 = z_7 = 0$. Note that $\psi_1 = \beta_0 + \beta_1 = \gamma_0, \psi_2 = \gamma_0 + \gamma_1 = \beta_0$ and hence $\psi_1 - \psi_2 = \beta_1$ and $\psi_2 - \psi_1 = \gamma_1$. Since γ_1 and β_1 are non-negative, we have $\psi_1 = \psi_2$ contradicting $\psi_1 \neq \psi_2$.

Assume $z_0 = z_1 = z_7 = 1, z_2 = z_3 = z_4 = z_5 = z_6 = 0$. Note that $\psi_1 = \beta_0 + \beta_1 = \gamma_3, \psi_2 = \gamma_0 + \gamma_1 = \beta_3$ and hence $\psi_1 + \psi_2 = \beta_0 + \beta_1 + \beta_3 \leq 1$ contradicting $\psi_1 + \psi_2 > 1$.

Case 6: $m = 1, l = 1$. Assume $z_0 = z_4 = 1, z_1 = z_2 = z_3 = z_5 = z_6 = z_7 = 0$. Note that $\psi_1 = beta_0 = \gamma_0, \psi_2 = \gamma_0 = \beta_0$, thus contradicting $\psi_1 \neq \psi_2$.

Assume $z_0 = z_5 = 1, z_1 = z_2 = z_3 = z_4 = z_6 = z_7 = 0$. Note that $\psi_1 = \beta_0 = \gamma_1, \psi_2 = \gamma_0 = \beta_1$, and hence $\psi_1 + \psi_2 = \beta_0 + \beta_1 \leq 1$, thus contradicting $\psi_1 + \psi_2 > 1$. ■

Appendix I

Proof of lemma 5.10.2

Since $A - B - Y$ and $AB - X - Y$ are Markov chains, to prove $A - B - XY$ is a Markov chain, it suffices to prove $A - B - X$ is a Markov chain. We therefore need to prove $p_{XA|B}(x_k, a_i|b_j) = p_{X|B}(x_k|b_j)p_{A|B}(a_i|b_j)$ for every $(x_k, a_i, b_j) \in \{0, 1\} \times \mathcal{A} \times \mathcal{B}$ such that $p_B(b_j) > 0$. It suffices to prove $p_{XA|B}(0, a_i|b_j) = p_{X|B}(0|b_j)p_{A|B}(a_i|b_j)$ for every $(a_i, b_j) \in \mathcal{A} \times \mathcal{B}$ such that $p_B(b_j) > 0$.[1]

Fix a b_j for which $p_B(b_j) > 0$. Let $p_{A|B}(a_i|b_j) = \alpha_i$ for each $i \in \mathbb{N}$ and $p_{XA|B}(0, a_i|b_j) = \chi_i$ for each $(i, j) \in \mathbb{N} \times \mathbb{N}$. It can be verified $p_{XYA|B}(\cdot, \cdot, \cdot|b_j)$ is as in table I.1. From table I.1, we infer $p_{AY|B}(a_i0|b_j) = \chi_i(1-\eta) + (\alpha_i - \chi_i)\eta = \alpha_i\eta + \chi_i(1-2\eta)$. From the Markov chain $A - B - Y$, we have $p_{AY|B}(a_i0|b_j) = p_{A|B}(a_i|b_j)p_{Y|B}(0|b_j) = \alpha_i p_{Y|B}(0|b_j)$. Therefore, $\alpha_i p_{Y|B}(0|b_j) = \alpha_i\eta + \chi_i(1-2\eta)$. Since $1 - 2\eta \neq 0$, we substitute for χ_i and α_i in terms of their definitions to conclude

$$p_{XA|B}(0, a_i|b_j) = \chi_i = \alpha_i \cdot \frac{p_{Y|B}(0|b_j) - \eta}{1 - 2\eta} = p_{A|B}(a_i|b_j)\frac{p_{Y|B}(0|b_j) - \eta}{1 - 2\eta}.$$

Since $\frac{p_{Y|B}(0|b_j)-\eta}{1-2\eta}$ is independent of i and b_j was an arbitrary element in $\mathcal{B}$ that satisfies $p_B(b_j) > 0$, we have established Markov chain $A - B - X$.

[1] Indeed, $p_{XA|B}(1, a_i|b_j) = p_{A|B}(a_i|b_j) - p_{XA|B}(0, a_i|b_j) = p_{A|B}(a_i|b_j)(1 - p_{X|B}(0|b_j)) = p_{A|B}(a_i|b_j)p_{X|B}(1|b_j)$.

AXY	$p_{AXY\|B}(\cdot, \cdot, \cdot\|b_j)$	AXY	$p_{AXY\|B}(\cdot, \cdot, \cdot\|b_j)$	AXY	$p_{AXY\|B}(\cdot, \cdot, \cdot\|b_j)$	AXY	$p_{AXY\|B}(\cdot, \cdot, \cdot\|b_j)$
a_i00	$\chi_i(1-\eta)$	a_i01	$\chi_i\eta$	a_i10	$(\alpha_i - \chi_i)\eta$	a_i11	$(\alpha_i - \chi_i)(1-\eta) =$

Table I.1: $p_{AXY|B}(\cdot, \cdot, \cdot|b_j)$

Appendix J

Upper bound on $P(\epsilon_l)$

From (5.25), it suffices to derive upper and lower bounds on $\mathrm{Var}\left\{\phi(M_1, M_2^{t_2}, M_3^{t_3})\right\}$ and $\mathbb{E}\left\{\phi(M_1, M_2^{t_2}, M_3^{t_3})\right\}$ respectively. Note that $\mathbb{E}\left\{\phi^2(m_1, m_2^{t_2}, m_3^{t_3})\right\} = \sum_{l=0}^{7} \mathscr{T}_l$, where

$$\mathscr{T}_0 = \mathbb{E}\left\{\phi(M_1, M_2^{t_2}, M_3^{t_3})\right\} = \sum_{\substack{(b_1, a_2^{s_2}, a_3^{s_3}) \in \\ \mathcal{B}_1 \times \mathcal{F}_\pi^{s_2} \times \mathcal{F}_\pi^{s_3}}} \sum_{\substack{(v_1^n, u_2^n, u_3^n) \in \\ T_{2\eta_2}(V_1, U_2, U_3 | q^n)}} P\left(V_1^n(m_1, b_1) = v_1^n, U_j(a_j^{s_j}) = u_j^n, I(a_j^{s_j}) = m_j^{t_j} : j = 2,3\right), \tag{J.1}$$

$$\mathscr{T}_1 = \sum_{\substack{(b_1, a_2^{s_2}, a_3^{s_3}) \in \\ \mathcal{B}_1 \times \mathcal{F}_\pi^{s_2} \times \mathcal{F}_\pi^{s_3}}} \sum_{\substack{\tilde{a}_3^{s_3} \in \mathcal{F}_\pi^{s_3} \\ \tilde{a}_3^{s_3} \neq a_3^{s_3}}} \sum_{\substack{(v_1^n, u_2^n, u_3^n) \in \\ T_{2\eta_2}(V_1, \underline{U} | q^n)}} \sum_{\substack{\tilde{u}_3^n \in \\ T_{2\eta_2}(U_3 | q^n, v_1^n, u_2^n)}} P\left(\begin{matrix} V_1^n(m_1, b_1) = v_1^n, U_j(a_j^{s_j}) = u_j^n, I(a_j^{s_j}) = m_j^{t_j} : j = 2,3, \\ U_3^n(\tilde{a}_3^{s_3}) = \tilde{u}_3^n, I(\tilde{a}_3^{s_3}) = m_3^{t_3} \end{matrix}\right)$$

$$\mathscr{T}_2 = \sum_{\substack{(b_1, a_2^{s_2}, a_3^{s_3}) \in \\ \mathcal{B}_1 \times \mathcal{F}_\pi^{s_2} \times \mathcal{F}_\pi^{s_3}}} \sum_{\substack{\tilde{a}_2^{s_2} \in \mathcal{F}_\pi^{s_2} \\ \tilde{a}_2^{s_2} \neq a_2^{s_2}}} \sum_{\substack{(v_1^n, u_2^n, u_3^n) \in \\ T_{2\eta_2}(V_1, \underline{U} | q^n)}} \sum_{\substack{\tilde{u}_2^n \in \\ T_{2\eta_2}(U_2 | q^n, v_1^n, u_3^n)}} P\left(\begin{matrix} V_1^n(m_1, b_1) = v_1^n, U_j(a_j^{s_j}) = u_j^n, I(a_j^{s_j}) = m_j^{t_j} : j = 2,3, \\ U_2^n(\tilde{a}_2^{s_2}) = \tilde{u}_2^n, I(\tilde{a}_2^{s_2}) = m_2^{t_2} \end{matrix}\right)$$

$$\mathscr{T}_3 = \sum_{\substack{(b_1, a_2^{s_2}, a_3^{s_3}) \in \\ \mathcal{B}_1 \times \mathcal{F}_\pi^{s_2} \times \mathcal{F}_\pi^{s_3}}} \sum_{\substack{(\tilde{a}_2^{s_2}, \tilde{a}_3^{s_3}) \in \mathcal{F}_\pi^{s_2} \times \mathcal{F}_\pi^{s_3} \\ \tilde{a}_2^{s_2} \neq a_2^{s_2}, \tilde{a}_3^{s_3} \neq a_3^{s_3}}} \sum_{\substack{(v_1^n, u_2^n, u_3^n) \in \\ T_{2\eta_2}(V_1, \underline{U} | q^n)}} \sum_{\substack{(\tilde{u}_2^n, \tilde{u}_3^n) \in \\ T_{2\eta_2}(\underline{U} | q^n, v_1^n)}} P\left(\begin{matrix} V_1^n(m_1, b_1) = v_1^n, U_j(a_j^{s_j}) = u_j^n, I(a_j^{s_j}) = m_j^{t_j} : j = 2,3, \\ U_j(\tilde{a}_j^{s_j}) = \tilde{u}_j^n, I(\tilde{a}_j^{s_j}) = m_j^{t_j} : j = 2,3 \end{matrix}\right)$$

$$\mathscr{T}_4 = \sum_{\substack{(b_1, a_2^{s_2}, a_3^{s_3}) \in \\ \mathcal{B}_1 \times \mathcal{F}_\pi^{s_2} \times \mathcal{F}_\pi^{s_3}}} \sum_{\substack{\tilde{b}_1 \in \mathcal{B}_1 \\ \tilde{b}_1 \neq b_1}} \sum_{\substack{(v_1^n, u_2^n, u_3^n) \in \\ T_{2\eta_2}(V_1, \underline{U} | q^n)}} \sum_{\substack{\tilde{v}_1^n \in \\ T_{2\eta_2}(V_1 | q^n, \underline{u}^n)}} P\left(V_1^n(m_1, b_1) = v_1^n, U_j(a_j^{s_j}) = u_j^n, I(a_j^{s_j}) = m_j^{t_j} : j = 2,3, V_1^n(m_1, \tilde{b}_1) = \tilde{v}_1^n\right)$$

$$\mathscr{T}_5 = \sum_{\substack{(b_1, a_2^{s_2}, a_3^{s_3}) \in \\ \mathcal{B}_1 \times \mathcal{F}_\pi^{s_2} \times \mathcal{F}_\pi^{s_3}}} \sum_{\substack{(\tilde{b}_1, \tilde{a}_3^{s_3}) \in \mathcal{B}_1 \times \mathcal{F}_\pi^{s_3} \\ \tilde{b}_1 \neq b_1 \tilde{a}_3^{s_3} \neq a_3^{s_3}}} \sum_{\substack{(v_1^n, u_2^n, u_3^n) \in \\ T_{2\eta_2}(V_1, \underline{U} | q^n)}} \sum_{\substack{\tilde{v}_1^n, \tilde{u}_3^n \in \\ T_{2\eta_2}(V_1, U_3 | q^n, u_2^n)}} P\left(\begin{matrix} V_1^n(m_1, b_1) = v_1^n, U_j(a_j^{s_j}) = u_j^n, I(a_j^{s_j}) = m_j^{t_j} : j = 2,3, \\ V_1(m_1, \tilde{b}_1) = \tilde{v}_1^n, U_3^n(\tilde{a}_3^{s_3}) = \tilde{u}_3^n, I(\tilde{a}_3^{s_3}) = m_3^{t_3} \end{matrix}\right)$$

$$\mathscr{T}_6 = \sum_{\substack{(b_1, a_2^{s_2}, a_3^{s_3}) \in \\ \mathcal{B}_1 \times \mathcal{F}_\pi^{s_2} \times \mathcal{F}_\pi^{s_3}}} \sum_{\substack{(\tilde{b}_1, \tilde{a}_2^{s_2}) \in \mathcal{B}_1 \times \mathcal{F}_\pi^{s_2} \\ \tilde{b}_1 \neq b_1 \tilde{a}_2^{s_2} \neq a_2^{s_2}}} \sum_{\substack{(v_1^n, u_2^n, u_3^n) \in \\ T_{2\eta_2}(V_1, \underline{U} | q^n)}} \sum_{\substack{\tilde{v}_1^n, \tilde{u}_2^n \in \\ T_{2\eta_2}(V_1, U_2 | q^n, u_3^n)}} P\left(\begin{matrix} V_1^n(m_1, b_1) = v_1^n, U_j(a_j^{s_j}) = u_j^n, I(a_j^{s_j}) = m_j^{t_j} : j = 2,3, \\ V_1(m_1, \tilde{b}_1) = \tilde{v}_1^n, U_2^n(\tilde{a}_2^{s_2}) = \tilde{u}_2^n, I(\tilde{a}_2^{s_2}) = m_2^{t_2} \end{matrix}\right)$$

$$\mathscr{T}_7 = \sum_{\substack{(b_1, a_2^{s_2}, a_3^{s_3}) \in \\ \mathcal{B}_1 \times \mathcal{F}_\pi^{s_2} \times \mathcal{F}_\pi^{s_3}}} \sum_{\substack{(\tilde{b}_1, \tilde{a}_2^{s_2}, \tilde{a}_3^{s_3}) \in \mathcal{B}_1 \times \mathcal{F}_\pi^{s_2} \times \mathcal{F}_\pi^{s_3} \\ \tilde{b}_1 \neq b_1, \tilde{a}_2^{s_2} \neq a_2^{s_2}, \tilde{a}_3^{s_3} \neq a_3^{s_3}}} \sum_{\substack{(v_1^n, u_2^n, u_3^n) \in \\ T_{2\eta_2}(V_1, \underline{U} | q^n)}} \sum_{\substack{(\tilde{v}_1^n, \tilde{u}_2^n, \tilde{u}_3^n) \in \\ T_{2\eta_2}(V_1, \underline{U} | q^n)}} P\left(\begin{matrix} V_1^n(m_1, b_1) = v_1^n, U_j(a_j^{s_j}) = u_j^n, I(a_j^{s_j}) = m_j^{t_j} : j = 2,3, \\ V_1(m_1, \tilde{b}_1) = \tilde{v}_1^n, U_j^n(\tilde{a}_j^{s_j}) = \tilde{u}_j^n, I(\tilde{a}_j^{s_j}) = m_j^{t_j} : j = 2,3 \end{matrix}\right).$$

