

An Achievable rate region for the 3–user interference channel based on coset codes

Arun Padakandla, Aria G. Sahebi and S. Sandeep Pradhan, *Member, IEEE*

Abstract

We consider the problem of communication over a three user discrete memoryless interference channel (3–IC). The current known coding techniques for communicating over an arbitrary 3–IC are based on message splitting, superposition coding and binning using independent and identically distributed (iid) random codebooks. In this work, we propose a new ensemble of codes - partitioned coset codes (PCC) - that possess an appropriate mix of empirical and algebraic closure properties. We develop coding techniques that exploit algebraic closure property of PCC to enable efficient communication over 3–IC. We analyze the performance of the proposed coding technique to derive an achievable rate region for the general discrete 3–IC. Additive and non-additive examples are identified for which the derived achievable rate region is the capacity, and moreover, strictly larger than current known largest achievable rate regions based on iid random codebooks.

I. INTRODUCTION

An interference channel (IC) is a model for communication between multiple transmitter receiver (Tx-Rx) pairs that share a common communication medium. Each transmitter wishes to communicate specific information to its corresponding receiver. 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 combating 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 [1] [2] [3], 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 [4], Carleial proposed the technique in which each receiver decodes a part of the signal sent by the interfering transmitter. To enable this, Carleial employed superposition coding [5] [6]. Each transmitter splits its 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.

The authors are with the Department of Electrical and Computer Engineering, University of Michigan, Ann Arbor 48109-2122, USA.

This work was supported by NSF grant CCF-1116021.

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 [7] 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. This coding technique is optimal under strong interference [4], [8]. El Gamal and Costa [9] prove that this technique is optimal for a class of deterministic ICs. Recently Etkin, Tse and Wang [10] prove that it is within 1 bit of the optimal for the Gaussian IC.

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. Whether the rate region proved achievable in [7], 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 article, we consider the problem of communicating over a three user interference channel (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* built over finite fields and Abelian groups for interference management. Coset codes are algebraically closed. The sum of any two codewords of a coset lies in another coset. Moreover, two cosets of a group or linear code, when added result in another coset of the same code. As against to adding two random codebooks whose codewords are statistically independent, we emphasize that the sum of two random cosets of a random group or 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 signals sent by the two interfering transmitters.

The use of lattice codes [11], [12], [13] and interference alignment techniques [14] have been proposed for efficient interference management over Gaussian ICs with three or more Tx-Rx pairs. Cadambe and Jafar [14] propose the technique of interference alignment for the Gaussian IC and thereby harness the available degrees of freedom in an IC with several Tx-Rx pairs more efficiently. Jafarian and Vishwanath [11] propose an achievable scheme for communicating over K -user Gaussian IC based on lattice codes and characterize its *efficiency*. Bresler, Parekh and Tse [12] employ lattice codes to align interference and thereby characterize the capacity of Gaussian ICs within a constant number of bits. While these works are restricted to *additive* ICs, 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 field and group 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 to 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 to 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-ICs 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 approach - use of statistically correlated codes possessing algebraic closure properties - it is natural to enquire the need for the same. Indeed, one can employ unstructured codes for communication over an arbitrary 3-IC and optimally stitch together all current known relevant coding techniques - message splitting, binning and superposition - 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 \mathcal{USB} -region, compare to the PCC rate region?

An important element of our findings is the strict sub-optimality of the \mathcal{USB} -technique for communicating over 3-ICs, 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 \mathcal{USB} -region in its corresponding capacity region. Moreover, for these 3-ICs 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 \mathcal{USB} -region for an arbitrary 3-IC.¹

In addition to efficiently decoding a bivariate function of the two interfering transmissions, which the proposed coding technique based on PCC accomplishes, it is necessary to enable receivers efficiently decode individual parts of interfering 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 or group multiplication - of transmissions of the two interfering transmitters. Such a technique is strictly sub-optimal for the purpose of decoding individual parts of interfering transmissions, when compared to the conventional technique based on unstructured independent codes. This leads us to enhance the PCC coding technique by incorporating the \mathcal{USB} -technique. This enables us to characterize a new achievable rate region for an arbitrary discrete 3-IC that contains PCC rate region and strictly enlarges the \mathcal{USB} -region.

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.² Example 3 illustrates the utility of this technique. We begin with preliminaries - notation, definitions and the precise statement of the problem - in section II. In section III, we provide a characterization of the CHK rate region for 2-IC. The first main finding of this article 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 II-B), and derive,

¹A little thought will convince an alert reader, that if 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.

²We note that the technique of interference alignment proposed by Cadambe and Jafar was studied in the context of Gaussian fading channels and achieve rates corresponding Gaussian input distributions.

in section III-B, an achievable rate region for the same, called \mathcal{USB} -region, that employs current known coding techniques based on unstructured codes. In section IV, we identify an additive 3-to-1 IC and propose a strategy based on correlated linear codes that is analytically proven to strictly outperform \mathcal{USB} -technique.

Our second main finding - a new achievable rate region for an arbitrary discrete 3-IC - is presented in three pedagogical steps. In the first step, we employ PCC to manage interference seen by only one receiver. In the second step, we employ PCC to manage interference seen by all three receivers. Finally, in the third step, we indicate how a coding technique that incorporates both \mathcal{USB} -region and PCC region can be developed for a general discrete 3-IC.

In this article, we develop coding techniques based on PCC built over finite fields and Abelian groups. Characterizing achievable rate regions for *arbitrary* 3-IC using statistically correlated codes endowed with algebraic properties, is a paradigm shift from the conventional techniques (based on unstructured codes) employed in information theory. The theory developed in this article contains several new elements. The rich structure of the finite field, and our fair understanding of coset codes³ provides us with the right setting to convey some of the new elements in it's simplest setting. Employing coset codes built over general Abelian groups involves, thanks to it's looser algebraic structure, a whole new set of tools and ideas. These being fairly group theoretic, we develop the same in a separate section. In section V, we develop PCC rate region by employing codes built over finite fields. Section VI contains our exposition in the context of Abelian groups, i.e., characterization of PCC rate region based on codes built over Abelian groups. While it is true that the PCC rate region using group codes contains PCC rate region based on finite fields, the above subdivision provides a pedagogical development of the involved techniques. Moreover, it also enables a reader unfamiliar with group theory to gather several of the key elements by a careful study of section V.

In sections V-A, VI-C, we present the first step that describes all the new elements of our framework in a simple setting. Here, we employ PCC built over finite fields, Abelian groups respectively, to manage interference seen by only one receiver. For this step, we furnish a complete and elaborate proof of achievability. In the second step, presented in section V-B, we employ PCC to manage interference seen by all three receivers. Finally, in section V-C, we incorporate unstructured codes via \mathcal{USB} -technique and PCC to derive an achievable rate region that is strictly larger than \mathcal{USB} -region.

Our third main finding is the identification of 3-IC's for which the proposed framework outperforms all current known coding techniques. In particular, we identify in section V-A, a non-additive 3-to-1 IC (Example 3) for which \mathcal{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 article - codes endowed with algebraic closure properties enable efficient communication over arbitrary 3-IC's not just additive, symmetric instances - and thereby justifies the framework developed herein. We strengthen this claim by identifying in section VI-C 3-IC's for which PCC codes

³While there has been several works that study performance of particular coset codes, a systematic information theoretic study of it's performance over an *arbitrary* instance of a multi-terminal channel has not been undertaken.

built over Abelian groups achieve capacity, while all known coding techniques, including PCC codes built over finite fields are sub-optimal. This illustrates the need to develop a fundamental theory of codes built over the various algebraic objects.

II. PRELIMINARIES: NOTATION AND DEFINITIONS

A. Notation

- We let \mathbb{N}, \mathbb{R} denote the set of natural numbers and real numbers respectively. Calligraphic letters such as \mathcal{X}, \mathcal{Y} exclusively to denote finite sets.
- For $K \in \mathbb{N}$, we let $[K] := \{1, 2, \dots, K\}$.
- In this article, we will need to define multiple objects, mostly triples, 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 $y_j \in \mathcal{Y}_j : j \in [3]$, we let $\underline{y} \in \underline{\mathcal{Y}}$ abbreviate $(y_1, y_2, y_3) \in \underline{\mathcal{Y}}$, (iii) if $d_j : \mathcal{Y}_j^n \rightarrow \mathcal{M}_j : j \in [3]$ denote (decoding) maps, then we let $\underline{d}(\underline{y}^n)$ denote $(d_1(y_1^n), d_2(y_2^n), d_3(y_3^n))$.
- If $j \in \{1, 2\}$, then $\dot{j} \in \{1, 2\} \setminus \{j\}$ is the other index.
- Unless otherwise mentioned, we let θ denote an integral power of a prime. Throughout, \mathcal{F}_θ will denote the finite field of cardinality θ .
- We employ the notion of typicality as in [15]. In particular, if U, V are random variables distributed with respect to p_{UV} , then $T_\eta(U, V) \in \mathcal{U}^n \times \mathcal{V}^n$ denotes the typical set with respect to p_{UV} and deviation parameter η . For any $v^n \in \mathcal{V}^n$, $T_\eta(U|v^n) = \{u^n : (u^n, v^n) \in T_\eta(U, V)\}$ denotes the conditional typical set.

B. Definitions: 3-IC, 3-to-1IC, 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 $\underline{X}_1, \dots, \underline{X}_{t-1}$, past outputs $\underline{Y}_1, \dots, \underline{Y}_{t-1}$ and (iii) used without feedback, i.e., encoders have no information of the symbols received by decoders. Let $W_{\underline{Y}|\underline{X}}(\underline{y}|\underline{x}) = W_{Y_1 Y_2 Y_3 | X_1 X_2 X_3}(y_1, y_2, y_3 | x_1, x_2, x_3)$ 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 bounded cost functions $\kappa_j : \mathcal{X}_j \rightarrow [0, \infty) : j \in [3]$. 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 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 $e_j : \mathcal{M}_j \rightarrow \mathcal{X}_j^n : j \in [3]$, and (iii) three decoder maps $d_j : \mathcal{Y}_j^n \rightarrow \mathcal{M}_j : j \in [3]$.

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

Definition 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) $\frac{\log |\underline{\mathcal{M}}_j^{(n)}|}{n} \geq R_j - \eta : j \in [3]$, (ii) $\bar{\xi}(\underline{e}^{(n)}, \underline{d}^{(n)}) \leq \eta$, and (iii) average cost $\underline{\tau}(\underline{e}^{(n)})_j \leq \tau_j + \eta$. The capacity region is $\mathbb{C}(\underline{\tau}) := \{\underline{R} \in \mathbb{R}^3 : (\underline{R}, \underline{\tau}) \text{ is achievable}\}$.

We now consider 3-to-1 IC, a class of 3-IC's that was studied in [16]. 3-to-1 IC enables us to prove strict sub-optimality of coding techniques based on unstructured codes. A 3-to-1 IC is a 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) $W_{Y_2 | \underline{\mathcal{X}}}(y_2 | \underline{x}) := \sum_{(y_1, y_3) \in \mathcal{Y}_1 \times \mathcal{Y}_3} W_{\underline{Y} | \underline{X}}(y | \underline{x})$ is independent of $(x_1, x_3) \in \mathcal{X}_1 \times \mathcal{X}_3$, and (ii) $W_{Y_3 | \underline{\mathcal{X}}}(y_3 | \underline{x}) := \sum_{(y_1, y_2) \in \mathcal{Y}_1 \times \mathcal{Y}_2} W_{\underline{Y} | \underline{X}}(y | \underline{x})$ 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

$$W_{\underline{Y} | \underline{X}}(y | \underline{x}) = W_{Y_1 | X_1}(y_1 | x_1) W_{Y_2 | X_2}(y_2 | x_2) W_{Y_3 | X_3}(y_3 | x_3)$$

for some conditional pmfs $W_{Y_1 | X_1}$, $W_{Y_2 | X_2}$ and $W_{Y_3 | X_3}$. We also note that $X_1 X_3 - X_2 - Y_2$ and $X_1 X_2 - X_3 - Y_3$ are Markov chains for any distribution $p_{X_1} p_{X_2} p_{X_3} W_{\underline{Y} | \underline{X}}$.⁴

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 \mathcal{USB} -region for 3-to-1 IC.

III. MESSAGE SPLITTING AND SUPERPOSITION USING UNSTRUCTURED CODES

A. CHK-technique for 2-IC

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 its satellite codebook.

⁴Any 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.

A standard information theoretic analysis of probability of error yields an achievable rate region referred to herein as CHK rate region for 2-IC.

Definition 4 and theorem 1 provide a characterization of rate pairs achievable using CHK-technique. We omit restating the definitions analogous to definitions 1, 2, 3 for a 2-IC.

Definition 4: Let $\mathbb{D}_{HK}(\mathcal{T})$ 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) $p_{Y|XW} = W_{Y|X}$, (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 $p_{QWXY} \in \mathbb{D}_{HK}(\mathcal{T})$, let $\alpha_{HK}(p_{QWXY})$ denote the set of rate pairs $(R_1, R_2) \in [0, \infty]^2$ that satisfy

$$\begin{aligned} R_j &< \min \{I(X_j; Y_j | QW_{\check{j}}), I(X_j; Y_j | Q\overline{W}) + I(W_j X_{\check{j}}; Y_{\check{j}} | QW_{\check{j}})\} : j \in [2] \\ R_1 + R_2 &< \min \left\{ I(X_j; Y_j | Q\overline{W}) + I(W_j X_{\check{j}}; Y_{\check{j}} | Q) : j \in [2], \sum_{j=1}^2 I(W_j X_{\check{j}}; Y_{\check{j}} | QW_{\check{j}}) \right\} \\ 2R_j + R_{\check{j}} &< I(X_j; Y_j | Q\overline{W}) + I(W_j X_{\check{j}}; Y_{\check{j}} | QW_{\check{j}}) + I(W_{\check{j}} X_j; Y_j | Q) : j \in [2] \end{aligned}$$

and

$$\alpha_{HK}(\mathcal{T}) = \text{cl} \left(\bigcup_{\substack{p_{QWXY} \in \\ \mathbb{D}_{HK}(\mathcal{T})}} \alpha_{HK}(p_{QWXY}) \right).$$

Theorem 1: For 2-IC $(\mathcal{X}, \mathcal{Y}, W_{Y|X}, \kappa)$, $\alpha_{HK}(\mathcal{T})$ is achievable, i.e., $\alpha_{HK}(\mathcal{T}) \subseteq \mathbb{C}(\mathcal{T})$.

Remark 1: Recently, several efforts [17], [18], [19] have yielded simplified descriptions [20] of $\alpha_{HK}(\mathcal{T})$. The description stated above involving fewer auxiliary random variables and tighter bounds on their cardinalities is due to Chong et. al. [17].

B. 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 - message splitting, superposition coding and precoding via binning - 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 \mathcal{USB} -technique.

One can characterize \mathcal{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 \mathcal{USB} -technique, it suffices to obtain a characterization of \mathcal{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 its message. This can be proved using the 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 with 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 point-to-point (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. 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 its 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 its semi-private part and (ii) $M_{2X} \in \mathcal{M}_{2X}$ of rate T_2 denote its 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 $(x_2^n(m_{21}, m_{2X}) : m_{2X} \in \mathcal{M}_{2X})$ is built over \mathcal{X}_2 . The codebooks over \mathcal{X}_2 form the private layer. The codeword $x_2^n(M_{21}, M_{2X})$ corresponding to message $M_2 = (M_{21}, M_{2X})$ is input on the channel.

Decoders 2 and 3 enjoying interference free reception perform simple PTP 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 $(u_2^n(\hat{m}_{21}), u_3^n(\hat{m}_{31}))$ such that $(x_1^n(\hat{m}_1), u_2^n(\hat{m}_{21}), u_3^n(\hat{m}_{31}), Y_1^n)$ 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 \mathcal{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 5 and theorem 2. 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 the time sharing sequence q^n , the three collections of codebooks, one corresponding to each user,⁵ are mutually independent. Let $p_{QP_{X_1|Q}P_{U_2X_2|Q}P_{U_3X_3|Q}W_{Y|X}}$ be a pmf on $\mathcal{Q} \times \mathcal{U}_2 \times \mathcal{U}_3 \times \mathcal{X} \times \mathcal{Y}$. 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

⁵Here, the collection of user j 's codebooks refers to the entire collection of codebooks employed by encoder j .