We have

$$\frac{4\text{Var}\left\{\phi(M_1, M_2^{t_2}, M_3^{t_3})\right\}}{\left(\mathbb{E}\left\{\phi(M_1, M_2^{t_2}, M_3^{t_3})\right\}\right)^2} = 4\frac{\left(\sum_{l=0}^{7}\mathscr{T}_l\right) - \mathscr{T}_0^2}{\mathscr{T}_0^2}.$$

We take a closer look at $\mathscr{T}_7$. For $\theta \in \mathcal{F}_\pi$, let

$$\mathscr{D}_\theta(a_2^{s_2}, a_3^{s_3}) := \left\{(\tilde{a}_2^{s_2}, \tilde{a}_3^{s_3}) : \tilde{a}_{3l}^{s_3} - a_{3l}^{s_3} = \theta(\tilde{a}_{2l}^{s_2} - a_{2l}^{s_2}) \text{ for } 1 \leq l \leq s_2 \text{ and } \tilde{a}_{3l}^{s_3} - a_{3l}^{s_3} = 0 \text{ for } s_2 + 1 \leq l \leq s_3\right\},$$

$\mathscr{D}(a_2^{s_2}, a_3^{s_3}) := \underset{\theta\in\mathcal{F}_\pi}{\cup}\mathscr{D}_\theta(a_2^{s_2}, a_3^{s_3})$ and $\mathscr{I}(a_2^{s_2}, a_3^{s_3}) = \mathcal{F}_\pi^{s_2} \times \mathcal{F}_\pi^{s_3} \setminus \mathscr{D}(a_2^{s_2}, a_3^{s_3})$. The reader may verify that for $(\tilde{a}_2^{s_2}, \tilde{a}_3^{s_3}) \in \mathscr{D}_\theta(a_2^{s_2}, a_3^{s_3})$

$$P\begin{pmatrix} V_1^n(m_1,b_1)=v_1^n, U_j(a_j^{s_j})=u_j^n, I(a_j^{s_j})=m_j^{t_j}:j=2,3, \\ V_1(m_1,\tilde{b}_1)=\tilde{v}_1^n, U_j^n(\tilde{a}_j^{s_j})=\tilde{u}_j^n, I(\tilde{a}_j^{s_j})=m_j^{t_j}:j=2,3 \end{pmatrix} = \begin{cases} \frac{P(V_1^n(m_1,b_1)=v_1^n, V_1(m_1,\tilde{b}_1)=\tilde{v}_1^n)}{\pi^{3n+2t_2+2t_3}} & \text{if } \tilde{u}_3^n - \theta\tilde{u}_2^n = u_3^n - \theta u_2^n \\ 0 & \text{otherwise} \end{cases}$$

For $(\tilde{a}_2^{s_2}, \tilde{a}_3^{s_3}) \in \mathscr{I}(a_2^{s_2}, a_3^{s_3})$, we claim

$$P\begin{pmatrix} V_1^n(m_1,b_1)=v_1^n, U_j(a_j^{s_j})=u_j^n, I(a_j^{s_j})=m_j^{t_j}:j=2,3, \\ V_1(m_1,\tilde{b}_1)=\tilde{v}_1^n, U_j^n(\tilde{a}_j^{s_j})=\tilde{u}_j^n, I(\tilde{a}_j^{s_j})=m_j^{t_j}:j=2,3 \end{pmatrix} = P\begin{pmatrix} V_1^n(m_1,b_1)=v_1^n, U_j(a_j^{s_j})=u_j^n \\ I(a_j^{s_j})=m_j^{t_j}:j=2,3 \end{pmatrix} P\begin{pmatrix} V_1(m_1,\tilde{b}_1)=\tilde{v}_1^n, U_j^n(\tilde{a}_j^{s_j})=\tilde{u}_j^n \\ I(\tilde{a}_j^{s_j})=m_j^{t_j}:j=2,3 \end{pmatrix}.$$

In order to prove this claim, it suffices to prove

$$P\begin{pmatrix} V_1^n(m_1,b_1)=v_1^n, U_j(a_j^{s_j})=u_j^n, I(a_j^{s_j})=m_j^{t_j}:j=2,3, \\ V_1(m_1,\tilde{b}_1)=\tilde{v}_1^n, U_j^n(\tilde{a}_j^{s_j})=\tilde{u}_j^n, I(\tilde{a}_j^{s_j})=m_j^{t_j}:j=2,3 \end{pmatrix} = \frac{P(V_1^n(m_1,b_1)=v_1^n, V_1(m_1,\tilde{b}_1)=\tilde{v}_1^n)}{\pi^{4n+2t_2+2t_3}}.$$

which can be verified through a counting process similar to that employed in lemma M.0.18. We therefore have $\mathscr{T}_7 = \mathscr{T}_{7I} + \mathscr{T}_{7D}$, where

$$\mathscr{T}_{7I} = \sum_{\substack{(b_1,a_2^{s_2},a_3^{s_3})\in \\ \mathcal{B}_1\times\mathcal{F}_\pi^{s_2}\times\mathcal{F}_\pi^{s_3}}} \sum_{\substack{(\tilde{b}_1,\tilde{a}_2^{s_2},\tilde{a}_3^{s_3})\in \\ \mathcal{B}_1\times\mathscr{I}(a_2^{s_2},a_3^{s_3})}} \sum_{\substack{(v_1^n,u_2^n,u_3^n)\in \\ T_{2\eta_2}(V_1,\underline{U}|q^n)}} \sum_{\substack{(\tilde{v}_1^n,\tilde{u}_2^n,\tilde{u}_3^n)\in \\ T_{2\eta_2}(V_1,\underline{U}|q^n)}} P\begin{pmatrix} V_1^n(m_1,b_1)=v_1^n, U_j(a_j^{s_j})=u_j^n \\ I(a_j^{s_j})=m_j^{t_j}:j=2,3 \end{pmatrix} P\begin{pmatrix} V_1(m_1,\tilde{b}_1)=\tilde{v}_1^n, U_j^n(\tilde{a}_j^{s_j})=\tilde{u}_j^n \\ I(\tilde{a}_j^{s_j})=m_j^{t_j}:j=2,3 \end{pmatrix} \quad \text{(J.2)}$$

$$\mathscr{T}_{7D} = \sum_{\substack{(b_1,a_2^{s_2},a_3^{s_3})\in \\ \mathcal{B}_1\times\mathcal{F}_\pi^{s_2}\times\mathcal{F}_\pi^{s_3}}} \sum_{\substack{(\tilde{a}_2^{s_2},\tilde{a}_3^{s_3})\in \\ \mathscr{D}(a_2^{s_2},a_3^{s_3})}} \sum_{\substack{u^n\in \\ T_{2\eta_2}(U_3\ominus\theta U_2|q^n)}} \sum_{\substack{(v_1^n,u_2^n,u^n\oplus\theta u_2^n)\in \\ T_{2\eta_2}(V_1,\underline{U}|q^n)}} \sum_{\substack{(\tilde{v}_1^n,\tilde{u}_2^n,u^n\oplus\theta\tilde{u}_2^n)\in \\ T_{2\eta_2}(V_1,\underline{U}|q^n)}} \frac{P(V_1^n(m_1,b_1)=v_1^n, V_1(m_1,\tilde{b}_1)=\tilde{v}_1^n)}{\pi^{3n+2t_2+2t_3}}.$$

Verify that $\mathscr{T}_{7I} \leq \mathscr{T}_0^2$. We therefore have

$$\frac{4\text{Var}\left\{\phi(M_1, M_2^{t_2}, M_3^{t_3})\right\}}{\mathbb{E}\left\{\phi(M_1, M_2^{t_2}, M_3^{t_3})\right\}} \leq 4\frac{\left(\sum_{l=0}^{6}\mathscr{T}_l\right) + \mathscr{T}_{7D}}{\mathscr{T}_0^2}. \quad \text{(J.3)}$$

and it suffices to derive lower bound on $\mathscr{T}_0$ and upper bounds on $\mathscr{T}_l : l \in [6]$ and $\mathscr{T}_{7D}$.

Just as we split $\mathscr{T}_7$, we split $\mathscr{T}_3$ as $\mathscr{T}_3 = \mathscr{T}_{3I} + \mathscr{T}_{3D}$. We let the reader fill in the details and confirm the following

bounds. From lemmas 2.2.3, 2.4.2, there exists $N_2(\eta_2) \in \mathbb{N}$, such that for all $n \geq N_2(\eta_2)$,

$$\begin{aligned}
\mathscr{T}_0 &\geq \frac{|\mathcal{B}_1|\pi^{s_2+s_3}\exp\{nH(V_1,\underline{U}|Q)-4n\eta_2\}}{\pi^{2n+t_2+t_3}\exp\{nH(V_1|Q)+4n\eta_2\}} \\
\mathscr{T}_1 &\leq \frac{|\mathcal{B}_1|\pi^{s_2+2s_3}\exp\{nH(V_1,\underline{U}|Q)+4n\eta_2+nH(U_3|Q,V_1,U_2)+8n\eta_2\}}{\pi^{3n+t_2+2t_3}\exp\{nH(V_1|Q)-4n\eta_2\}} \\
\mathscr{T}_2 &\leq \frac{|\mathcal{B}_1|\pi^{2s_2+s_3}\exp\{nH(V_1,\underline{U}|Q)+4n\eta_2+nH(U_2|Q,V_1,U_3)+8n\eta_2\}}{\pi^{3n+2t_2+t_3}\exp\{nH(V_1|Q)-4n\eta_2\}} \\
\mathscr{T}_{3I} &\leq \frac{|\mathcal{B}_1|\pi^{2s_2+2s_3}\exp\{nH(V_1,\underline{U}|Q)+4n\eta_2+nH(U_2,U_3|Q,V_1)+8n\eta_2\}}{\pi^{4n+2t_2+2t_3}\exp\{nH(V_1|Q)-4n\eta_2\}} \\
\mathscr{T}_{3D} &\leq \pi\frac{|\mathcal{B}_1|\pi^{2s_2+s_3}\exp\{nH(V_1,\underline{U}|Q,U_3\ominus\theta U_2)+8n\eta_2+nH(U_3\ominus\theta U_2|Q)+4n\eta_2\}}{\pi^{3n+2t_2+2t_3}\exp\{nH(V_1|Q)-4n\eta_2-nH(\underline{U}|Q,V_1,U_3\ominus\theta U_2)-16n\eta_2\}} \\
\mathscr{T}_4 &\leq \frac{|\mathcal{B}_1|^2\pi^{s_2+s_3}\exp\{nH(V_1,\underline{U}|Q)+4n\eta_2+nH(V_1|Q,U_2,U_3)+8n\eta_2\}}{\pi^{2n+t_2+t_3}\exp\{2nH(V_1|Q)-8n\eta_2\}} \\
\mathscr{T}_5 &\leq \frac{|\mathcal{B}_1|^2\pi^{s_2+2s_3}\exp\{nH(V_1,\underline{U}|Q)+4n\eta_2+nH(V_1,U_3|Q,U_2)+8n\eta_2\}}{\pi^{3n+t_2+2t_3}\exp\{2nH(V_1|Q)-8n\eta_2\}} \\
\mathscr{T}_6 &\leq \frac{|\mathcal{B}_1|^2\pi^{2s_2+s_3}\exp\{nH(V_1,\underline{U}|Q)+4n\eta_2+nH(V_1,U_2|Q,U_3)+8n\eta_2\}}{\pi^{3n+2t_2+t_3}\exp\{2nH(V_1|Q)-8n\eta_2\}} \\
\mathscr{T}_{7D} &\leq \frac{|\mathcal{B}_1|^2\pi^{2s_2+s_3}\exp\{2nH(V_1,\underline{U}|Q,U_3\ominus\theta U_2)+16n\eta_2+nH(U_3\ominus\theta U_2|Q)+4n\eta_2\}}{\pi^{3n+2t_2+2t_3}\exp\{2nH(V_1|Q)-8n\eta_2\}}
\end{aligned}$$

We now employ the bounds on the parameters of the code ((5.22) - (5.24)). It maybe verified that, for $n \geq \max\{N_1(\eta), N_2(\eta_2)\}$,

$$\begin{aligned}
\frac{\mathscr{T}_0}{\mathscr{T}_0^2} &\leq \exp\left\{-n\left(\frac{\log|\mathcal{B}_1|}{n}+\left(\sum_{l=2}^{3}\frac{s_l-t_l}{n}\right)\log\pi-[2\log\pi-H(\underline{U}|Q,V_1)+16\eta_2]\right)\right\} \leq \exp\left\{-n\left(\begin{smallmatrix}\delta_1+\frac{\eta}{8}\\-16\eta_2\end{smallmatrix}\right)\right\} \quad \text{(J.4)} \\
\frac{\mathscr{T}_1}{\mathscr{T}_0^2} &\leq \exp\left\{-n\left(\frac{\log|\mathcal{B}_1|}{n}+\frac{s_2-t_2}{n}\log\pi-[1-H(U_2|Q,V_1)+32\eta_2]\right)\right\} \leq \exp\left\{-n\left(\delta_1+\frac{\eta}{8}-32\eta_2\right)\right\} \\
\frac{\mathscr{T}_2}{\mathscr{T}_0^2} &\leq \exp\left\{-n\left(\frac{\log|\mathcal{B}_1|}{n}+\frac{s_3-t_3}{n}\log\pi-[1-H(U_3|Q,V_1)+32\eta_2]\right)\right\} \leq \exp\left\{-n\left(\delta_1+\frac{\eta}{8}-32\eta_2\right)\right\} \\
\frac{\mathscr{T}_{3I}}{\mathscr{T}_0^2} &\leq \exp\left\{-n\left(\frac{\log|\mathcal{B}_1|}{n}-32\eta_2\right)\right\} \leq \exp\left\{-n\left(\delta_1+\frac{\eta}{8}-32\eta_2\right)\right\} \\
\frac{\mathscr{T}_{3D}}{\mathscr{T}_0^2} &\leq \max_{\theta\neq 0}\pi\exp\left\{-n\left(\frac{\log|\mathcal{B}_1|}{n}+\frac{s_3}{n}\log\pi-[1-H(U_3\ominus\theta U_2|Q,V_1)+48\eta_2]\right)\right\} \leq \pi\exp\{-n(\delta_1-48\eta_2)\} \\
\frac{\mathscr{T}_4}{\mathscr{T}_0^2} &\leq \exp\left\{-n\left(\left(\sum_{l=2}^{3}\frac{s_l-t_l}{n}\right)\log\pi-[2-H(\underline{U}|Q)+36\eta_2]\right)\right\} \leq \exp\{-n(\delta_1-36\eta_2)\} \\
\frac{\mathscr{T}_5}{\mathscr{T}_0^2} &\leq \exp\left\{-n\left(\frac{s_2-t_2}{n}\log\pi-[1-H(U_2|Q)+36\eta_2]\right)\right\} \leq \exp\{-n(\delta_1-36\eta_2)\} \\
\frac{\mathscr{T}_6}{\mathscr{T}_0^2} &\leq \exp\left\{-n\left(\frac{s_3-t_3}{n}\log\pi-[1-H(U_3|Q)+36\eta_2]\right)\right\} \leq \exp\{-n(\delta_1-36\eta_2)\} \\
\frac{\mathscr{T}_{7D}}{\mathscr{T}_0^2} &\leq \max_{\theta\neq 0}\pi\exp\left\{-n\left(\frac{s_3}{n}\log\pi-[1-H(U_3\ominus\theta U_2|Q)+48\eta_2]\right)\right\} \leq \pi\exp\left\{-n\left(\delta_1-\frac{\eta}{8}-48\eta_2\right)\right\}.
\end{aligned}$$