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 $(x_2^n(m_{21}, m_{2X}) : m_{2X} \in \mathcal{M}_{2X})$ in the private codebook corresponding to semi-private message m_{21}^U are independent and identically distributed according to $\prod_{t=1}^n p_{X_2|U_2Q}(\cdot|(u_2^n(m_{21}^U))_t, q_t)$. 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

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

The probability of decoder 1 decoding erroneously decays exponentially if

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

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 5: Let $\mathbb{D}_u(\mathcal{T})$ denote the collection of pmfs $p_{QU_2U_3XY}$ defined on $\mathcal{Q} \times \mathcal{U}_2 \times \mathcal{U}_3 \times \mathcal{X} \times \mathcal{Y}$, where $\mathcal{Q}, \mathcal{U}_2, \mathcal{U}_3$ are finite sets, such that (i) $p_{Y|XU_2U_3Q} = W_{Y|X}$, (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 $p_{QU_2U_3XY} \in \mathbb{D}_u(\mathcal{T})$, let $\alpha_u(p_{QU_2U_3XY})$ denote the set of rate triples $(R_1, R_2, R_3) \in [0, \infty)^3$ that satisfy

$$0 \leq R_1 < I(X_1; Y_1 | Q, U_2, U_3), \quad 0 \leq R_j < I(U_j X_j; Y_j | Q) : j = 2, 3 \quad (1)$$

$$R_1 + R_2 < I(U_2 X_1; Y_1 | QU_3) + I(X_2; Y_2 | QU_2), \quad R_1 + R_3 < I(U_3 X_1; Y_1 | QU_2) + I(X_3; Y_3 | QU_3)$$

$$R_1 + R_2 + R_3 < I(U_2 U_3 X_1; Y_1 | Q) + I(X_2; Y_2 | QU_2) + I(X_3; Y_3 | QU_3), \quad (2)$$

and

$$\alpha_u(\mathcal{T}) = \text{cl} \left(\bigcup_{p_{QU_2U_3XY} \in \mathbb{D}_u(\mathcal{T})} \alpha_u(p_{QU_2U_3XY}) \right).$$

Theorem 2: For 3-to-1 IC $(\mathcal{X}, \mathcal{Y}, W_{Y|X}, \kappa)$, $\alpha_u(\mathcal{T})$ is achievable, i.e., $\alpha_u(\mathcal{T}) \subseteq \mathcal{C}(\mathcal{T})$.

The reader will also recognize that $\alpha_u(\mathcal{T})$ is also achievable over an arbitrary 3-IC.⁶ This is stated below.

Theorem 3: For 3-IC $(\mathcal{X}, \mathcal{Y}, W_{Y|X}, \kappa)$, $\alpha_u(\mathcal{T})$ is achievable, i.e., $\alpha_u(\mathcal{T}) \subseteq \mathcal{C}(\mathcal{T})$.

IV. STRICT SUB-OPTIMALITY OF \mathcal{USB} -REGION FOR 3-TO-1 IC

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

⁶Unless the 3-IC $(\mathcal{X}, \mathcal{Y}, W_{Y|X}, \kappa)$ is a 3-to-1IC, $\alpha_u(\mathcal{T})$ is *not* its \mathcal{USB} -region.

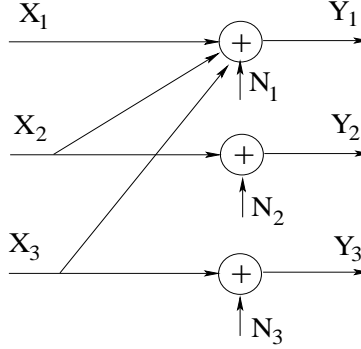


Fig. 1. A binary additive 3-to-1 IC described in example 1.

Example 1: Consider a binary additive 3-to-1 IC illustrated in figure 1 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

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

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 1, 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 respective capacities $1 - h_b(\delta)$. Constrained to average Hamming weight of τ , user 1 cannot hope to achieve a rate larger than $h_b(\tau * \delta_1) - h_b(\delta_1)$.⁷ 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, \mathcal{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 $h_b(\tau * \delta_1) - h_b(\delta_1)$.⁸ This claim and the strict sub-optimality of \mathcal{USB} -technique is stated in theorem 4.

⁷If 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 .

⁸An informed reader will be able to reason this by relating this situation to a PTP channel with partial state observed at the receiver.

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 the 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 $\tau * \delta_1 \leq \delta$.

In theorem 4, 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 1 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 \mathcal{WSB} -technique is strictly suboptimal for the 3-to-1 IC presented in example 1.

Theorem 4: Consider the 3-to-1 IC described in example 1. If $\tau * \delta_1 \leq \min\{\delta_2, \delta_3\}$, then $\mathbb{C}(\tau) = \beta(\tau, \frac{1}{2}, \frac{1}{2}, \delta)$, where $\beta(\underline{\tau}, \underline{\delta})$ is given by (3). If $h_b(\delta_2) + h_b(\delta_3) < 1 + h_b(\tau * \delta_1)$, then $(h_b(\tau * \delta_1) - h_b(\delta_1), 1 - h_b(\delta_2), 1 - h_b(\delta_3)) \notin \alpha_u(\tau, 0, 0)$.

Proof: We only need to prove the second statement. If $H(X_j|Q, U_j) = 0$ for $j = 2, 3$, then the upper bound in (2) reduces to $R_1 + R_2 + R_3 \leq I(X_2 X_3 X_1; Y_1|Q) \leq 1 - h_b(\delta_1)$. 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, 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

$$0 < H(X_j|Q, U_j) = H(X_j|X_j, Q, U_2, U_3) = H(X_2 \oplus X_3|X_j, Q, U_2, U_3) \leq H(X_2 \oplus X_3|Q, U_2 U_3).$$

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 $H(X_2 \oplus X_3|(Q, U_2 U_3) = \tilde{q}^*) > 0$. Under this condition, we prove that the upper bound (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 1: 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 $f : [0, \frac{1}{2}] \rightarrow [0, 1]$ 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 may be 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 (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

$$I(X_1; Y_1 | \tilde{Q}) = \sum_{\tilde{q}} p_{\tilde{Q}}(\tilde{q}) H(X_1 \oplus N_1 \oplus X_2 \oplus X_3 | \tilde{Q} = \tilde{q}) - \sum_{x_1, \tilde{q}} p_{X_1 \tilde{Q}}(x_1, \tilde{q}) H(N_1 \oplus X_2 \oplus X_3 | \tilde{Q} = \tilde{q}) \quad (4)$$

$$= \sum_{\tilde{q}} p_{\tilde{Q}}(\tilde{q}) H(X_1 \oplus N_1 \oplus X_2 \oplus X_3 | \tilde{Q} = \tilde{q}) - \sum_{\tilde{q}} p_{\tilde{Q}}(\tilde{q}) H(N_1 \oplus X_2 \oplus X_3 | \tilde{Q} = \tilde{q})$$

$$< \sum_{\tilde{q}} p_{\tilde{Q}}(\tilde{q}) H(X_1 \oplus N_1 | \tilde{Q} = \tilde{q}) - \sum_{\tilde{q}} p_{\tilde{Q}}(\tilde{q}) H(N_1 | \tilde{Q} = \tilde{q}) = \sum_q p_Q(q) H(X_1 \oplus N_1 | Q = q) - h_b(\delta_1) \quad (5)$$

$$= \sum_q p_Q(q) h_b(p_{X_1|Q}(1|q) * \delta_1) - h_b(\delta_1) \leq h_b(\mathbb{E}_Q \{p_{X_1|Q}(1|q) * \delta_1\}) - h_b(\delta_1) \leq h_b(\tau * \delta_1) - h_b(\delta_1), \quad (6)$$

where (i) (4) follows from independence of (N_1, X_2, X_3) and X_1 conditioned on realization of Q , (ii) (5) follows from the 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 1, and noting that $p_{X_1 \oplus N_1 | \tilde{Q}}(1 | \tilde{q}^*) > p_{N_1 | \tilde{Q}}(1 | q^*)$, (iii) the first inequality in (6) follows from Jensen's inequality and the second follows from the cost constraint that any test channel in $\mathbb{D}_u(\tau, 0, 0)$ must satisfy. ■

Corollary 1: Consider the 3-to-1 IC in example 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 $(h_b(\tau * \delta_1) - h_b(\delta_1), 1 - h_b(\delta), 1 - h_b(\delta)) \notin \alpha_u(\tau, 0, 0)$ but $(h_b(\tau * \delta_1) - h_b(\delta_1), 1 - h_b(\delta), 1 - h_b(\delta)) \in \mathbb{C}(\tau)$ and thus $\alpha_u(\tau, 0, 0) \neq \mathbb{C}(\tau)$. In particular, if $\delta_1 = 0.01$ and $\delta_2 \in (0.1325, 0.21)$, then $\alpha_u(\frac{1}{8}, 0, 0) \neq \mathbb{C}(\frac{1}{8})$.

Comparison with Bandemer and El Gamal [21] :- We refer the reader to [21, 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 1 with a notation consistent with that employed in [21].

Example 2: Consider a binary additive 3-to-1 IC illustrated in figure 1 with $\mathcal{X}_j = \mathcal{Z}_j = \{0, 1\} : j \in [3]$ with channel transition probabilities $W_{\underline{Z}|\underline{X}}(\underline{z}|\underline{x}) = BSC_{\delta_1}(z_1|x_1 \oplus x_2 \oplus x_3)BSC_{\delta_2}(z_2|x_2)BSC_{\delta_3}(z_3|x_3)$. 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 [21]. It may be 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 arguments⁹. We now employ the bound

$$R_1 + \min \{R_2 + H(X_{31}|Q), R_3 + H(X_{21}|Q), R_2 + R_3, H(S_1|Q)\} \leq I(X_1, S_1; Z_1|Q) \quad (7)$$

present in the description of $\mathfrak{R}_1(Q, X_1, X_2, X_3)$. Clearly, the right hand side of (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.

V. ACHIEVABLE RATE REGION USING PCC BUILT OVER FINITE FIELDS

In this section we present our second main finding - a new achievable rate region for 3-IC - in the context of finite fields. In other words, we propose a coding technique based on PCC built over finite fields. Characterizing its information-theoretic performance enables us to derive an achievable rate region, henceforth referred to as PCC-region.¹⁰ We derive PCC rate region in three pedagogical steps. In the first step, presented in section V-A, 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 proof of achievability. In this section, we also identify a *non-additive* 3-to-1 IC (Example 3) for which we *analytically* prove (i) strict sub-optimality of \mathcal{WSB} -technique and (ii) optimality of PCC rate region. This example indeed illustrates the central theme of this article - 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 V-B, we employ PCC to manage interference seen by every receiver and thereby provide a characterization of PCC rate region. In the third step we provide a unification of PCC rate region and \mathcal{WSB} -rate region along the lines of [22, Section VI].

A. Step I : Managing interference seen by one receiver using PCC built over fields

The linear coding technique proposed for example 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.

⁹This 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 is trivial for these j .

¹⁰We employ the same terminology for the rate region achievable using PCC built over Abelian groups in section VI.

Definition 6: Let $\mathbb{D}_f(\mathcal{T})$ denote the collection of distributions $p_{QU_2U_3XY} \in \mathbb{D}_u(\mathcal{T})$ defined over $\mathcal{Q} \times \mathcal{U}_2 \times \mathcal{U}_3 \times \mathcal{X} \times \mathcal{Y}$, where $\mathcal{U}_2 = \mathcal{U}_3$ is a finite field. For $p_{QU_2U_3XY} \in \mathbb{D}_f(\mathcal{T})$, let $\alpha_f^{3-1}(p_{QU_2U_3XY})$ 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

$$\alpha_f^{3-1}(\mathcal{T}) = \text{cocl} \left(\bigcup_{\substack{p_{QU_2U_3XY} \in \\ \mathbb{D}_f(\mathcal{T})}} \alpha_f^{3-1}(p_{QU_2U_3XY}) \right).$$

Theorem 5: For 3-IC $(\mathcal{X}, \mathcal{Y}, W_{Y|X}, \kappa)$, $\alpha_f^{3-1}(\mathcal{T})$ is achievable, i.e., $\alpha_f^{3-1}(\mathcal{T}) \subseteq \mathbb{C}(\mathcal{T})$.

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 $p_{QU_2U_3XY} \in \mathbb{D}_f(\mathcal{T})$ with $\mathcal{Q} = \phi^{11}$ and $\mathcal{U}_2 = \mathcal{U}_3 = \mathcal{F}_\theta$. 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 a generic index $j \in \{2, 3\}$. As in section III-B, 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) $M_{j1} \in \mathcal{M}_{j1} := [\theta^{t_j}]$ 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 III-B 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 a coset $\lambda_j \subseteq \mathcal{U}_j^n$ of a linear code $\bar{\lambda}_j \subseteq \mathcal{U}_j^n$ denote user j 's cloud center codebook. In particular, let $g_j \in \mathcal{U}_j^{s_j \times n}$ denote generator matrix of $\bar{\lambda}_j$ and coset λ_j correspond to shift $b_j^n \in \mathcal{U}_j^n$. We let the cloud center codebooks of users' 2 and 3 overlap, i.e., the larger of $\bar{\lambda}_2, \bar{\lambda}_3$ contains the other. For example, if $s_{j2} \leq s_{j3}$, then $\bar{\lambda}_{j2} \subseteq \bar{\lambda}_{j3}$. We therefore let $g_{j3}^T = \begin{bmatrix} g_{j2}^T & g_{j3/j2}^T \end{bmatrix}$.

Since codewords of a uniformly distributed coset code are uniformly distributed, we need to partition the coset code λ_j into θ^{t_j} bins to induce a non-uniform distribution over the auxiliary alphabet \mathcal{U}_j . In particular, for each codeword $u_j^n(a^{s_j}) := a^{s_j} g_j \oplus b_j^n$, where $a^{s_j} \in \mathcal{U}_j^{s_j}$, a binning function $i_j(a^{s_j}) \in [\theta^{t_j}]$ 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 and we formalize the same through the following definition.

Definition 7: A coset code λ is completely specified by the generator matrix $g \in \mathcal{F}_\theta^{k \times n}$ and a bias vector $b_j^n \in \mathcal{F}_\theta^n$. Consider a partition of λ into θ^l bins. Each codeword $a^k g \oplus b^n$ is assigned an index $i(a^k) \in [\theta^l]$. This

¹¹Since the time sharing random variable Q is employed in a standard way, we choose to omit the same in this description.

coset code λ with its 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 $m \in [\theta^l]$, let $c(m) := \{a^k \in \mathcal{F}_\theta^k : i(a^k) = m\}$.

User j 'th satellite codebook \mathcal{C}_j , 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 $(u_j^n(a^{sj}), x_j^n(M_{jX}, b_{jX}))$ of jointly typical codewords with $(a^{sj}, b_{jX}) \in c_{j1}(M_{j1}) \times c_{jX}$. 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 a generic index $j \in \{2, 3\}$. Decoder j identifies all $(\hat{m}_{j1}, \hat{m}_{jX})$ for which there exists $(a^{sj}, b_{jX}) \in c_{j1}(\hat{m}_{j1}) \times c_{jX}$ such that $(u_j^n(a^{sj}), x_j^n(\hat{m}_{jX}, b_{jX}), Y_j^n)$ is jointly typical with respect to $p_{U_j X_j Y_j}^n$. If there is exactly one such pair $(\hat{m}_{j1}, \hat{m}_{jX})$, this is declared as the message of user j . Otherwise an error is signaled.