Substituting, the above bounds in (J.3), we conclude $P(\epsilon_l) \leq (28 + 8\log\pi)\exp\left\{-n\left(\delta_1 - \frac{\eta}{8} - 48\eta_2\right)\right\}$ for $n \geq \max\{N_1(\eta), N_2(\eta_2)\}$. In the sequel, we derive a lower bound on $\mathcal{L}(n)$ and prove that for large n, $\mathcal{L}(n) > 1$, thereby establishing $\epsilon_1 \subseteq \epsilon_l$. From the definition of $\mathcal{L}(n)$, (J.1), we have

$$\mathcal{L}(n) = \frac{\mathscr{T}_0}{2} \geq \frac{|\mathcal{B}_1|\pi^{s_2+s_3}|T_{2\eta_2}(V_1, \underline{U}|q^n)|}{2\pi^{2n+t_2+t_3}\exp\{nH(V_1|Q) + 4n\eta_2\}}, \tag{J.5}$$

for sufficiently large n. Moreover, from (J.4), we note that $\mathcal{L}(n) \geq \frac{1}{2}\exp\left\{n\left(\delta_1 + \frac{\eta}{8} - 16\eta_2\right)\right\}$ for $n \geq \max\{N_1(\eta), N_2(\eta_2)\}$. By our choice of η, η_2, for sufficiently large n, we have $\mathcal{L}(n) > 1$.

Appendix K

Upper bound on $P\big((\epsilon_1 \cup \epsilon_2 \cup \epsilon_3)^c \cap \epsilon_{41}\big)$

We begin by introducing some compact notation. We let $\underline{M}^{\underline{t}}$ denote the pair $(M_2^{t_2}, M_3^{t_3})$ of message random variables. We let $\underline{m}^{\underline{t}}$ denote a generic element $(m_2^{t_2}, m_3^{t_3}) \in \mathcal{F}_\pi^{\underline{t}} := \mathcal{F}_\pi^{t_2} \times \mathcal{F}_\pi^{t_3}$, and similarly $\underline{a}^{\underline{s}}$ denote $(a_2^{s_2}, a_3^{s_3}) \in \mathcal{F}_\pi^{\underline{s}} := \mathcal{F}_\pi^{s_2} \times \mathcal{F}_\pi^{s_3}$. We abbreviate $T_{8\eta_2}(V_1, U_2 \oplus U_3 | q^n, y_1^n)$ as $T_{8\eta_2}(V_1, \oplus | q^n, y_1^n)$ and the vector $X^n(M_1, M_2^{t_2}, M_3^{t_3})$ input on the channel as X^n. Let

$$\tilde{T}_{\eta_2}(q^n) := \{(v_1^n, \underline{u}^n, x^n, y_1^n) \in T_{8\eta_2}(V_1, \underline{U}, X, Y_1 | q^n) : (v_1^n, \underline{u}^n) \in T_{2\eta_2}(V_1, \underline{U} | q^n), (v_1^n, \underline{u}^n, x^n) \in T_{4\eta_2}(V_1, \underline{U}, X | q^n)\},$$
$$\tilde{T}_{\eta_2}(q^n | v_1^n, \underline{u}^n) = \Big\{(x^n, y_1^n) : (v_1^n, \underline{u}^n, x^n, y_1^n) \in \tilde{T}_{\eta_2}(q^n)\Big\}$$

We begin by characterizing the event under question. Denoting $\tilde{\epsilon}_{41} = (\epsilon_l \cup \epsilon_2 \cup \epsilon_3)^c \cap \epsilon_{41}$, we have

$$P(\tilde{\epsilon}_{41}) \leq \sum_{m_1} \sum_{\hat{m}_1 \neq m_1} \sum_{\hat{b}_1 \in \mathcal{B}_1} \sum_{\hat{a}_3^{s_3}} \sum_{\substack{(v_1^n, \underline{u}^n, x^n, y_1^n) \\ \in \tilde{T}_{\eta_2}(q^n)}} \sum_{\substack{(\hat{v}_1^n, \hat{u}^n) \in \\ T_{8\eta_2}(V_1, \oplus | q^n, y_1^n)}} P\left(\left\{\begin{smallmatrix} M_1 = m_1, V_1^n(m_1, B_1) = v_1^n, U_l^n(A_l^{s_l}) = u_l^n \\ I_l(A_l^{s_l}) = M_l^{t_l} : l = 2,3, Y_1^n = y_1^n, X^n = x^n \\ U_\oplus^n(\hat{a}_3^{s_3}) = \hat{u}^n, V_1^n(\hat{m}_1, \hat{b}_1) = \hat{v}_1^n \end{smallmatrix}\right\} \cap \epsilon_l^c\right) \tag{K.1}$$

We consider a generic term in the above sum. Observe that

$$P\left(\begin{smallmatrix} Y_1^n = y_1^n \\ X^n = x^n \end{smallmatrix} \middle| \left\{\begin{smallmatrix} M_1 = m_1, V_1^n(m_1, B_1) = v_1^n, U_l^n(A_l^{s_l}) = u_l^n \\ I_l(A_l^{s_l}) = M_l^{t_l} : l = 2,3, U_\oplus^n(\hat{a}_3^{s_3}) = \hat{u}^n, V_1^n(\hat{m}_1, \hat{b}_1) = \hat{v}_1^n \end{smallmatrix}\right\} \cap \epsilon_l^c\right) = P\left(\begin{smallmatrix} Y_1^n = y_1^n \\ X^n = x^n \end{smallmatrix} \middle| \begin{smallmatrix} V_1^n(M_1, B_1) = v_1^n \\ U_l^n(A_l^{s_l}) = u_l^n : l = 2,3 \end{smallmatrix}\right) =: \theta(y_1^n, x^n | v_1^n, \underline{u}^n), \tag{K.2}$$

$$P\left(\left\{\begin{smallmatrix} M_1 = m_1, V_1^n(m_1, B_1) = v_1^n \\ U_l^n(A_l^{s_l}) = u_l^n, I_l(A_l^{s_l}) = M_l^{t_l} : l = 2,3 \\ U_\oplus^n(\hat{a}_3^{s_3}) = \hat{u}^n, V_1^n(\hat{m}_1, \hat{b}_1) = \hat{v}_1^n \end{smallmatrix}\right\} \cap \epsilon_l^c\right) = \sum_{\underline{m}^{\underline{t}} \in \mathcal{F}_\pi^{\underline{t}}} \sum_{\substack{(b_1, \underline{a}^{\underline{s}}) \in \\ \mathcal{B}_1 \times \mathcal{F}_\pi^{\underline{s}}}} P\left(\left\{\begin{smallmatrix} M_1 = m_1, V_1^n(m_1, b_1) = v_1^n, U_l^n(a_l^{s_l}) = u_l^n \\ M_l^{t_l} = m_l^{t_l}, A_l^{s_l} = a_l^{s_l}, I_l(a_l^{s_l}) = m_l^{t_l} : l = 2,3 \\ B_1 = b_1, U_\oplus^n(\hat{a}_3^{s_3}) = \hat{u}^n, V_1^n(\hat{m}_1, \hat{b}_1) = \hat{v}_1^n \end{smallmatrix}\right\} \cap \epsilon_l^c\right), \tag{K.3}$$

and the product of left hand sides of (K.2) and (K.3) is a generic term in (K.1). We now consider a generic term on the right hand side of (K.3). Note that

$$P\left(E \cap \left\{B_1 = b_1, A_l^{s_l} = a_l^{s_l}\right\} \cap \epsilon_l^c\right) \leq P(E) P\left(\left\{B_1 = b_1, A_l^{s_l} = a_l^{s_l}\right\} | E \cap \epsilon_l^c\right) = \frac{P(E)}{\mathcal{L}(n)},$$

where E abbreviates the event $\{M_1{=}m_1,V_1^n(m_1,b_1){=}v_1^n,U_l^n(a_l^{s_l}){=}u_l^n,M_l^{t_l}{=}m_l^{t_l},I_l(a_l^{s_l}){=}m_l^{t_l}:l{=}2,3,U_\oplus^n(\hat{a}_3^{s_3}){=}\hat{u}^n,V_1^n(\hat{m}_1,\hat{b}_1){=}\hat{v}_1^n\}$. Substituting the above in (K.3), we have

$$P\left(\left\{\begin{smallmatrix}M_1=m_1,V_1^n(m_1,B_1)=v_1^n\\ U_l^n(A_l^{s_l})=u_l^n,I_l(A_l^{s_l})=M_l^{t_l}:l=2,3\\ U_\oplus^n(\hat{a}_3^{s_3})=\hat{u}^n,V_1^n(\hat{m}_1,\hat{b}_1)=\hat{v}_1^n\end{smallmatrix}\right\}\cap\epsilon_l^c\right)\le\frac{1}{\mathcal{L}(n)}\sum_{\underline{m}^t\in\mathcal{F}_\pi^t}\sum_{\substack{(b_1,\underline{a}^s)\\ \in\mathcal{B}_1\times\mathscr{D}(\hat{a}^{s_3})}}P\left(\begin{smallmatrix}M_1=m_1,V_1^n(m_1,b_1)=v_1^n,U_l^n(a_l^{s_l})=u_l^n,M_l^{t_l}=m_l^{t_l}\\ I_l(a_l^{s_l})=m_l^{t_l}:l=2,3,U_\oplus^n(\hat{a}_3^{s_3})=\hat{u}^n,V_1^n(\hat{m}_1,\hat{b}_1)=\hat{v}_1^n\end{smallmatrix}\right)$$
$$+\frac{1}{\mathcal{L}(n)}\sum_{\underline{m}^t\in\mathcal{F}_\pi^t}\sum_{\substack{(b_1,\underline{a}^s)\\ \in\mathcal{B}_1\times\mathscr{I}(\hat{a}^{s_3})}}P\left(\begin{smallmatrix}M_1=m_1,V_1^n(m_1,b_1)=v_1^n,U_l^n(a_l^{s_l})=u_l^n,M_l^{t_l}=m_l^{t_l}\\ I_l(a_l^{s_l})=m_l^{t_l}:l=2,3,U_\oplus^n(\hat{a}_3^{s_3})=\hat{u}^n,V_1^n(\hat{m}_1,\hat{b}_1)=\hat{v}_1^n\end{smallmatrix}\right)\quad\text{(K.4)}$$

where $\mathscr{D}(\hat{a}^{s_3}):=\{\underline{a}^s:(a_2^{s_2}0^{s_+})\oplus a_3^{s_3}=\hat{a}^{s_3}\}$, $s_+=s_3-s_2$ and $\mathscr{I}(\hat{a}^{s_3}):=\mathcal{F}_\pi^{s_2}\times\mathcal{F}_\pi^{s_3}\setminus\mathscr{D}(\hat{a}^{s_3})$. Let us evaluate a generic term in the right hand side of (K.4). The collection $M_1,M_2^{t_2},M_3^{t_3},V_1^n(m_1,b_1),I_2(a^{s_2}),I_3(a^{s_3}),(U_l(a_l^{s_l}):l=2,3,U_\oplus(\hat{a}_3^{s_3})),V_1^n(\hat{m}_1,\hat{b}_1)$ are mutually independent, where $(U_l(a_l^{s_l}):l=2,3,U_\oplus(\hat{a}_3^{s_3}))$ is treated as a single random object. If $(a_2^{s_2},a_3^{s_3})\in\mathscr{D}(\hat{a}^{s_3})$, then

$$P(U_l(a_l^{s_l})=u_l^n:l=2,3,U_\oplus(\hat{a}_3^{s_3})=\hat{u}^n)=\begin{cases}\frac{1}{\pi^{2n}} & \text{if } u_2^n\oplus u_3^n=\hat{u}^n\\ 0 & \text{otherwise.}\end{cases}.$$

Otherwise, i.e., $(a_2^{s_2},a_3^{s_3})\in\mathscr{I}(\hat{a}^{s_3})$, a counting argument similar to that employed in appendix L proves $P(U_l(a_l^{s_l})=u_l^n:l=2,3,U_\oplus(\hat{a}_3^{s_3})=\hat{u}^n)=\frac{1}{\pi^{3n}}$. We therefore have

$$P\left(\begin{smallmatrix}M_1=m_1,V_1^n(m_1,b_1)=v_1^n,U_l^n(a_l^{s_l})=u_l^n,M_l^{t_l}=m_l^{t_l}\\ I_l(a_l^{s_l})=m_l^{t_l}:l=2,3,U_\oplus^n(\hat{a}_3^{s_3})=\hat{u}^n,V_1^n(\hat{m}_1,\hat{b}_1)=\hat{v}_1^n\end{smallmatrix}\right)=\begin{cases}\frac{P\left(\begin{smallmatrix}M_1=m_1,V_1^n(m_1,b_1)=v_1^n\\ \underline{M}^t=\underline{m}^t,V_1^n(\hat{m}_1,\hat{b}_1)=\hat{v}_1^n\end{smallmatrix}\right)}{\pi^{2n+t_2+t_3}} & \text{if } (a_2^{s_2},a_3^{s_3})\in\mathscr{D}(\hat{a}^{s_3})\\ & \text{and } u_2^n\oplus u_3^n=\hat{u}^n\\ \frac{P\left(\begin{smallmatrix}M_1=m_1,V_1^n(m_1,b_1)=v_1^n\\ \underline{M}^t=\underline{m}^t,V_1^n(\hat{m}_1,\hat{b}_1)=\hat{v}_1^n\end{smallmatrix}\right)}{\pi^{3n+t_2+t_3}} & \text{if } (a_2^{s_2},a_3^{s_3})\in\mathscr{I}(\hat{a}^{s_3})\end{cases}\quad\text{(K.5)}$$