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

We characterize the performance of the proposed coding technique in the following proof by averaging over the ensemble of codebooks. Since the distribution induced on the codebooks is such that codebooks of users 2 and 3 are statistically correlated and moreover, contain correlated codewords, this involves new elements.

Proof: Let $p_{QU_2U_3XY} \in \mathbb{D}_f(\mathcal{T})$, $\underline{R} \in \alpha_f^{3-1}(p_{QU_2U_3XY})$ and $\tilde{\eta} > 0$. Let us assume $\mathcal{U}_2 = \mathcal{U}_3 = \mathcal{F}_\theta$ is the finite field of size θ . For each $n \in \mathbb{N}$ sufficiently large, we prove existence of a 3-IC code $(n, \underline{\mathcal{M}}, \underline{e}, \underline{d})$ for which $\frac{\log \mathcal{M}_k}{n} \geq R_k - \tilde{\eta}$, $\tau_k(e_k) \leq \tau_k + \tilde{\eta}$ for $k \in [3]$ and $\bar{\xi}(\underline{e}, \underline{d}) \leq \tilde{\eta}$.

Taking a cue from the above coding technique, we begin with an alternative characterization of $\alpha_f^{3-1}(p_{QU_2U_3XY})$ in terms of the parameters of the code.

Definition 8: Consider $p_{QU_2U_3XY} \in \mathbb{D}_f(\mathcal{T})$ and let $\mathcal{F}_\theta := \mathcal{U}_2 = \mathcal{U}_3$. Let $\tilde{\alpha}_f^{3-1}(p_{QU_2U_3XY})$ be defined as the set of rate triples $(R_1, R_2, R_3) \in [0, \infty)^3$ for which $\bigcup_{\delta > 0} \mathcal{S}(\underline{R}, p_{QU_2U_3XY}, \delta)$ is non-empty, where $\mathcal{S}(\underline{R}, p_{QU_2U_3XY}, \delta)$ 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

$$R_j = T_j + L_j, \quad K_j > \delta, \quad (S_j - T_j) > \log \theta - H(U_j|Q) + \delta,$$

$$(S_j - T_j) + K_j > \log \theta + H(X_j|Q) - H(U_j, X_j|Q) + \delta$$

$$T_j > \delta, \quad L_j > \delta, \quad K_j + L_j < I(X_j; Y_j, U_j|Q) - \delta, \quad S_j < \log \theta - H(U_j|X_j, Y_j, Q) - \delta,$$

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

$$R_1 + S_j < \log \theta + H(X_1|Q) - H(X_1, U_2 \oplus U_3|Y_1, Q) - \delta$$

for $j = 2, 3$.

Lemma 2: $\tilde{\alpha}_f^{3-1}(p_{QU_2U_3XY}) = \alpha_f^{3-1}(p_{QU_2U_3XY})$.

Proof: The proof follows by substituting $R_j = T_j + L_j$ in the bounds characterizing $\mathcal{S}(\underline{R}, p_{QU_2U_3XY})$ 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 $\alpha_f^{3-1}(p_{QU_2U_3XY})$. The presence of strict inequalities in the bounds characterizing $\alpha_f^{3-1}(p_{QU_2U_3XY})$ and $\mathcal{S}(\underline{R}, p_{QU_2U_3XY}, \delta)$ enables one to prove $\bigcup_{\delta>0} \mathcal{S}(\underline{R}, p_{QU_2U_3XY}, \delta)$ is non-empty for every $\underline{R} \in \alpha_f^{3-1}(p_{QU_2U_3XY})$. \blacksquare Lemma 7 provides us with $\delta > 0$ and parameters $(S_j, T_j, K_j, L_j, : j = 2, 3) \in \mathcal{S}(\underline{R}, p_{QU_2U_3XY}, \delta)$ of the code whose existence we seek to prove. Define $\eta = \frac{1}{2d} \min\{\delta, \tilde{\eta}\}$, where $d \in \mathbb{N}$ will be specified in due course. Let $q^n \in T_\eta(Q)$ denote the time sharing sequence. User 1's code contains $\exp\{nR_1\}$ codewords $(x_1^n(m_1) \in \mathcal{X}_1^n : m_1 \in \mathcal{M}_1)$, where $\mathcal{M}_1 := [\exp\{nR_1\}]$. For $j \in \{2, 3\}$, user j 'th cloud center codebook λ_j is the PCC $(n, s_j, t_j, g_j, b_j^n, i_j)$ built over $\mathcal{U}_j^n = \mathcal{F}_\theta^n$ where $s_j := \frac{nS_j}{\log \theta}$ and $t_j := \frac{nT_j}{\log \theta}$. We refer the reader to the coding technique described prior to the proof for the definitions of $u_j^n(a^{s_j})$ and $c_{j1}(m_{j1})$. The PCCs *overlap*, and without loss of generality, we assume $s_2 \leq s_3$ and therefore $g_3^T = [g_2^T \ g_{3/2}^T]$.

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

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

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 $\lambda_2 \oplus \lambda_3 := \{u_2^n \oplus u_3^n : u_j^n \in \lambda_j, j = 2, 3\}$ of the cloud center codebooks. Let $u_\oplus^n(a^{s_3}) := a^{s_3}g_3 \oplus b_2^n \oplus b_3^n$ denote a generic codeword in $\lambda_2 \oplus \lambda_3$. Note that $\lambda_2 \oplus \lambda_3 = \{u_\oplus^n(a^{s_3}) : a^{s_3} \in \mathcal{U}_3^{s_3}\}$.¹² Having received Y_1^n , it looks for all potential message \hat{m}_1 for which there exists a $a^{s_3} \in \mathcal{U}_3^{s_3}$ such that $(q^n, u_\oplus^n(a^{s_3}), x_1^n(\hat{m}_1), Y_1^n) \in T_{2\eta_1}(Q, U_2 \oplus U_3, X_1, Y_1)$.¹³ 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 $(q^n, u_j^n(a^{s_j}), x_j^n(\hat{m}_{jX}, b_{jX}), Y_j^n) \in T_{2\eta_1}(Q, U_j, X_j, Y_j)$, where Y_j^n is the received vector. If there is exactly one such pair $(\hat{m}_{j1}, \hat{m}_{jX})$, this is declared as message of user j . 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 $(n, \underline{\mathcal{M}}, \underline{e}, \underline{d})$ of rate $\frac{\log |\mathcal{M}_1|}{n} = R_1, \frac{\log |\mathcal{M}_j|}{n} = \frac{t_j}{n} \log \theta + L_j = T_j + L_j = R_j : j \in \{2, 3\}$, 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 $\mathcal{C}_1 := (X_1^n(m_1) : m_1 \in \mathcal{M}_1)$, generator matrices $G_2, G_{3/2}$,¹⁴ bias vectors B_1^n, B_2^n , bin indices $(I_j(a^{s_j}) : a^{s_j} \in \mathcal{U}_j^{s_j}) : j = 2, 3$ and codewords of $\mathcal{C}_j = (X_j^n(m_{jX}, b_{jX}) : (m_{jX}, b_{jX}) \in \mathcal{M}_{jX} \times c_{jX}) : j = 2, 3$ be mutually independent, (ii) the codewords of $\mathcal{C}_j : j = 1, 2, 3$ are identically distributed according to $\prod_{t=1}^n p_{X_j|Q}(\cdot|q_t)$, (iii) generator matrices $G_{j_1}, G_{j_2/j_1}$, bias vectors B_1^n, B_2^n , bin indices

¹²Here we have used the assumption $s_2 \leq s_3$. In general, if $s_{j_1} \leq s_{j_2}$, we have $\lambda_2 \oplus \lambda_3 = \{u_\oplus^n(a^{s_{j_2}}) : a^{s_{j_2}} \in \mathcal{U}_{j_2}^{s_{j_2}}\}$, where $u_\oplus^n(a^{s_{j_2}}) := a^{s_{j_2}}g_{j_2} \oplus b_2^n \oplus b_3^n$ denotes a generic codeword.

¹³The choice for η_1 is indicated at the end of the proof.