Substituting (K.5) in (K.4) and recognizing that product of right hand sides of (K.3), (K.2) is a generic term in the sum (K.1), we have

$$P(\tilde{\epsilon}_{41})\le\sum_{(m_1,\underline{m}^t)}\sum_{\hat{m}_1\neq m_1}\sum_{\hat{b}_1\in\mathcal{B}_1}\sum_{\hat{a}_3^{s_3}}\sum_{\substack{(b_1,\underline{a}^s)\\ \in\mathcal{B}_1\times\mathscr{D}(\hat{a}^{s_3})}}\sum_{\substack{(v_1^n,\underline{u}^n,x^n,y_1^n)\\ \in\tilde{T}_{\eta_2}(q^n)}}\theta(y_1^n,x^n|v_1^n,\underline{u}^n)\sum_{\substack{(\hat{v}_1^n,u_2^n\oplus u_3^n)\in\\ T_{8\eta_2}(V_1,\oplus|q^n,y_1^n)}}\frac{P\left(\begin{smallmatrix}M_1=m_1,V_1^n(m_1,b_1)=v_1^n\\ \underline{M}^t=\underline{m}^t,V_1^n(\hat{m}_1,\hat{b}_1)=\hat{v}_1^n\end{smallmatrix}\right)}{\pi^{2n+t_2+t_3}\mathcal{L}(n)}$$
$$+\sum_{(m_1,\underline{m}^t)}\sum_{\hat{m}_1\neq m_1}\sum_{\hat{b}_1\in\mathcal{B}_1}\sum_{\hat{a}_3^{s_3}}\sum_{\substack{(b_1,\underline{a}^s)\\ \in\mathcal{B}_1\times\mathscr{I}(\hat{a}^{s_3})}}\sum_{\substack{(v_1^n,\underline{u}^n,x^n,y_1^n)\\ \in\tilde{T}_{\eta_2}(q^n)}}\theta(y_1^n,x^n|v_1^n,\underline{u}^n)\sum_{\substack{(\hat{v}_1^n,\hat{u}^n)\in\\ T_{8\eta_2}(V_1,\oplus|q^n,y_1^n)}}\frac{P\left(\begin{smallmatrix}M_1=m_1,V_1^n(m_1,b_1)=v_1^n\\ \underline{M}^t=\underline{m}^t,V_1^n(\hat{m}_1,\hat{b}_1)=\hat{v}_1^n\end{smallmatrix}\right)}{\pi^{3n+t_2+t_3}\mathcal{L}(n)}$$

The codewords over $\mathcal{V}^n$ are picked independently and identically with respect to $p^n_{V_1|Q}(\cdot|q^n)$ and hence by conditional frequency typicality (lemma 2.2.3), we have

$$P\left(M_1 = m_1, V_1^n(m_1, b_1) = v_1^n, \underline{M^t} = \underline{m^t}, V_1^n(\hat{m}_1, \hat{b}_1) = \hat{v}_1^n\right) \leq \exp\left\{-n(2H(V_1|Q) - 20\eta_2)\right\} P(M_1 = m_1, \underline{M^t} = \underline{m^t})$$

for the pairs $(v_1^n, \hat{v}_1^n)$ in question. This upper bound being independent of the arguments in the summation, we only need to compute the number of terms in the summations. For a fixed pair (u_2^n, u_3^n), lemma 2.4.2 guarantees existence of $N_4(\eta_2) \in \mathbb{N}$ such that for all $n \geq N_4(\eta_2)$, we have $|\left\{v_1^n : (v_1^n, u_2 \oplus u_3^n) \in T_{8\eta_2}(V_1, U_2 \oplus U_3|q^n, y_1^n)\right\}| \leq \exp\left\{n(H(V_1|Q, U_2 \oplus U_3, Y_1) + 32\eta_2)\right\}$ and $|T_{8\eta_2}(V_1, U_2 \oplus U_3|q^n, y_1^n)| \leq \exp\left\{n(H(V_1, U_2 \oplus U_3|Q, Y_1) + 32\eta_2)\right\}$. Substituting this upper bound, the inner most summation turns out to be

$$\sum_{\substack{(\hat{v}_1^n, u_2^n \oplus u_3^n) \in \\ T_{8\eta_2}(V_1, \oplus|q^n, y_1^n)}} \frac{P\left(\substack{M_1 = m_1, V_1^n(m_1, b_1) = v_1^n \\ \underline{M^t} = \underline{m^t}, V_1^n(\hat{m}_1, \hat{b}_1) = \hat{v}_1^n}\right)}{\pi^{2n+t_2+t_3}} \leq \exp\left\{-n\left(\substack{2H(V_1|Q) - 52\eta_2 \\ -H(V_1|Q, U_2 \oplus U_3, Y_1)}\right)\right\} \frac{P(M_1 = m_1, \underline{M^t} = \underline{m^t})}{\pi^{2n+t_2+t_3}\mathcal{L}(n)} =: \beta_1,$$

$$\sum_{\substack{(\hat{v}_1^n, \hat{u}^n) \in \\ T_{8\eta_2}(V_1, \oplus|q^n, y_1^n)}} \frac{P\left(\substack{M_1 = m_1, V_1^n(m_1, b_1) = v_1^n \\ \underline{M^t} = \underline{m^t}, V_1^n(\hat{m}_1, \hat{b}_1) = \hat{v}_1^n}\right)}{\pi^{3n+t_2+t_3}} \leq \exp\left\{-n\left(\substack{2H(V_1|Q) - 52\eta_2 \\ -H(V_1, U_2 \oplus U_3|Q, Y_1)}\right)\right\} \frac{P(M_1 = m_1, \underline{M^t} = \underline{m^t})}{\pi^{3n+t_2+t_3}\mathcal{L}(n)} =: \beta_2$$

Substituting β_1 and β_2, we have

$$\begin{aligned}
P(\tilde{\epsilon}_{41}) \leq & \sum_{(m_1, \underline{m^t})} \sum_{\hat{m}_1 \neq m_1} \sum_{\substack{\hat{b}_1 \in \mathcal{B}_1 \\ \hat{a}^{s_3} \in \mathcal{F}_\pi^{s_3}}} \sum_{\substack{(b_1, \underline{a^s}) \in \\ \mathcal{B}_1 \times \mathscr{D}(\hat{a}^{s_3})}} \sum_{\substack{(v_1^n, \underline{u}^n) \in \\ T_{2\eta_2}(V_1, \underline{U}|q^n)}} \sum_{\substack{(x^n, y_1^n) \in \\ \tilde{T}_{\eta_2}(q^n|v_1^n, \underline{u}^n)}} \theta(y_1^n, x^n|v_1^n, \underline{u}^n)\beta_1 \\
& + \sum_{(m_1, \underline{m^t})} \sum_{\hat{m}_1 \neq m_1} \sum_{\hat{b}_1 \in \mathcal{B}_1} \sum_{\hat{a}_3^{s_3}} \sum_{\substack{(b_1, \underline{a^s}) \in \\ \mathcal{B}_1 \times \mathscr{I}(\hat{a}^{s_3})}} \sum_{\substack{(v_1^n, \underline{u}^n) \in \\ T_{2\eta_2}(V_1, \underline{U}|q^n)}} \sum_{\substack{(x^n, y_1^n) \in \\ \tilde{T}_{\eta_2}(q^n|v_1^n, \underline{u}^n)}} \theta(y_1^n, x^n|v_1^n, \underline{u}^n)\beta_2 \\
\leq & \sum_{(m_1, \underline{m^t})} \sum_{\hat{m}_1 \neq m_1} \sum_{\substack{\hat{b}_1 \in \mathcal{B}_1 \\ \hat{a}^{s_3} \in \mathcal{F}_\pi^{s_3}}} \sum_{\substack{(b_1, \underline{a^s}) \in \\ \mathcal{B}_1 \times \mathscr{D}(\hat{a}^{s_3})}} \sum_{\substack{(v_1^n, \underline{u}^n) \in \\ T_{2\eta_2}(V_1, \underline{U}|q^n)}} \beta_1 + \sum_{(m_1, \underline{m^t})} \sum_{\hat{m}_1 \neq m_1} \sum_{\hat{b}_1 \in \mathcal{B}_1} \sum_{\hat{a}_3^{s_3}} \sum_{\substack{(b_1, \underline{a^s}) \in \\ \mathcal{B}_1 \times \mathscr{I}(\hat{a}^{s_3})}} \sum_{\substack{(v_1^n, \underline{u}^n) \in \\ T_{2\eta_2}(V_1, \underline{U}|q^n)}} \beta_2
\end{aligned}$$

The terms in the first and second summation are identical to β_1 and β_2 respectively. Multiplying each with the corresponding number of terms, employing the lower bound for $\mathcal{L}(n)$ derived in (J.5), it maybe verified that $P(\tilde{\epsilon}_{41}) \leq \mathscr{T}_1 + \mathscr{T}_2$, where

$$\begin{aligned}
\mathscr{T}_1 &= 2\exp\left\{-n\left([I(V_1; U_2 \oplus U_3, Y_1|Q) - 56\eta_2] - \left[\frac{\log|\mathcal{B}_1|}{n} + \frac{\log|\mathcal{M}_1|}{n}\right]\right)\right\} \\
\mathscr{T}_2 &= 2\exp\left\{-n\left([\log\pi + H(V_1|Q) - H(V_1, U_2 \oplus U_3|Q, Y_1) - 56\eta_2] - \left[\frac{\log|\mathcal{B}_1|}{n} + \frac{\log|\mathcal{M}_1|}{n} + \frac{s_3\log\pi}{n}\right]\right)\right\}.
\end{aligned}$$

From bounds on the parameters of the code ((5.22) - (5.24)), it maybe verified that for $n \geq \max\{N_1(\eta), N_j(\eta_2) : j = 2, 3, 4\}$, $P((\epsilon_l \cup \epsilon_2 \cup \epsilon_3)^c \cap \epsilon_{41}) \leq 4\exp\left\{-n\left(\delta_1 + \frac{\eta}{4} - 56\eta_2\right)\right\}$.

Appendix L

Upper bound on $P\big((\epsilon_1 \cup \epsilon_2 \cup \epsilon_3)^c \cap \epsilon_{4j}\big)$ for 3−DBC

We begin by introducing some compact notation similar to that introduced in appendix K. We let $\underline{M^t}$ denote the pair $(M_2^{t_2}, M_3^{t_3})$ of message random variables. We let $\underline{m^t}$ denote a generic element $(m_2^{t_2}, m_3^{t_3}) \in \mathcal{F}_\pi^{\underline{t}} := \mathcal{F}_\pi^{t_2} \times \mathcal{F}_\pi^{t_3}$, and similarly $\underline{a^s}$ denote $(a_2^{s_2}, a_3^{s_3}) \in \mathcal{F}_\pi^{\underline{s}} := \mathcal{F}_\pi^{s_2} \times \mathcal{F}_\pi^{s_3}$. We let

$$\hat{T}_{\eta_2}(q^n) := \left\{(v_1^n, \underline{u}^n, x^n, y_j^n) \in T_{8\eta_2}(V_1, \underline{U}, X, Y_j|q^n) : (v_1^n, \underline{u}^n) \in T_{2\eta_2}(V_1, \underline{U}|q^n), (v_1^n, \underline{u}^n, x^n) \in T_{4\eta_2}(V_1, \underline{U}, X|q^n)\right\},$$
$$\hat{T}_{\eta_2}(q^n|v_1^n, \underline{u}^n) = \left\{(x^n, y_j^n) : (v_1^n, \underline{u}^n, x^n, y_j^n) \in \hat{T}_{\eta_2}(q^n)\right\}$$

We begin by characterizing the event under question. For $j = 2, 3$, denoting $\tilde{\epsilon}_{4j} := (\epsilon_l \cup \epsilon_2 \cup \epsilon_3)^c \cap \epsilon_{4j}$, we have

$$P(\tilde{\epsilon}_{4j}) \leq \sum_{(m_1, \underline{m^t})} \sum_{\hat{m}_j^{t_j} \neq M_j^{t_j}} \sum_{\hat{a}_j^{s_j}} \sum_{\substack{(v_1^n, \underline{u}^n, x^n, y_j^n) \\ \in \hat{T}_{\eta_2}(q^n)}} \sum_{\substack{\hat{u}_j^{s_j} \in \\ T_{8\eta_2}(U_j|q^n, y_j^n)}} P\left(\left\{\begin{smallmatrix} M_1=m_1, \underline{M^t}=\underline{m^t}, V_1^n(m_1,B_1)=v_1^n \\ U_l^n(A_l^{s_l})=u_l^n, I_l(A^{s_l})=m_l^{t_l}: l=2,3, Y_j^n=y_j^n \\ X^n=x^n, U_j^n(\hat{a}_j^{s_j})=\hat{u}_j^n, I_j(\hat{a}_j^{s_j})=\hat{m}_j^{t_j} \end{smallmatrix}\right\} \cap \epsilon_l^c\right), \tag{L.1}$$

where X^n abbreviates $X^n(M_1, \underline{M^t})$, the random vector input on the channel. We consider a generic term in the above sum. Observe that

$$P\left(\begin{smallmatrix} Y_j^n=y_j^n \\ X^n=x^n \end{smallmatrix}\middle| \left\{\begin{smallmatrix} M_1=m_1, \underline{M^t}=\underline{m^t}, V_1^n(m_1,B_1)=v_1^n \\ U_l^n(A_l^{s_l})=u_l^n, I_l(A^{s_l})=m_l^{t_l}: l=2,3 \\ U_j^n(\hat{a}_j^{s_j})=\hat{u}_j^n, I_j(\hat{a}_j^{s_j})=\hat{m}_j^{t_j} \end{smallmatrix}\right\} \cap \epsilon_l^c\right) = P\left(\begin{smallmatrix} Y_j^n=y_j^n \\ X^n=x^n \end{smallmatrix}\middle| \begin{smallmatrix} V_1^n(M_1,B_1)=v_1^n \\ U_l^n(A_l^{s_l})=u_l^n: l=2,3 \end{smallmatrix}\right) =: \theta(y^n, x^n|v_1^n, \underline{u}^n), \tag{L.2}$$

$$P\left(\left\{\begin{smallmatrix} M_1=m_1, \underline{M^t}=\underline{m^t}, V_1^n(m_1,B_1)=v_1^n \\ U_l^n(A_l^{s_l})=u_l^n, I_l(A^{s_l})=m_l^{t_l}: l=2,3 \\ U_j^n(\hat{a}_j^{s_j})=\hat{u}_j^n, I_j(\hat{a}_j^{s_j})=\hat{m}_j^{t_j} \end{smallmatrix}\right\} \cap \epsilon_l^c\right) = \sum_{\substack{(b_1, \underline{a^s}) \\ \in \mathcal{B}_1 \times \mathcal{F}_\pi^{\underline{s}}}} P\left(E \cap \left\{\begin{smallmatrix} B_1=b_1 \\ \underline{A^s}=\underline{a^s} \end{smallmatrix}\right\} \cap \epsilon_l^c\right) \leq \sum_{\substack{(b_1, \underline{a^s}) \\ \in \mathcal{B}_1 \times \mathcal{F}_\pi^{\underline{s}}}} P(E) P\left(\left\{\begin{smallmatrix} B_1=b_1 \\ \underline{A^s}=\underline{a^s} \end{smallmatrix}\right\} | E \cap \epsilon_l^c\right), \tag{L.3}$$

where E abbreviates the event $\left\{M_1{=}m_1,\ \underline{M}^{t}{=}\underline{m}^{t},\ V_1^n(m_1,b_1){=}v_1^n,\ U_l^n(a_l^{s_l}){=}u_l^n,\ I_l(a^{s_l}){=}m_l^{t_l}{:}l{=}2,3,\ U_j^n(\hat{a}_j^{s_j}){=}\hat{u}_j^n,\ I_j(\hat{a}_j^{s_j}){=}\hat{m}_j^{t_j}\right\}$. We now focus on the terms on the right hand side of (L.3). By the encoding rule, $P\left(\left\{B_1{=}b_1,\ \underline{A}^{s}{=}\underline{a}^{s}\right\}|E\cap\epsilon_l^c\right) = \frac{1}{\mathcal{L}(n)}$. We are left to evaluate $P(E)$. The collection $M_1, M_2^{t_2}, M_3^{t_3}, V_1^n(m_1,b_1), I_2(a^{s_2}), I_3(a^{s_3}), I_j(\hat{a}^{s_j}), (U_l(a_l^{s_l}) : l=2,3, U_j(\hat{a}_j^{s_j}))$ are mutually independent, where $(U_l(a_l^{s_l}) : l=2,3, U_j(\hat{a}_j^{s_j}))$ is treated as a single random object. The following counting argument proves the triplet $U_l(a_l^{s_l}) : l=2,3, U_j(\hat{a}_j^{s_j})$ also to be mutually independent. Let $\{j,\not{j}\}=\{2,3\}$. For any $u_j^n, u_{\not{j}}^n$ and $\hat{u}_j^n$, let us study

$$\left|\left\{(g_2, g_{3/2}, b_2^n, b_3^n) : a_j^{s_j} g_j \oplus b_j^n = u_j^n, a_{\not{j}}^{s_{\not{j}}} g_{\not{j}} \oplus b_{\not{j}}^n = u_{\not{j}}^n, (\hat{a}_j^{s_j} \ominus a_j^{s_j}) g_j = \hat{u}_j^n - u_j^n\right\}\right|.$$

There exists a t such that $\hat{a}_{jt}^{s_j} \neq a_{jt}^{s_j}$. For any choice of rows $1, 2, \cdots, t-1, t+1, \cdots, s_3$ of g_3, one can choose the tth row of g_j and b_2^n, b_3^n such that the above conditions are satisfied. The cardinality of the above set is $\pi^{(s_3-1)n}$. The uniform distribution and mutual independence guarantee $P(U_l(a_l^{s_l}) = u_l^n : l=2,3, U_j(\hat{a}_j^{s_j}) = \hat{u}_j^n) = \frac{1}{\pi^{3n}}$.

We therefore have

$$P\begin{pmatrix} M_1{=}m_1, \underline{M}^{t}{=}\underline{m}^{t}, V_1^n(m_1,b_1){=}v_1^n, \\ U_l^n(a_l^{s_l}){=}u_l^n, I_l(a^{s_l}){=}m_l^{t_l}{:}l{=}2,3, \\ U_j^n(\hat{a}_j^{s_j}){=}\hat{u}_j^n, I_j(\hat{a}_j^{s_j}){=}\hat{m}_j^{t_j} \end{pmatrix} = \frac{P(M_1 = m_1, \underline{M}^{t} = \underline{m}^{t}, V_1^n(m_1,b_1) = v_1^n)}{\pi^{3n+t_2+t_3+t_j}} \tag{L.4}$$

Substituting (L.4), (L.3) and (L.2) in (L.1), we have

$$P(\tilde{\epsilon}_{4j}) \leq \sum_{(m_1,\underline{m}^{t})} \sum_{(b_1,\underline{a}^{s})} \sum_{\hat{m}_j^{t_j} \neq m_j^{t_j}} \sum_{\hat{a}_j^{s_j}} \sum_{\substack{(v_1^n,\underline{u}^n,x^n,y_j^n) \\ \in \hat{T}_{\eta_2}(q^n)}} \theta(y^n,x^n|v_1^n,\underline{u}^n) \sum_{\substack{\hat{u}_j^n \in \\ T_{16\eta_2}(U_j|q^n,y_j^n)}} \frac{P(M_1 = m_1, \underline{M}^{t} = \underline{m}^{t}, V_1^n(m_1,b_1) = v_1^n)}{\pi^{3n+t_2+t_3+t_j}\mathcal{L}(n)}.$$

Note that terms in the innermost sum do not depend on the arguments of the sum. We now employ the bounds on the cardinality of conditional typical sets (lemma 2.4.2). There exists $N_5(\eta_2) \in \mathbb{N}$ such that for all $n \geq N_5(\eta_2)$, we have $|T_{16\eta_2}(U_j|q^n,y_j^n)| \leq \exp\{n(H(U_j|Q,Y_j)+32\eta_2)\}$ for all $(q^n,y_j^n) \in T_{8\eta_2}(Q,Y_j)$. For $n \geq \max\{N_1(\eta), N_5(\eta_2)\}$, we therefore have

$$\begin{aligned} P(\tilde{\epsilon}_{4j}) &\leq \sum_{(m_1,\underline{m}^{t})} \sum_{(b_1,\underline{a}^{s})} \sum_{\hat{m}_j^{t_j} \neq m_j^{t_j}} \sum_{\hat{a}_j^{s_j}} \sum_{\substack{(v_1^n,\underline{u}^n) \\ \in T_{2\eta_2}(V_1,\underline{U}|q^n)}} \frac{P\begin{pmatrix} V_1(m_1,b_1){=}v_1^n, M_1{=}m_1 \\ M_l^{t_l}{=}m_l^{t_l}{:}l{=}2,3 \end{pmatrix} \exp\{n32\eta_2\}}{\pi^{3n+t_2+t_3+t_j}\exp\{-nH(U_j|Q,Y_j)\}} \sum_{\substack{(x^n,y_j^n)\in \\ \hat{T}_{\eta_2}(q^n|v_1^n,\underline{u}^n)}} \frac{\theta(y^n,x^n|v_1^n,\underline{u}^n)}{\mathcal{L}(n)} \\ &\leq \sum_{(m_1,\underline{m}^{t})} \sum_{(b_1,\underline{a}^{s})} \sum_{\hat{m}_j^{t_j} \neq m_j^{t_j}} \sum_{\hat{a}_j^{s_j}} \sum_{\substack{(v_1^n,\underline{u}^n) \\ \in T_{2\eta_2}(V_1,\underline{U}|q^n)}} \frac{P\begin{pmatrix} V_1(m_1,b_1){=}v_1^n, M_1{=}m_1 \\ M_l^{t_l}{=}m_l^{t_l}{:}l{=}2,3 \end{pmatrix} \exp\{n32\eta_2\}}{\pi^{3n+t_2+t_3+t_j}\exp\{-nH(U_j|Q,Y_j)\}} \frac{1}{\mathcal{L}(n)} \\ &\leq 2\exp\{s_j\log\pi - n(\log\pi - H(U_j|Q,Y_j) - 32\eta_2)\} \leq 2\exp\{-n(\delta_1 - 32\eta_2)\}, \end{aligned} \tag{L.5}$$

where (L.5) follows from definition of $\mathcal{L}(n)$, (J.1) and the bounds on the parameters of the code derived in (5.22) - (5.24).

Appendix M

An upper bound on $P(\epsilon_5)$

In this appendix, we derive an upper bound on $P(\epsilon_5)$. As is typical in proofs of channel coding theorems, this step involves establishing statistical independence of cosets $C_j(M_j^{l_j}) : j = 1, 2$ corresponding to the message pair and any codeword $V^n(\hat{a}^k, \hat{m}^l)$ in a competing coset. We establish this in lemma M.0.19. We begin with the necessary spadework. Throughout this appendix, we employ the notation introduced in proof of theorem 6.2.2.

Lemma M.0.18 *If $m^l \neq \hat{m}^l$, then for any triple $\nu_1, \nu_2, \hat{\nu} \in \mathcal{V}^n$,*

$$P\left(\begin{smallmatrix} V_j^n(0^{k_j}, m_j^{l_j}) = \nu_j^n : j = 1,2, \\ V^n(0^k, \hat{m}^l) = \hat{\nu}^n \end{smallmatrix}\right) = P\left(V_j^n(0^{k_j}, m_j^{l_j}) = \nu_j^n : j = 1, 2\right) P\left(V^n(0^k, \hat{m}^l) = \hat{\nu}^n\right)$$

□

Proof: By definition of $V_j(0_{k_j}, m_j^{l_j}) : j = 1, 2$ and $V(0^k, m^l)$,

$$\begin{aligned} P\left(\begin{smallmatrix} V_j^n(0^{k_j}, m_j^{l_j}) = \nu_j^n : j = 1,2, \\ V^n(0^k, \hat{m}^l) = \hat{\nu}^n \end{smallmatrix}\right) &= P\left(\begin{smallmatrix} \left[m_1^{l_1} \ 0^{l_2}\right] G_{O/I} \oplus B_1^n = \nu_1^n, \left[0^{l_1} \ m_2^{l_2}\right] G_{O/I} \oplus B_2^n = \nu_2^n \\ \left[\hat{m}_1^{l_1} \ \hat{m}_2^{l_2}\right] G_{O/I} \oplus B_1^n \oplus B_2^n = \hat{\nu}^n \end{smallmatrix}\right) \\ &= P\left(\begin{smallmatrix} \left[m_1^{l_1} \ 0^{l_2}\right] G_{O/I} \oplus B_1^n = \nu_1^n, \left[0^{l_1} \ m_2^{l_2}\right] G_{O/I} \oplus B_2^n = \nu_2^n \\ \left[\tilde{m}_1^{l_1} \ \tilde{m}_2^{l_2}\right] G_{O/I} = \hat{\nu}^n \end{smallmatrix}\right) \end{aligned} \quad \text{(M.1)}$$

where $\tilde{m}_j^{l_j} = \hat{m}_j^{l_j} - m_j^{l_j}$. We now prove, using a counting argument similar to that employed in proof of lemma A.0.1, the term on right hand side of (M.1) is $\frac{1}{\pi^{3n}}$. Since $\hat{m}^l \neq m^l$, there exists $t \in [l]$ such that $\hat{m}_t \neq m_t$. Given any $(l-1)$ vectors $g_{O/I,j} \in \mathcal{V}^n : j \in [l] \setminus \{t\}$, there exists a unique triple of vectors $(g_{O/I,t}, b_1^n, b_2^n) \in \mathcal{V}^n \times \mathcal{V}^n \times \mathcal{V}^n$ such that $\left[m_1^{l_1} \ 0^{l_2}\right] g_{O/I} \oplus b_1^n = \nu_1^n$, $\left[0^{l_1} \ m_2^{l_2}\right] g_{O/I} \oplus b_2^n = \nu_2^n$ and $\left[\tilde{m}_1^{l_1} \ \tilde{m}_2^{l_2}\right] g_{O/I} = \hat{\nu}^n$, where row j of $g_{O/I}$ is $g_{O/I,j}$. Hence

$$\left| \left\{ (g_{O/I}, b_1^n, b_2^n) \in \mathcal{V}^{k \times n} \times \mathcal{V}^n \times \mathcal{V}^n : \begin{smallmatrix} \left[m_1^{l_1} \ 0^{l_2}\right] g_{O/I} \oplus B_1^n = \nu_1, \left[0^{l_1} \ m_2^{l_2}\right] g_{O/I} \oplus B_2^n = \nu_2 \\ \left[\tilde{m}_1^{l_1} \ \tilde{m}_2^{l_2}\right] g_{O/I} = \hat{\nu}^n \end{smallmatrix} \right\} \right| = \pi^{(l-1)n}.$$

The mutual independence and uniform distribution of $G_{O/I}, B_1, B_2^n$ implies the term on RHS of (M.1) is indeed $\frac{1}{\pi^{3n}}$. It remains to prove

$$P\left(V_j^n(0^{k_j}, m_j^{l_j}) = \nu_j^n : j = 1,2\right) P\left(V^n(0^k, \hat{m}^l) = \hat{\nu}^n\right) = \frac{1}{\pi^{3n}}.$$

It follows from lemma A.0.1 that $P\left(V^n(0^k, \hat{m}^l) = \hat{\nu}^n\right) = \frac{1}{\pi^n}$. Using the definition of $V^n(0^k, \hat{m}^l)$, we only need to prove

$$P\left(\begin{smallmatrix}\left[m_1^{l_1}\ 0^{l_2}\right] G_{O/I} \oplus B_1^n = \nu_1, \\ \left[0^{l_1}\ m_2^{l_2}\right] G_{O/I} \oplus B_2^n = \nu_2\end{smallmatrix}\right) = \frac{1}{\pi^{2n}}.$$

This follows again from a counting argument. For every matrix $g_{O/I} \in \mathcal{V}^{l\times n}$, there exists a unique pair of vectors $b_1^n, b_2^n \in \mathcal{V}^n$ such that $\left[m_1^{l_1}\ 0^{l_2}\right] G_{O/I} \oplus B_1^n = \nu_1$, and $\left[0^{l_1}\ m_2^{l_2}\right] G_{O/I} \oplus B_2^n = \nu_2$ thus yielding

$$\left|\left\{(g_{O/I}, b_1^n, b_2^n) \in \mathcal{V}^{k\times n} \times \mathcal{V}^n \times \mathcal{V}^n : \begin{smallmatrix}\left[m_1^{l_1}\ 0^{l_2}\right] G_{O/I} \oplus B_1^n = \nu_1, \\ \left[0^{l_1}\ m_2^{l_2}\right] G_{O/I} \oplus B_2^n = \nu_2\end{smallmatrix}\right\}\right| = \pi^{ln}, \tag{M.2}$$

and the proof is completed using the mutual independence and uniform distribution of $G_{O/I}, B_1^n, B_2^n$. ■

Lemma M.0.19 *For any $\hat{m}^l \neq m^l$, and any $\hat{a}^k \in \mathcal{V}^k$, the pair of cosets $C_j(m_j^{l_j}) : j = 1,2$ is statistically independent of $V^n(\hat{a}^k, \hat{m}^l)$.* □

Proof: For $j = 1,2$, let $\nu_j^n(a_j^{k_j}) \in \mathcal{V}^n$ for each $a_j^{k_j} \in \mathcal{V}^{k_j}$, and $\hat{\nu}^n \in \mathcal{V}^n$. We need to prove

$$P\left(\begin{array}{l} C_1^n(m_1^{l_1}) = (\nu_1(a_1^{k_1}) : a_1^{k_1} \in \mathcal{V}^{k_1}) \\ C_2^n(m_2^{l_2}) = (\nu_2(a_2^{k_2}) : a_2^{k_2} \in \mathcal{V}^{k_2}) \\ V^n(\hat{a}^k, \hat{m}^l) = \hat{\nu}^n \end{array}\right) = P\left(\begin{array}{l} C_1^n(m_1^{l_1}) = (\nu_1(a_1^{k_1}) : a_1^{k_1} \in \mathcal{V}^{k_1}) \\ C_2^n(m_2^{l_2}) = (\nu_2(a_2^{k_2}) : a_2^{k_2} \in \mathcal{V}^{k_2}) \end{array}\right) P(\substack{V^n(\hat{a}^k, \hat{m}^l) \\ = \hat{\nu}^n})$$

for every choice of $\nu_j(a_j^{k_j}) \in \mathcal{V}^n : a_j^{k_j} \in \mathcal{V}^{k_j}, j = 1,2$ and $\hat{\nu}^n \in \mathcal{V}^n$.