¹⁴Recall, that we have assumed $s_2 \leq s_3$.

$(I_j(a^{s_j}) : a^{s_j} \in \mathcal{U}_j^{s_j}) : j = 2, 3$ be uniformly distributed over their respective range spaces. We denote the random partitioned coset code $(n, s_j, t_j, G_j, B_j^n, I_j)$ of user j as Λ_j and let (i) $U_j^n(a^{s_j}) := a^{s_j} G_j \oplus B_j^n$ denote a generic random codeword in Λ_j , (ii) $U_{\oplus}^n(a^{s_3}) := a^{s_3} G_3 \oplus B_2^n \oplus B_3^n$ denote a generic codeword in $\Lambda_2 \oplus \Lambda_3$, and (iii) $C_{j1}(M_{j1}) = \{a^{s_j} \in \mathcal{U}_j^{s_j} : I_j(a^{s_j}) = M_{j1}\}$ 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

$$\begin{aligned} \epsilon_{11} &:= \{(q^n, X_1^n(M_1)) \notin T_{2\eta}(Q, X_1)\} \\ \epsilon_{1j} &:= \bigcap_{\substack{(a^{s_j}, b_{jX}) \in \\ C_{j1}(M_{j1}) \times c_{jX}}} \{(q^n, U_j^n(a^{s_j}), X_j^n(M_{jX}, b_{jX})) \notin T_{2\eta}(Q, U_j, X_j)\}, \text{ for } j = 2, 3 \\ \epsilon_2 &:= \{(q^n, U_2^n(A^{s_2}), U_3^n(A^{s_3}), X_1^n(M_1), X_2^n(M_{2X}, B_{2X}), X_3^n(M_{3X}, B_{3X})) \notin T_{\eta_1}(Q, U_2, U_3, \underline{X})\} \quad (8) \\ \epsilon_3 &:= \{(q^n, U_2^n(A^{s_2}), U_3^n(A^{s_3}), X_1^n(M_1), X_2^n(M_{2X}, B_{2X}), X_3^n(M_{3X}, B_{3X}), \underline{Y}^n) \notin T_{2\eta_1}(Q, X_1, U_2, U_3, \underline{X}, \underline{Y})\} \\ \epsilon_{41} &:= \bigcup_{\hat{m}_1 \neq M_1} \bigcup_{a^{s_3} \in \mathcal{U}_3^{s_3}} \{(q^n, U_{\oplus}^n(a^{s_3}), X_1^n(\hat{m}_1), Y_1^n) \in T_{2\eta_1}(Q, U_2 \oplus U_3, X_1, Y_1)\} \\ \epsilon_{4j} &:= \bigcup_{\hat{m}_j \neq M_j} \bigcup_{\substack{a^{s_j} \in \\ C_{j1}(\hat{m}_{j1})}} \bigcup_{b_{jX} \in c_{jX}} \{(q^n, U_j^n(a^{s_j}), X_j^n(\hat{m}_{jX}, b_{jX}), Y_j^n) \in T_{2\eta_1}(Q, U_j, V_j, Y_j)\} \text{ for } j = 2, 3. \end{aligned}$$

Note that $\epsilon := \bigcup_{j=1}^3 (\epsilon_{1j} \cup \epsilon_2 \cup \epsilon_3 \cup \epsilon_{4j})$ 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

$\epsilon_l := \epsilon_{l2} \cup \epsilon_{l3}$, where

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

$\mathcal{L}_j(n)$ is half of the expected number of jointly typical pairs in the indexed pair of bins.¹⁵ For sufficiently large n , we prove $\mathcal{L}_j(n) > 2$. For such an n , $\epsilon_{1j} \subseteq \epsilon_{lj} : j = 2, 3$. Since, we can choose n sufficiently large, we will henceforth assume $\epsilon_{1j} \subseteq \epsilon_{lj} : j = 2, 3$. It therefore suffices to derive upper bounds on $P(\epsilon_{11}), P(\epsilon_{lj}) : j = 2, 3, P(\tilde{\epsilon}_1 \cap \epsilon_2), P((\tilde{\epsilon}_1 \cup \epsilon_2)^c \cap \epsilon_3), P((\tilde{\epsilon}_1 \cup \epsilon_2 \cup \epsilon_3)^c \cap \epsilon_{4j}) : j = 1, 2, 3$ where $\tilde{\epsilon}_1 := \epsilon_{11} \cup \epsilon_l = \epsilon_{11} \cup \epsilon_{l2} \cup \epsilon_{l3}$.

Upper bound on $P(\epsilon_{11})$:- By conditional frequency typicality, for sufficiently large n , $P(\epsilon_{11}) \leq \frac{\eta}{32}$.

Upper bound on $P(\epsilon_{lj})$:- Using a second moment method similar to that employed in [15, Appendix A], we derive an upper bound on $P(\epsilon_{lj})$ in appendix A. In particular, we prove

$$P(\epsilon_{lj}) \leq 12 \exp\{-n(\delta - 32\eta)\} \quad (10)$$

for sufficiently large n . In deriving the above upper bound, we employed, among others, the bounds

$$\begin{aligned} K_j &> \delta > 0, \quad (S_j - T_j) - [\log \theta - H(U_j|Q)] > \delta > 0 \\ (S_j - T_j) + K_j - [\log \theta + H(X_j|Q) - H(U_j, X_j|Q)] &> \delta > 0. \end{aligned}$$

¹⁵Since the precise value of $\mathcal{L}_j(n)$ is necessary only in the derivation of the upper bound, it is provided in appendix A.

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 upper bounds on $P(\tilde{\epsilon}_1^c \cap \epsilon_2)$, $P((\tilde{\epsilon}_1 \cup \epsilon_2)^c \cap \epsilon_3)$, 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 codes, 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 B. In particular, we prove that for sufficiently large n ,

$$P(\tilde{\epsilon}_1^c \cap \epsilon_2) + P((\tilde{\epsilon}_1 \cup \epsilon_2)^c \cap \epsilon_3) \leq 2 \exp\{-n(n^2 \mu \eta_1^2 - 32\eta)\} + \frac{\eta}{32}. \quad (11)$$

Upper bound on $P((\tilde{\epsilon}_1 \cup \epsilon_2 \cup \epsilon_3)^c \cap \epsilon_{41})$:- In appendix C, we prove

$$P((\tilde{\epsilon}_1 \cup \epsilon_2 \cup \epsilon_3)^c \cap \epsilon_{41}) \leq 4 \exp\{-n[\delta - 28\eta_1 - 12\eta]\} \quad (12)$$

for sufficiently large n . In deriving (12), we employed, among others, the bounds

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

Upper bound on $P((\tilde{\epsilon}_1 \cup \epsilon_3)^c \cap \epsilon_{4j})$:- In appendix D, we prove

$$P((\tilde{\epsilon}_1 \cup \epsilon_2 \cup \epsilon_3)^c \cap \epsilon_{4j}) \leq 10 \exp\{-n(\delta - (9\eta + 16\eta_1))\} \quad (13)$$

for sufficiently large n . In deriving (13), we employed, among others, the bounds

$$(\log \theta - H(U_j|X_j, Y_j, Q)) - S_j > \delta > 0, \quad (\log \theta + H(X_j|Q) - H(U_j, X_j|Y_j, Q)) - (S_j + K_j) > \delta > 0,$$

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

$$(\log \theta + H(X_j|Q) - H(X_j, U_j|Y_j, Q)) - (K_j + L_j + S_j) > \delta > 0.$$

We now collect the derived upper bounds. From (10), (11), (12) and (13), we have

$$P\left(\bigcup_{j=1}^3 (\epsilon_{1j} \cup \epsilon_{3j} \cup \epsilon_{4j})\right) \leq \frac{\eta}{32} + 24 \exp\{-n(\delta - 32\eta)\} + 2 \exp\{-n(n^2 \mu \eta_1^2 - 32\eta)\} + \frac{\eta}{32} \\ + 4 \exp\{-n[\delta - 28\eta_1 - 12\eta]\} + 20 \exp\{-n(\delta - (9\eta + 16\eta_1))\}$$

The reader may recall that we need $\eta = \frac{1}{2^d} \min\{\tilde{\eta}, \delta\}$ and that $\eta_1 \geq 4\eta$ for the above bounds to hold. The reader may verify that, by choosing d sufficiently large, one can choose η and $\eta_1 \geq 4\eta$ such that the upper bound above decays exponentially. This completes the 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 that the chosen input codewords are jointly typical with respect to $p_{QU_2U_3\underline{X}\underline{Y}}$, 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. ■

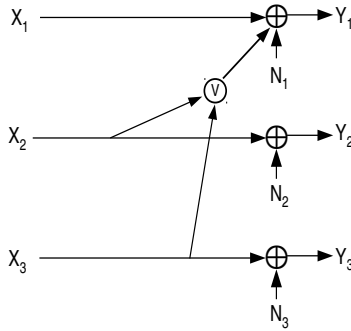


Fig. 2. A binary 3-to-1 IC described in example 3.