If (i) for some $j = 1$ or $j = 2$, $(\nu_j(a_j^{k_j} \oplus \tilde{a}_j^{k_j}) - \nu_j(0^{k_j})) \neq (\nu_j(a_j^{k_j}) - \nu_j(0^{k_j})) \oplus (\nu_j(\tilde{a}_j^{k_j}) - \nu_j(0^{k_j}))$ for any pair $a_j^{k_j}, \tilde{a_j}^{k_j} \in \mathcal{V}^{k_j}$, or (ii) $\nu_1(a^{k_1}) - v_1(0^{k_1}) \neq \nu_2(a_1^{k_1} 0^{k_+}) - v_2(0^{k_2})$ for some $a_1^{k_1} \in \mathcal{V}^{k_1}$, then LHS and first term of RHS are zero and equality holds. Otherwise,

$$\begin{aligned}
& P\left(C_j^n(m_j^{l_j}) = (\nu_j(a_j^{k_j}) : a_j^{k_j} \in \mathcal{V}^{k_j}) : j = 1,2, V^n(\hat{a}^k, \hat{m}^l) = \hat{\nu}^n\right) \\
&= P\left(\begin{smallmatrix} a_2^{k_2} G_{I_2} = \nu_2(a^{k_2}) - \nu_2(0^{k_2}) : a_2^{k_2} \in \mathcal{V}^{k_2}, V_j^n(0^{k_j}, m_j^{l_j}) = \nu_j(0^{k_j}) : j = 1,2, \\ V^n(0^k, \hat{m}^l) = \hat{\nu}^n - (\nu_2(\hat{a}^k) - \nu_2(0^{k_2}))\end{smallmatrix}\right) && \text{(M.3)} \\
&= P\left(\begin{smallmatrix} a_2^{k_2} G_{I_2} = \nu_2(a^{k_2}) - \\ \nu_2(0^{k_2}) : a_2^{k_2} \in \mathcal{V}^{k_2}\end{smallmatrix}\right) P\left(\begin{smallmatrix} V_j^n(0^{k_j}, m_j^{l_j}) = \nu_j(0^{k_j}) : j = 1,2, \\ V^n(0^k, \hat{m}^l) = \hat{\nu}^n - (\nu_2(\hat{a}^k) - \nu_2(0^{k_2}))\end{smallmatrix}\right) && \text{(M.4)} \\
&= P\left(\begin{smallmatrix} a_2^{k_2} G_{I_2} = \nu_2(a^{k_2}) - \\ \nu_2(0^{k_2}) : a_2^{k_2} \in \mathcal{V}^{k_2}\end{smallmatrix}\right) P\left(\begin{smallmatrix}\left[m_1^{l_1}\ 0^{l_2}\right] G_{O/I} \oplus B_1^n = \nu_1(0^{k_1}), \\ \left[0^{l_1}\ m_2^{l_2}\right] G_{O/I} \oplus B_2^n = \nu_2(0^{k_2})\end{smallmatrix}\right) P\left(V^n(\hat{a}^k, \hat{m}^l) = \hat{\nu}^n\right) && \text{(M.5)} \\
&= P\left(\begin{smallmatrix} a_2^{k_2} G_{I_2} = \nu_2(a^{k_2}) - \nu_2(0^{k_2}) : a_2^{k_2} \in \mathcal{V}^{k_2} \\ \left[m_1^{l_1}\ 0^{l_2}\right] G_{O/I} \oplus B_1^n = \nu_1(0^{k_1}), \left[0^{l_1}\ m_2^{l_2}\right] G_{O/I} \oplus B_2^n = \nu_2(0^{k_2})\end{smallmatrix}\right) P\left(V^n(\hat{a}^k, \hat{m}^l) = \hat{\nu}^n\right) && \text{(M.6)} \\
&= P\left(C_j^n(m_j^{l_j}) = (\nu_j(a_j^{k_j}) : a_j^{k_j} \in \mathcal{V}^{k_j}) : j = 1,2\right) P\left(V^n(\hat{a}^k, \hat{m}^l) = \hat{\nu}^n\right) && \text{(M.7)}
\end{aligned}$$

where i) (M.5) and (M.7) follow from definition of cosets $C_j(m_j^{l_j})$, (ii) (M.4) and (M.6) follow from independence of G_{I_2} and the collection $(G_{O/I}, B_1^n, B_2^n)$ and (iii) (M.3) follows from lemma M.0.18. ∎

We emphasize consequence of lemma M.0.19 in the following remark.

Remark M.0.20 *If $m^l \neq \hat{m}^l$, then conditioned on event $\{M^l = m^l\}$, received vector Y^n is statistically independent of $V^n(\hat{a}^k, \hat{m}^l)$ for any $\hat{a}^k \in \mathcal{V}^k$. We establish truth of this statement in the sequel. Let $\mathcal{C}_j$ denote the set of all ordered π^{k_j}-tuples of vectors in $\mathcal{V}^n$. Observe that*

$$\begin{aligned}
P\left(\substack{M^l=m^l, Y^n=y^n, \\ V^n(\hat{a}^k,\hat{m}^l)=\hat{\nu}^n}\right) &= \sum_{C_1\in\mathcal{C}_1}\sum_{C_2\in\mathcal{C}_2}\sum_{\boldsymbol{s}^n\in\mathcal{S}^n} P\left(\substack{M^l=m^l, C_j(m_j^{l_j})=C_j:j=1,2, S^n=s^n \\ V^n(\hat{a}^k,\hat{m}^l)=\hat{v}^n, Y^n=y^n}\right) \\
&= \sum_{C_1\in\mathcal{C}_1}\sum_{C_2\in\mathcal{C}_2}\sum_{\boldsymbol{s}^n\in\mathcal{S}^n} P\left(\substack{M^l=m^l \\ S^n=s^n}\right) P\left(\substack{C_1(m_1^{l_1})=C_1 \\ C_2(m_2^{l_2})=C_2}\right) P\left(V^n(\hat{a}^k,\hat{m}^l)=\hat{v}^n\right) P\left(Y^n=y^n \middle|^{C_j(m_j^{l_j})=C_j:j=1,2}_{S^n=s^n, M^l=m^l}\right) \qquad \text{(M.8)} \\
&= \sum_{C_1\in\mathcal{C}_1}\sum_{C_2\in\mathcal{C}_2}\sum_{\boldsymbol{s}^n\in\mathcal{S}^n} P\left(\substack{M^l=m^l, Y^n=y^n, S^n=s^n \\ C_j(m_j^{l_j})=C_j:j=1,2}\right) P\left(V^n(\hat{a}^k,\hat{m}^l)=\hat{v}^n\right) \\
&= P\left(M^l=m^l, Y^n=y^n\right) P\left(V^n(\hat{a}^k,\hat{m}^l)=\hat{\nu}^n\right)
\end{aligned}$$

where (N.5) follows from (i) independence of random objects that characterize codebook and (S^n, M^l), (ii) lemma M.0.19 and (iii) statistical independence of the inputs $X_j(M_j^{l_j}, S_j^n) : j = 1, 2$ to the channel and the codeword $V^n(\hat{a}^k, \hat{m}^l)$ conditioned on the specific realization of cosets $(C_j(M_j^{l_j}) : j = 1, 2)$ and the event $\{M^l = m^l\}$. Moreover, since $P(V^n(\hat{a}^k, \hat{m}^l) = \hat{\nu}^n) = \frac{1}{\pi^n}$, we have $P(M^l = m^l, Y^n = y^n, V^n(\hat{a}^k, \hat{m}^l) = \hat{\nu}^n) = \frac{1}{\pi^n} P(M^l = m^l, Y^n = y^n)$.

We are now equipped to derive an upper bound on $P(\epsilon_5)$. Observe that

$$\begin{aligned}
P(\epsilon_5) &\leq P\left(\bigcup_{\hat{a}^k\in\mathcal{V}^k}\bigcup_{\substack{m^l,\hat{m}^l \\ m^l\neq\hat{m}^l}}\left\{\substack{(V^n(\hat{a}^k,\hat{m}^l),Y^n)\in T_{\eta_5(\eta)}(p_{V_1\oplus V_2,Y}) \\ M^l=m^l}\right\}\right) \\
&\leq \sum_{\hat{a}^k\in\mathcal{V}^k}\sum_{\substack{m^l,\hat{m}^l \\ m^l\neq\hat{m}^l}}\sum_{y^n\in T_{\eta_5(\eta)}(Y)}\sum_{v^n\in T_{\eta_5(\eta)}(V_1\oplus V_2|y^n)} P\left(\substack{V^n(a^k,\hat{m}^l)=v^n \\ M^l=m^l, Y^n=y^n}\right) \\
&\leq \sum_{\hat{a}^k\in\mathcal{V}^k}\sum_{\substack{m^l,\hat{m}^l \\ m^l\neq\hat{m}^l}}\sum_{y^n\in T_{\eta_5(\eta)}(Y)}\sum_{v^n\in T_{\eta_5(\eta)}(V_1\oplus V_2|y^n)} P\left(V^n(a^k,\hat{m}^l)=v^n\right) P(M^l=m^l, Y^n=y^n) \\
&\leq \sum_{\hat{a}^k\in\mathcal{V}^k}\sum_{\hat{m}^l\in\mathcal{V}^l}\sum_{y^n\in T_{\eta_5(\eta)}(Y)}\sum_{v^n\in T_{\eta_5(\eta)}(V_1\oplus V_2|y^n)} \frac{P(Y^n=y^n)}{\pi^n} \\
&\leq \sum_{y^n\in T_{\eta_5(\eta)}(Y)} \frac{\pi^{k+l}|T_{2\eta_5(\eta)}(V_1\oplus V_2|y^n)|}{\pi^n} \leq \exp\left\{-n\log\pi\left(1-\frac{H(V_1\oplus V_2|Y)+3\eta_5(\eta)}{\log\pi}-\frac{k+l}{n}\right)\right\}. \qquad \text{(M.9)}
\end{aligned}$$

where (N.6) follows from the uniform bound of $\exp\{n(H(V_1 \oplus V_2|Y) + 3\eta_5(\eta))\}$ on $|T_{2\eta_5(\eta)}(V_1 \oplus V_2|y^n)|$ for any $y^n \in T_{\eta_5(\eta)}(Y)$, $n \geq N_6(\eta)$ provided by lemma 2.4.2 for $n \geq N_6(\eta)$. Substituting the upper bound for $\frac{k+l}{n}$ in (6.13), we have

$$P(\epsilon_5) \leq \exp\{-n(\eta_2(\eta) + \eta_3(\eta) - 3\eta_5(\eta))\} \text{ for all } n \geq \max\{N_1(\eta), N_6(\eta)\}. \tag{M.10}$$

Appendix N

An upper bound on $P(\epsilon_3)$

In this appendix, we derive an upper bound on $P(\epsilon_3)$. As is typical in proofs of channel coding theorems, this step involves establishing statistical independence of $C_j(hS_j^n) : j = 1, 2$ and any codeword $V^n(a^k, \hat{m}^l)$ in a competing coset $\hat{m}^l \neq hS_1^n \oplus hS_2^n$. We establish this in lemma N.0.22. We begin with the necessary spadework. The following lemmas holds for any $\mathcal{F}_q$ and we state it in this generality.

Lemma N.0.21 *Let $\mathcal{F}_q$ be a finite field. Let $G_I \in \mathcal{F}_q^{k\times n}$, $G_{O/I} \in \mathcal{F}_q^{l\times n}$, $B_j^n \in \mathcal{F}_q^n : j = 1, 2$ be mutually independent and uniformly distributed on their respective range spaces. Then the following hold.*

(a) $P(V^n(a^k, m^l) = v^n) = \frac{1}{q^n}$ *for any* $a^k \in \mathcal{F}_q^k$, $m^l \in \mathcal{F}_q^l$ *and* $v^n \in \mathcal{F}_q^n$,

(b) $P(V_j^n(a_j^k, m_j^l) = v_j^n : j = 1, 2) = \frac{1}{q^{2n}}$ *for any* $a_j^k \in \mathcal{F}_q^k$, $m_j^l \in \mathcal{F}_q^l$ *and* $v_j^n \in \mathcal{F}_q^n : j = 1, 2$, *and*

(c) $P\begin{pmatrix} V_j^n(0^k, m_j^l) = v_{j,0^k}^n : j=1,2, \\ V^n(0^k, \hat{m}^l) = v^n \end{pmatrix} = \frac{1}{q^{3n}}$ *for any* $\hat{m}^l \neq m_1^l \oplus m_2^l$ *and* $v_{j,0^k}^n : j = 1, 2$, *and* v^n.

□

Proof: The proof follows from a counting argument similar to that employed in [63, Remarks 1,2].

(i) For any $g_I \in \mathcal{F}_q^{k\times n}$, $g_{O/I} \in \mathcal{F}_q^{l\times n}$, $v^n \in \mathcal{F}_q^n$, there exists a unique $b^n \in \mathcal{F}_q^n$ such that $a^k g_I \oplus m^l g_{O/I} \oplus b^n = v^n$. Since G_I, $G_{O/I}$ and B^n are mutually independent and uniformly distributed $P(V^n(a^k, m^l) = v^n) = \frac{q^{kn}q^{ln}}{q^{kn}q^{ln}q^n} = \frac{1}{q^n}$.

(ii) We first note $P(V_j^n(a_j^k, m_j^l) = v_j^n : j = 1, 2) = P(a_j^k G_I \oplus m_j^l G_{O/I} \oplus B_j^n = v_j^n : j = 1, 2)$. For any choice of g_I and $g_{O/I}$, there exists unique $b_j^n : j = 1, 2$ such that $a_j^k g_I \oplus m_j^l g_{O/I} \oplus b_j^n = v_j^n : j = 1, 2$. Since G_I, $G_{O/I}$ and B^n are mutually independent and uniformly distributed, the probability in question is therefore $\frac{q^{kn}q^{ln}}{q^{kn}q^{ln}q^{2n}} = \frac{1}{q^{2n}}$.

(iii) Note that

$$P\begin{pmatrix} V_j^n(0^k, m_j^l) = v_{j,0^k}^n : j=1,2, \\ V^n(0^k, \hat{m}^l) = v^n \end{pmatrix} = P\begin{pmatrix} m_j^l G_{O/I} \oplus B_j^n = v_{j,0^k}^n : \\ j=1,2, \hat{m}^l G_{O/I} \oplus B^n = v^n \end{pmatrix} = P\begin{pmatrix} m_j^l G_{O/I} \oplus B_j^n = v_{j,0^k}^n : j=1,2, \\ (\hat{m}^l \ominus (m_1^l \oplus m_2^l)) G_{O/I} = v^n \end{pmatrix}$$

Since $\hat{m}^l \neq m_1^l \oplus m_2^l$, there exists an index t such that $\hat{m}_t \neq m_{1t} \oplus m_{2t}$. Therefore, given any set of rows $\underline{g}_{O/I,1} \cdots, \underline{g}_{O/I,t-1}, \underline{g}_{O/I,t+1}, \cdots, \underline{g}_{O/I,l}$, there exists a unique selection for row $\underline{g}_{O/I,t}$ such that $(\hat{m}^l \ominus (m_1^l \oplus m_2^l))g_{O/I} = v^n$. Having chosen this, choose $b_j^n = v_{j,0^k}^n \ominus m_j^l g_{O/I}$. Since G_I, $G_{O/I}$ and $B_j^n : j = 1, 2$ are mutually independent and uniformly distributed, the probability in question is therefore $\frac{q^{(l-1)n}}{q^{ln}q^{2n}} = \frac{1}{q^{3n}}$.

■

Lemma N.0.22 *If generator matrices $G_I \in \mathcal{F}_q^{k\times n}$, $G_{O/I} \in \mathcal{F}_q^{l\times n}$ and $B_j^n \in \mathcal{F}_q^n : j = 1, 2$ are mutually independent and uniformly distributed over their respective range spaces, then the pair of cosets $C_j(m_j^l) : j = 1, 2$ is independent of $V^n(\hat{a}^k, \hat{m}^l)$ whenever $\hat{m}^l \neq (m_1^l \oplus m_2^l)$.* □

Proof: Let $v_{j,a^k}^n \in \mathcal{F}_q^n$ for each $a^k \in \mathcal{F}_q^k$, $j = 1, 2$ and $\hat{v}^n \in \mathcal{F}_q^n$. We need to prove

$$\begin{aligned} &P(C_j^n(m_j^l) = (v_{j,a^k}^n : a^k \in \mathcal{F}_q^k) : j = 1, 2, V^n(\hat{a}^k, \hat{m}^l) = \hat{v}^n) \\ &= P(C_j^n(m_j^l) = (v_{j,a^k} : a^k \in \mathcal{F}_q^k) : j = 1, 2) P(V^n(\hat{a}^k, \hat{m}^l) = \hat{v}^n) \end{aligned} \tag{N.1}$$

for every choice of $v_{j,a^k} \in \mathcal{F}_q^n : a^k \in \mathcal{F}_q^k, j = 1, 2$ and $\hat{v}^n \in \mathcal{F}_q^n$.

If (i) for some $j = 1$ or $j = 2$, $(v_{j,a^k\oplus\tilde{a}^k}^n - v_{j,0^k}^n) \neq (v_{j,a^k}^n - v_{j,0^k}^n) \oplus (v_{j,\tilde{a}^k}^n - v_{j,0^k}^n)$ for any pair a^k, $\tilde{a}^k \in \mathcal{F}_q^k$, or (ii) $v_{1,a^k}^n - v_{1,0^k}^n \neq v_{2,a^k}^n - v_{2,0^k}^n$ for some $a^k \in \mathcal{F}_q^k$, then LHS and first term of RHS are zero and equality holds.

Otherwise, LHS of (N.1) is

$$\begin{aligned} P\left(C_j^n(m_j^l)=(v_{j,a^k}^n : a^k\in\mathcal{F}_q^k) : j=1,2, V^n(\hat{a}^k,\hat{m}^l)=\hat{v}^n\right) &= P\begin{pmatrix} a^kG_I = v_{1,a^k}^n - v_{1,0^k}^n : a^k \in \mathcal{F}_q^k, V_j^n(0^k, m_j^l) = v_{j,0^k}^n : j=1,2, \\ V^n(0^k,\hat{m}^l) = \hat{v}^n - (v_{1,\hat{a}^k}^n - v_{1,0^k}^n) \end{pmatrix} \\ &= P\begin{pmatrix} a^kG_I = v_{1,a^k}^n - \\ v_{1,0^k}^n : a^k \in \mathcal{F}_q^k \end{pmatrix} P\begin{pmatrix} V_j^n(0^k,m_j^l) = v_{j,0^k}^n : j=1,2, \\ V^n(0^k,\hat{m}^l) = \hat{v}^n - (v_{1,\hat{a}^k}^n - v_{1,0^k}^n) \end{pmatrix}, \end{aligned} \tag{N.2}$$

where we have used independence of G_I and $(G_{O/I}, B_1^n, B_2^n)$ in arriving at (N.2). Similarly RHS of (N.1) is

$$\begin{aligned} P\left(C_j^n(m_j^l)=(v_{j,a^k}^n : a^k\in\mathcal{F}_q^k) : j=1,2\right) P\left(V^n(\hat{a}^k,\hat{m}^l)=\hat{v}^n\right) &= P\begin{pmatrix} a^kG_I = v_{1,a^k}^n - v_{1,0^k}^n : a^k \in \mathcal{F}_q^k, \\ V_j^n(0^k, m_j^l) = v_{j,0^k}^n : j=1,2 \end{pmatrix} P\begin{pmatrix} a^kG_I \oplus \hat{m}^l G_{O/I} \oplus B^n = \\ \hat{v}^n \end{pmatrix} \\ &= P\begin{pmatrix} a^kG_I = v_{1,a^k}^n - \\ v_{1,0^k}^n : a^k \in \mathcal{F}_q^k \end{pmatrix} P\begin{pmatrix} V_j^n(0^k,m_j^l) = \\ v_{j,0^k}^n : j=1,2 \end{pmatrix} \cdot \frac{1}{q^n} \qquad \text{(N.3)} \\ &= P\begin{pmatrix} a^kG_I = v_{1,a^k}^n - \\ v_{1,0^k}^n : a^k \in \mathcal{F}_q^k \end{pmatrix} P\begin{pmatrix} m_j^l G_{O/I} \oplus B_j^n = \\ v_{j,0^k}^n : j=1,2 \end{pmatrix} \cdot \frac{1}{q^n} \\ &= P\left(a^kG_I = v_{1,a^k}^n - v_{1,0^k}^n : a^k \in \mathcal{F}_q^k\right) \cdot \frac{1}{q^{3n}}, \qquad \text{(N.4)} \end{aligned}$$

where (N.3), (N.4) follows from lemma N.0.21(a) and (b) respectively. Comparing simplified forms of LHS in (N.2) and RHS in (N.4), it suffices to prove

$$P\begin{pmatrix} V_j^n(0^k,m_j^l) = v_{j,0^k}^n : j=1,2, \\ V^n(0^k,\hat{m}^l) = \hat{v}^n - (v_{1,\hat{a}^k}^n - v_{1,0^k}^n) \end{pmatrix} = \frac{1}{q^{3n}}.$$

This follows from lemma N.0.21(c) ■

We emphasize consequence of lemma N.0.22 in the following.

Remark N.0.23 *If $\hat{m}^l \neq hs_1^n \oplus hs_2^n$, then conditioned on the event $\left\{S_j^n = s_j^n : j = 1,2\right\}$, received vector Y^n is statistically independent of $V^n(\hat{a}^k, \hat{m}^l)$ for any $\hat{a}^k \in \mathcal{S}^k$. We establish truth of this statement in the sequel. Let $\mathcal{C}$ denote the set of all ordered $|\mathcal{S}|^k$-tuples of vectors in $\mathcal{S}^n$. Observe that,*

$$\begin{aligned}
P\left(\begin{smallmatrix}\underline{s}^n=\underline{s}^n, Y^n=y^n,\\ V^n(\hat{a}^k,\hat{m}^l)=\hat{v}^n\end{smallmatrix}\right) &= \sum_{C_1\in\mathcal{C}}\sum_{C_2\in\mathcal{C}} P\left(\begin{smallmatrix}\underline{s}^n=\underline{s}^n, C_j(hs_j^n)=C_j:j=1,2,\\ V^n(\hat{a}^k,\hat{m}^l)=\hat{v}^n, Y^n=y^n\end{smallmatrix}\right)\\
&= \sum_{C_1\in\mathcal{C}_1}\sum_{C_2\in\mathcal{C}_2} P\left(\underline{s}^n=\underline{s}^n\right) P\left(\begin{smallmatrix}C_1(hs_1^n)=C_1\\ C_2(hs_2^n)=C_2\end{smallmatrix}\right) P\left(V^n(\hat{a}^k,\hat{m}^l)=\hat{v}^n\right)\cdot P\left(Y^n=y^n\big|^{C_j(hs_j^n)=C_j:j=1,2}_{\underline{s}^n=\underline{s}^n}\right) \qquad \text{(N.5)}\\
&= \sum_{C_1\in\mathcal{C}_1}\sum_{C_2\in\mathcal{C}_2} P\left(\begin{smallmatrix}\underline{s}^n=\underline{s}^n, Y^n=y^n,\\ C_j(hs_j^n)=C_j:j=1,2\end{smallmatrix}\right) P\left(V^n(\hat{a}^k,\hat{m}^l)=\hat{v}^n\right)\\
&= P\left(\underline{s}^n=\underline{s}^n, Y^n=y^n\right) P\left(V^n(\hat{a}^k,\hat{m}^l)=\hat{v}^n\right)
\end{aligned}$$

We have used (a) independence of $\underline{s}^n$ and random objects that characterize the codebook, (b) independence of $V^n(\hat{a}^k,\hat{m}^l)$ and $(C_j(hs_j^n): j=1,2)$ (lemma N.0.22), (c) $(\mu_1(hs_1^n),\mu_2(hs_2^n))$ being a function of $(C_1(hs_1^n),C_2(hs_2^n))$, is conditionally independent of $V^n(\hat{a}^k,\hat{m}^l)$ given $(C_1(hs_1^n),C_2(hs_2^n))$ in arriving at (N.5). Moreover, since $P(V^n(\hat{a}^k,\hat{m}^l)=\hat{v}^n)=\frac{1}{|\mathcal{S}|^n}$, we have $P\left(\underline{s}^n=\underline{s}^n, Y^n=y^n, V^n(\hat{a}^k,\hat{m}^l)=\hat{v}^n\right)=\frac{1}{|\mathcal{S}|^n}P(\underline{s}^n=\underline{s}^n, Y^n=y^n)$.

We are now equipped to derive an upper bound on $P(\epsilon_3)$. Observe that

$$\begin{aligned}
P(\epsilon_3) &\leq P\left(\bigcup_{\hat{a}^k\in\mathcal{S}^k}\bigcup_{\underline{s}^n=\underline{s}^n}\bigcup_{\substack{\hat{m}^l\neq\\ h(s_1^n\oplus s_2^n)}}\left\{\begin{smallmatrix}(V^n(\hat{a}^k,\hat{m}^l),Y^n)\in\\ T_{\eta_1}(p_{V_1\oplus V_2,Y}),\underline{s}^n=\underline{s}^n\end{smallmatrix}\right\}\right) \leq \sum_{\substack{\hat{a}^k\in\mathcal{S}^k,\\ \underline{s}^n=\underline{s}^n}}\sum_{\substack{\hat{m}^l\neq\\ h(s_1^n\oplus s_2^n)}}\sum_{\substack{y^n\in T_{\eta_1}(Y),v^n\in\\ T_{\eta_1}(V_1\oplus V_2|y^n)}} P\left(\begin{smallmatrix}V^n(a^k,\hat{m}^l)=v^n\\ \underline{s}^n=\underline{s}^n, Y^n=y^n\end{smallmatrix}\right)\\
&\leq \sum_{\substack{\hat{a}^k\in\mathcal{S}^k,\\ \underline{s}^n=\underline{s}^n}}\sum_{\substack{\hat{m}^l\neq\\ h(s_1^n\oplus s_2^n)}}\sum_{\substack{y^n\in T_{\eta_1}(Y),v^n\in\\ T_{\eta_1}(V_1\oplus V_2|y^n)}} P\left(\begin{smallmatrix}V^n(a^k,\hat{m}^l)\\ =v^n\end{smallmatrix}\right) P\left(\begin{smallmatrix}\underline{s}^n=\underline{s}^n,\\ Y^n=y^n\end{smallmatrix}\right)\\
&\leq \sum_{\hat{a}^k\in\mathcal{S}^k}\sum_{\substack{\hat{m}^l\neq\\ h(s_1^n\oplus s_2^n)}}\sum_{y^n\in T_{\eta_1}(Y)}\sum_{v^n\in T_{\eta_1}(V_1\oplus V_2|y^n)} \frac{P(Y^n=y^n)}{|\mathcal{S}|^n}\\
&\leq \sum_{y^n\in T_{\eta_1}(Y)} \frac{|\mathcal{S}|^{k+l}\,|T_{\eta_1}(V_1\oplus V_2|y^n)|}{|\mathcal{S}|^n} \leq \exp\left\{-n\log|\mathcal{S}|\left(1-\tfrac{H(V_1\oplus V_2|Y)+3\eta_1+k+l}{\log|\mathcal{S}|}\right)\right\}. \qquad \text{(N.6)}
\end{aligned}$$

where (N.6) follows from the uniform bound of $\exp\{n(H(V_1\oplus V_2|Y)+3\eta_1)\}$ on $|T_{\eta_1}(V_1\oplus V_2|y^n)|$ for any $y^n\in T_{\eta_1}(Y)$, $n\geq N_6(\eta)$ (Conditional frequency typicality) for $n\geq N_6(\eta)$.