The coding technique proposed in the proof of theorem 5 is indeed a generalization of that proposed for example 1, and moreover capacity achieving for the same. We formalize this through the following corollary.

Corollary 2: For the 3-to-1 IC in example 1, if $\tau * \delta_1 < \min\{\delta_2, \delta_3\}$, then $\alpha_f^{3-1}(\tau, \frac{1}{2}, \frac{1}{2}) = \mathbb{C}(\tau)$.

It can be verified that $\beta(\tau, \frac{1}{2}, \frac{1}{2}, \underline{\delta}) = \alpha_f^{3-1}(p_{QU_2U_3XY})$ 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 (3).

In the sequel, we illustrate through an example the central claim of this article 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 3: Consider a binary 3-to-1 IC illustrated in figure 2 with $\mathcal{X}_j = \mathcal{Y}_j = \{0, 1\} : j \in [3]$ with channel transition probabilities $W_{\underline{Y}|\underline{X}}(y|\underline{x}) = BSC_{\delta_1}(y_1|x_1 \oplus (x_2 \vee x_3))BSC_{\delta_2}(y_2|x_2)BSC_{\delta_3}(y_3|x_3)$, where \vee denotes logical OR.¹⁶ Users' inputs are constrained with respect to a Hamming cost function, i.e., $\kappa_j(x) = x$ for $x \in \{0, 1\}$. Assume user j th 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 \mathcal{USB} -technique.

Lemma 3: Consider example 3 with $\delta := \delta_2 = \delta_3 \in (0, \frac{1}{2})$ and $\tau := \tau_2 = \tau_3 \in (0, \frac{1}{2})$. Let $\beta := \delta_1 * (2\tau - \tau^2)$. 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

$$h_b(\tau_1 * \delta_1) - h_b(\delta_1) + 2(h_b(\tau * \delta) - h_b(\delta)) > h_b(\tau_1(1 - \beta) + (1 - \tau_1)\beta) - h_b(\delta_1) \quad (14)$$

In particular, if (14) is true, $\alpha_u(\underline{\tau}) \subsetneq \beta(\underline{\tau}, \underline{\delta})$, where $\beta(\underline{\tau}, \underline{\delta})$ is defined in (3).

Please refer to appendix E for a proof. 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 $\underline{\delta} = (\delta_1, \delta, \delta)$. It therefore suffices to derive conditions under which $(h_b(\tau_1 * \delta_1) - h_b(\delta_1), h_b(\tau * \delta) - h_b(\delta), h_b(\tau * \delta) - h_b(\delta)) \in \alpha_f^{3-1}(\tau_1, \tau, \tau)$.

¹⁶ $BSC(\cdot|\cdot)$ has been defined in example 1.

Lemma 4: Consider example 3 with $\delta := \delta_2 = \delta_3 \in (0, \frac{1}{2})$ and $\tau := \tau_2 = \tau_3 \in (0, \frac{1}{2})$. Let $\beta := \delta_1 * (2\tau - \tau^2)$. 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,

$$h_b(\tau * \delta) - h_b(\delta) \leq \theta, \quad (15)$$

where $\theta = h_b(\tau) - h_b((1-\tau)^2) - (2\tau - \tau^2)h_b(\frac{\tau^2}{2\tau - \tau^2}) - h_b(\tau_1 * \delta_1) + h_b(\tau_1 * \beta)$. We therefore have $\alpha_f^{3-1}(\tau_1, \tau, \tau) = \mathbb{C}(\tau_1, \tau, \tau)$ if (15) holds.

Proof: The proof only involves identifying the appropriate test channel $p_{QU_2U_3XY} \in \mathbb{D}_f^{3-1}(\tau_1, \tau, \tau)$. 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 $p_{U_j X_j}(0, 0) = 1 - p_{U_j X_j}(1, 1) = 1 - \tau$ and therefore $P(U_j = 2) = P(X_j \neq U_j) = 0$ for $j = 2, 3$. It is easily verified that $p_{QU_2U_3XY} \in \mathbb{D}_f^{3-1}(\tau_1, \tau, \tau)$, i.e., in particular respects the cost constraints.

The choice of this test channel, particularly the ternary field, 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 $p_{QU_2U_3XY}$, $\alpha_f^{3-1}(p_{QU_2U_3XY})$ is defined as the set of rate triples $(R_1, R_2, R_3) \in [0, \infty)^3$ that satisfy

$$\begin{aligned} R_1 &< \min\{0, \theta\} + h_b(\tau_1 * \delta_1) - h_b(\delta_1), \quad R_j < h_b(\tau * \delta) - h_b(\delta) : j = 2, 3 \\ R_1 + R_j &< h_b(\tau_1 * \delta_1) - h_b(\delta_1) + \theta, \end{aligned} \quad (16)$$

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, $(h_b(\tau_1 * \delta_1) - h_b(\delta_1), h_b(\tau * \delta) - h_b(\delta), h_b(\tau * \delta) - h_b(\delta)) \in \text{cocl}(\alpha_f^{3-1}(p_{U_2U_3XY}))$ if (15) is satisfied. ■

Conditions (14) and (15) 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.

A skeptical reader will wonder whether the utility of PCC hinges on the additive multiple access channel (MAC) $Y_1 = X_1 \oplus (X_2 \vee X_3) \oplus N_1$. The following example provides conclusive evidence that this is indeed not the case.

Example 4: Consider a binary 3-to-1 IC illustrated in figure 3 with $\mathcal{X}_j = \mathcal{Y}_j = \{0, 1\} : j \in [3]$ with channel transition probabilities $W_{Y|X}(y|x) = MAC(y_1|x_1, x_2 \vee x_3)BSC_\delta(y_2|x_2)BSC_\delta(y_3|x_3)$, where $MAC(0|0, 0) = 0.989$, $MAC(0|0, 1) = 0.01$, $MAC(0|1, 0) = 0.02$, $MAC(0|1, 1) = 0.993$ and $MAC(0|b, c) + MAC(1|b, c) = 1$ for each $(b, c) \in \{0, 1\}^2$. Users' inputs are constrained with respect to a Hamming cost function, i.e., $\kappa_j(x) = x$ for $x \in \{0, 1\}$. Assume user j th input is constrained to an average cost per symbol of $\tau_j \in (0, \frac{1}{2})$, where $\tau := \tau_2 = \tau_3$.

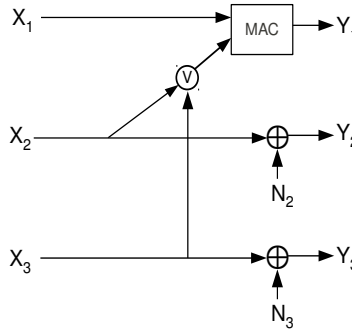


Fig. 3. A binary 3-to-1 IC described in example 4.