Bibliography

[1] C. E. Shannon, "A mathematical theory of communication," Bell System Technical Journal, vol. 27, pp. 379–423, 623–656, July and October 1948.

[2] I. Csiszár and J. Körner, Information Theory: Coding Theorems for Discrete Memoryless Systems, 2nd ed. Budapest: Cambridge University Press, June 2011.

[3] R. Ahlswede, "Multi-way communication channels," in Proceedings of 2nd ISIT, Thakadsor, Armenian SSR, Sept 1971, pp. 23–52.

[4] H. Liao, "A coding theorem for multiple access communications," in Proceedings of 2nd ISIT, Thakadsor, Armenian SSR, 1972.

[5] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," Information Theory, IEEE Transactions on, vol. 19, no. 4, pp. 471 – 480, July 1973.

[6] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," IEEE Trans. Inform. Theory, vol. 22, no. 1, pp. 1–10, January 1976.

[7] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," Problems of Ctrl. and Info. Th., vol. 19, no. 1, pp. 19–31, 1980.

[8] S. I. Gel'fand, "Capacity of one broadcast channel," Probl. Pered. Inform., vol. 13, no. 3, pp. 106108, JulySept. 1977; translated in Probl. Inform. Transm., pp. 240242, JulySept. 1977.

[9] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," IEEE Trans. Inform. Theory, vol. IT-25, no. 3, pp. 306–311, May 1979.

[10] J. L. Massey, "Applied digital information theory I," available at http://www.isiweb.ee.ethz.ch/archive/massey_scr/adit1.pdf.

[11] A. Feinstein, "A new basic theorem of information theory," Information Theory, Transactions of the IRE Professional Group on, vol. 4, no. 4, pp. 2–22, 1954.

[12] I. Csiszár and J. Korner, "Graph decomposition: A new key to coding theorems," Information Theory, IEEE Transactions on, vol. 27, no. 1, pp. 5–12, 1981.

[13] T. Han and K. Kobayashi, "A new achievable rate region for the interference channel," Information Theory, IEEE Transactions on, vol. 27, no. 1, pp. 49 – 60, jan 1981.

[14] R. G. Gallager, Information Theory and Reliable Communication. New York: John Wiley & Sons, 1968.

[15] T. Philosof and R. Zamir, "On the loss of single-letter characterization: The dirty multiple access channel," IEEE Trans. Inform. Theory, vol. 55, pp. 2442–2454, June 2009.

[16] B. Nazer and M. Gastpar, "Computation over multiple-access channels," IEEE Trans. on Info. Th., vol. 53, no. 10, pp. 3498 –3516, oct. 2007.

[17] V. Cadambe and S. Jafar, "Interference alignment and degrees of freedom of the k -user interference channel," Information Theory, IEEE Transactions on, vol. 54, no. 8, pp. 3425–3441, 2008.

[18] J. Körner and K. Marton, "How to encode the modulo-two sum of binary sources (corresp.)," IEEE Trans. Inform. Theory, vol. 25, no. 2, pp. 219 – 221, Mar 1979.

[19] S. Sridharan, A. Jafarian, S. Vishwanath, S. Jafar, and S. Shamai, "A layered lattice coding scheme for a class of three user gaussian interference channels," in 2008 46th Annual Allerton Conference Proceedings on, sept. 2008, pp. 531 –538.

[20] T. M. Cover, "Broadcast channels," IEEE Trans. Inform. Theory, vol. IT-18, no. 1, pp. 2–14, Jan. 1972.

[21] P. P. Bergmans, "Random coding theorems for the broadcast channels with degraded components," IEEE Trans. Inform. Theory, vol. IT-15, pp. 197–207, Mar. 1973.

[22] G. Bresler, A. Parekh, and D. Tse, "The approximate capacity of the many-to-one and one-to-many gaussian interference channels," Information Theory, IEEE Transactions on, vol. 56, no. 9, pp. 4566 –4592, sept. 2010.

[23] D. Krithivasan and S. Pradhan, "Distributed source coding using abelian group codes: A new achievable rate-distortion region," IEEE Trans. Inform. Theory, vol. 57, no. 3, pp. 1495 –1519, march 2011.

[24] R. Sundaresan, "Lecture notes for E2-301 Topics in multi-user communications," Jan-Apr 2008, available at http://ece.iisc.ernet.in/ kprem/e2301.tar.gz.

[25] A. Orlitsky and J. Roche, "Coding for computing," Information Theory, IEEE Transactions on, vol. 47, no. 3, pp. 903 –917, mar 2001.

[26] A. E. Gamal and Y.-H. Kim, Network Information Theory, 1st ed. New York: Cambridge University Press, 2012.

[27] T. M. Cover and J. A. Thomas, Elements of Information Theory, 2nd ed. New York: John Wiley & Sons, 2006.

[28] W. Hoeffding, "Asymptotically optimal tests for multinomial distributions," Annals of Mathematical Statistics, vol. 36, no. 2, pp. 369–401, 1965.

[29] I. Sanov, "On the probability of large deviations of random variables," Matematicheskii Sbornik, vol. 42(84), pp. 11–44, 1957, translated by Dana E. A. Quade, Institute of Statistics, Mimeograph Series No. 192, March 1958, available at.

[30] R. Ahlswede, "Group codes do not achieve shannon's channel capacity for general discrete channels," The Annals of Mathematical Statistics, vol. 42, no. 1, pp. 224–240, February 1971.

[31] R. Barron, B. Chen, and G. W. Wornell, "The duality between information embedding and source coding with side information and some applications," Information Theory, IEEE Transactions on, vol. 49, no. 5, pp. 1159–1180, 2003.

[32] A. Padakandla and S. Pradhan, "Achievable rate region for three user discrete broadcast channel based on coset codes," available at `http://arxiv.org/abs/1207.3146`.

[33] A. Sahebi and S. Pradhan, "Nested lattice codes for arbitrary continuous sources and channels," in Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on, 2012, pp. 626–630.

[34] C. E. Shannon, "Two-way communication channels," Proc. 4th Berkeley Symp. Mathematical Statistics and Probability, vol. 1, pp. 611 –644, 1961.

[35] R. Ahlswede, "The capacity region of a channel with two senders and two receivers," Annals of Probability, vol. 2, no. 5, pp. 805–814, 1974.

[36] H. Sato, "Two-user communication channels," Information Theory, IEEE Transactions on, vol. 23, no. 3, pp. 295–304, 1977.

[37] A. Carleial, "A case where interference does not reduce capacity (corresp.)," Information Theory, IEEE Transactions on, vol. 21, no. 5, pp. 569–570, 1975.

[38] ——, "Interference channels," Information Theory, IEEE Transactions on, vol. 24, no. 1, pp. 60–70, 1978.

[39] H. Sato, "The capacity of the gaussian interference channel under strong interference (corresp.)," Information Theory, IEEE Transactions on, vol. 27, no. 6, pp. 786–788, 1981.

[40] A. Gamal and M. Costa, "The capacity region of a class of deterministic interference channels (corresp.)," Information Theory, IEEE Transactions on, vol. 28, no. 2, pp. 343–346, 1982.

[41] R. Etkin, D. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit," Information Theory, IEEE Transactions on, vol. 54, no. 12, pp. 5534–5562, 2008.

[42] H.-F. Chong, M. Motani, H. K. Garg, and H. El Gamal, "On the Han-Kobayashi region for the Interference Channel," Information Theory, IEEE Transactions on, vol. 54, no. 7, pp. 3188–3195, 2008.

[43] H.-F. Chong, M. Motani, and H. K. Garg, "A comparison of two achievable rate regions for the interference channel," San Diego, Feb. 2006.

[44] G. Kramer, "Review of rate regions for interference channels," in Communications, 2006 International Zurich Seminar on, 2006, pp. 162–165.

[45] K. Kobayashi and T. Han, "A further consideration on the HK and the CMG regions for the interference channel," San Diego, Feb. 2007.

[46] B. Bandemer and A. El Gamal, "Interference decoding for deterministic channels," Information Theory, IEEE Transactions on, vol. 57, no. 5, pp. 2966–2975, 2011.

[47] A. Padakandla and S. Pradhan, "Achievable rate region based on coset codes for multiple access channel with states," available at `http://arxiv.org/abs/1301.5655`.

[48] 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 Trans. on Info. Th., vol. 29, no. 3, pp. 396 – 412, may 1983.

[49] B. E. Hajek and M. B. Pursley, "Evaluation of an achievable rate region for the broadcast channel," IEEE Trans. Inform. Theory, vol. IT-25, no. 1, pp. 36–46, Jan. 1979.

[50] T. M. Cover, "An achievable rate region for the broadcast channel," IEEE Trans. Inform. Theory, vol. IT-21, no. 4, pp. 399–404, Jul. 1975.

[51] P. P. Bergmans, "A simple converse for broadcast channels with additive white Gaussian noise," IEEE Trans. Inform. Theory, vol. IT-20, pp. 279–280, Mar. 1974.

[52] R. Ahlswede and J. Körner, "Source coding with side information and a converse for degraded broadcast channels," IEEE Trans. Inform. Theory, vol. IT-21, pp. 629–637, Nov. 1975.

[53] J. Körner and K. Marton, "General broadcast channels with degraded message sets," IEEE Trans. Inform. Theory, vol. IT-23, pp. 60–64, Jan. 1977.

[54] A. El Gamal, "The capacity of a class of broadcast channels," IEEE Trans. Inform. Theory, vol. IT-25, pp. 166–169, Mar. 1979.

[55] K. Marton, "The capacity region of deterministic broadcast channels," in Trans. Int. Symp. Inform. Theory, Paris-Cachan, France, 1977.

[56] M. S. Pinsker, "Capacity of noiseless broadcast channels," Probl. Pered. Inform., vol. 14, no. 2, pp. 28–34, Apr.-Jun. 1978, translated in Probl. Inform. Transm., pp. 97-102, Apr.-June 1978.

[57] S. I. Gel'fand and M. S. Pinsker, "Capacity of a broadcast channel with one deterministic component," Probl. Pered. Inform., vol. 16, no. 1, pp. 24–34, Jan.-Mar. 1980, ; translated in Probl. Inform. Transm., vol. 16, no. 1, pp. 17-25, Jan.-Mar. 1980.

[58] A. El Gamal, "The capacity of the product and sum of two reversely degraded broadcast channels," Probl. Pered. Inform., vol. 16, pp. 3–23, Jan.-Mar. 1980.

[59] A. El Gamal and E. Van der Meulen, "A proof of Marton's coding theorem for the discrete memoryless broadcast channel," IEEE Trans. Inform. Theory, vol. IT-27, no. 1, pp. 120–122, Jan. 1981.

[60] E. C. Van der Meulen, "Random coding theorems for the general discrete memoryless broadcast channel," IEEE Trans. Inform. Theory, vol. IT-21, no. 2, pp. 180–190, Mar. 1975.

[61] H. Weingarten, Y. Steinberg, and S. Shamai(Shitz), "The capacity region of the Gaussian MIMO broadcast channel," IEEE Trans. Inform. Theory, vol. 52, pp. 3936–3964, September 2006.

[62] A. Gohari and V. Anantharam, "Evaluation of Marton's inner bound for the general broadcast channel," IEEE Trans. Inform. Theory, vol. 58, no. 2, pp. 608 –619, Feb. 2012.

[63] A. Padakandla and S. Pradhan, "Nested linear codes achieve Marton's inner bound for general broadcast channels," in 2011 IEEE ISIT Proceedings, 31 2011-aug. 5 2011, pp. 1554 –1558.

[64] T. Berger, Multiterminal Source Coding. In: The Information Theory Approach to Communications (ed. G. Longo), CISM Courses and Lecture Notes No. 229. Springer, Wien-New York, 1977.

[65] P. Gács and J. Körner, "Common information is far less than mutual information," Problems of Control and Information Theory, vol. 2, no. 2, pp. 119–162, 1972.

[66] I. Csiszár, "Linear codes for sources and source networks: Error exponents, universal coding," Information Theory, IEEE Transactions on, vol. 28, no. 4, pp. 585–592, 1982.

[67] A. Padakandla and S. S. Pradhan, "Computing sum of sources over an arbitrary multiple access channel," to appear in proceedings of 2013 IEEE International Symposium on Information Theory Proceedings (ISIT), 2013.

[68] A. Padakandla, A. Sahebi, and S. Pradhan, "A new achievable rate region for the 3-user discrete memoryless interference channel," in 2012 IEEE ISIT Proceedings, july 2012, pp. 2256 –2260.

[69] D. Krithivasan and S. Pradhan, "Lattices for distributed source coding: Jointly gaussian sources and reconstruction of a linear function," Information Theory, IEEE Transactions on, vol. 55, no. 12, pp. 5628 –5651, dec. 2009.

[70] E. Haim, Y. Kochman, and U. Erez, "Expurgation for discrete multiple-access channels via linear codes," in Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on, 2012, pp. 31–35.

[71] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: Lossy source-channel communication revisited," IEEE Trans. Inform. Theory, vol. 49, no. 3, pp. 1147–1158, May 2003.

[72] S. Pradhan and K. Ramchandran, "On functional duality in multiuser source and channel coding problems with one-sided collaboration," Information Theory, IEEE Transactions on, vol. 52, no. 7, pp. 2986–3002, 2006.

[73] A. Sahebi and S. Pradhan, "On distributed source coding using Abelian group codes," in Proceedings of 2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton), oct. 2012.

[74] 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. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0019995871907753

[75] ——, "Bounds on algebraic code capacities for noisy channels. II," Information and Control, vol. 19, no. 2, pp. 146 – 158, 1971. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0019995871907832

[76] A. Sahebi and S. Pradhan, "Abelian group codes for source coding and channel coding," submitted to IEEE Trans. of Information theory, April 2013, available at `http://arxiv.org/abs/1305.1598`.

[77] A. Padakandla and S. Pradhan, "Computing sum of sources over an arbitrary multiple access channel," available at `http://arxiv.org/abs/1301.5684`.

[78] H. G. Eggleston, Convexity. Cambridge: Cambridge University Press, 1958.