Our study of example 4 closely mimics that of example 3. In particular, we derive conditions under which $\underline{C}^* := (C_1, h_b(\tau * \delta) - h_b(\delta), h_b(\tau * \delta) - h_b(\delta)) \in \alpha_f^{3-1}(\underline{\mathcal{T}})$ and $\underline{C}^* \notin \alpha_u(\underline{\mathcal{T}})$, where $\underline{\mathcal{T}} := (\tau_1, \tau, \tau)$,

$$C_1 := \sup_{p_{XY} \in \mathcal{D}(\underline{\mathcal{T}})} I(X_1; Y_1 | X_2 \vee X_3), \quad (17)$$

$$\mathcal{D}(\underline{\mathcal{T}}) := \left\{ \begin{array}{l} p_{XY} \text{ is a pmf on } \underline{\mathcal{X}} \times \underline{\mathcal{Y}} \text{ such that (i) } p_{Y|X} = W_{Y|X} \text{ is the channel transition probabilities} \\ \text{of example 4, (ii) } p_{X_j} = p_{X_1} p_{X_2} p_{X_3}, p_{X_j}(1) = \tau \text{ for } j = 2, 3 \text{ and (iii) } p_{X_1}(1) \leq \tau_1 \end{array} \right\}. \quad (18)$$

By strict concavity of $I(X_1; Y_1 | X_2 \vee X_3)$ in p_{X_1} , and the compactness of $\mathcal{D}(\underline{\mathcal{T}})$, there exists a unique p_{XY}^* with respect to which $I(X_1; Y_1 | X_2 \vee X_3) = C_1$. We are now set to state the conditions under which $\underline{C}^* \notin \alpha_u(\underline{\mathcal{T}})$.

Lemma 5: Consider example 4 and let $\underline{C}^*, C_1, \mathcal{D}(\underline{\mathcal{T}}), p_{XY}^*$ be defined as above. If

$$C_1 + 2(h_b(\tau * \delta) - h_b(\delta)) = I(X_1; Y_1 | X_2 \vee X_3) + 2(h_b(\tau * \delta) - h_b(\delta)) > I(\underline{X}; Y_1), \quad (19)$$

where the mutual information terms $I(X_1; Y_1 | X_2 \vee X_3), I(\underline{X}; Y_1)$ are evaluated with respect to p_{XY}^* , then $\underline{C}^* \notin \alpha_u(\underline{\mathcal{T}})$.

The reader will recognize that above lemma is the counterpart of lemma 3 for example 4. We provide a (sketch of the) proof in appendix F.

Lemma 6: Consider example 4 C_1, p_{XY}^* be as defined above. If $h_b(\tau^2) + (1 - \tau^2)h_b(\frac{(1-\tau)^2}{1-\tau^2}) + H(Y_1 | X_2 \vee X_3) - H(Y_1) \leq \min\{H(X_2 | Y_2)H(X_3 | Y_3)\}$, where the entropies are evaluated with respect to p_{XY}^* , then $(C_1, h_b(\delta * \tau) - h_b(\delta), h_b(\delta * \tau) - h_b(\delta)) \in \alpha_f^{3-1}(\underline{\mathcal{T}})$.

Proof: As in proof of lemma 4, we identify an appropriate test channel $p_{QU_2U_3XY} \in \mathbb{D}_f(\underline{\mathcal{T}})$ for which $(C_1, h_b(\delta * \tau) - h_b(\delta), h_b(\delta * \tau) - h_b(\delta)) \in \alpha_f^{3-1}(p_{QU_2U_3XY})$. Let $\mathcal{Q} = \phi$ be empty, $\mathcal{U}_2 = \mathcal{U}_3 = \{0, 1, 2\}$. Let $p_{\underline{X}} = p_{\underline{X}}^*$. Let $p_{U_j X_j}(0, 0) = 1 - p_{U_j X_j}(1, 1) = 1 - \tau$ and therefore $P(U_j = 2) = P(X_j \neq U_j) = 0$ for $j = 2, 3$. It is easily verified that $p_{QU_2U_3XY} \in \mathbb{D}_f^{3-1}(\underline{\mathcal{T}})$, i.e, in particular respects the cost constraints.

It maybe verified that the hypothesis $h_b(\tau^2) + (1 - \tau^2)h_b(\frac{(1-\tau)^2}{1-\tau^2}) + H(Y_1 | X_2 \vee X_3) - H(Y_1) = H(U_2 \oplus_3 U_3) + H(Y_1 | X_2 \vee X_3) - H(Y_1) = H(U_2 \oplus_3 U_3) + H(Y_1 | U_2 \oplus_3 U_3) - H(Y_1) = H(U_2 \oplus_3 U_3 | Y_1)$. we therefore have $H(U_2 \oplus_3 U_3 | Y_1) \leq \min\{H(X_2 | Y_2)H(X_3 | Y_3)\}$. This implies (i) $H(U_j) \geq H(U_2 \oplus_3 U_3 | Y_1)$ and (ii) $H(U_j) -$

$H(U_2 \oplus U_3|Y_1) \geq H(U_j) - H(U_j|Y_j) = I(U_j; Y_j) = I(X_j; Y_j) = h_b$. Employing these in bounds characterizing $\alpha_f^{3-1}(p_{QU_2U_3XY})$ and the marginal $p_{XY} = p_{XY}^*$, it can be verified that $(C_1, h_b(\delta * \tau) - h_b(\delta), h_b(\delta * \tau) - h_b(\delta)) \in \alpha_f^{3-1}(p_{QU_2U_3XY})$. ■

For example 4, with $\tau_1 = 0.01, \tau = \tau_2 = \tau_3 = 0.1525, \delta = 0.067$, the conditions stated in lemma 5 and 6 hold simultaneously. For this channel, $p_{X_1}^*(1) = 0.99$,

$$C_1 + 2(h_b(\tau * \delta) - h_b(\delta)) - I(X; Y_1) = 0.0048,$$

and

$$\min\{H(X_2|Y_2)H(X_3|Y_3)\} - [h_b(\tau^2) + (1 - \tau^2)h_b(\frac{(1 - \tau)^2}{1 - \tau^2}) + H(Y_1|X_2 \vee X_3) - H(Y_1)] = 0.0031.$$

A note on our choice of the MAC $X_1, X_2 \vee X_3 - Y_1$ is in order. The reader will recognize the MAC being ‘quite close’ to the additive scenario $Y_1 = X_1 \oplus (X_2 \vee X_3) \oplus N_1$ studied in example 3. In order for coset codes to outperform unstructured codes, we do not need the MAC to be so ‘close’ to the additive MAC. The need for the MAC to be ‘so close’ is a consequence of our desire to provide an *analytical* proof for strict sub-optimality of unstructured codes. Note that since we (i) do not resort to outer bounds, (ii) wish to provide analytical upper bounds to the rates achievable using unstructured codes, and (iii) cannot compute any of the associated rates in a reasonable time, we demand the MAC to be such that coset codes achieve the maximum possible rate for user 1, with users 2 and 3 constrained to achieve their PTP capacities,¹⁷ and unstructured codes to be strictly sub-optimal. This justifies the ‘closeness’ of the considered MAC to an additive MAC. Finally, the above documented findings indicate that if structured codes yield gains for particular additive scenarios, then one can reason out the presence of such gains for ‘close’ non-additive scenarios simply by appealing to the continuity of rate regions in the channel parameters.

B. Step II: PCC rate region for a general discrete 3-IC using codes built over finite fields

In this section, we employ PCC to manage interference seen by every 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 5.

User j splits its message M_j of rate $R_j = L_j + T_{ji} + T_{jk}$ into three parts $(M_{ji}^U, M_{jk}^U, M_j^X)$, where i, j, k are distinct indices in $\{1, 2, 3\}$. Let $\mathcal{U}_{ji} = \mathcal{F}_{\theta_i}, \mathcal{U}_{jk} = \mathcal{F}_{\theta_k}$ be finite fields. Let $\lambda_{ji} \subseteq \mathcal{U}_{ji}^n$ denote an (n, s_{ji}, t_{ji}) PCC and $\lambda_{jk} \subseteq \mathcal{U}_{jk}^n$ denote an (n, s_{jk}, t_{jk}) PCC. If we let $S_{ji} := \frac{s_{ji}}{n} \log \theta_i, T_{ji} := \frac{t_{ji}}{n} \log \theta_i$ and $S_{jk} := \frac{s_{jk}}{n} \log \theta_k, T_{jk} := \frac{t_{jk}}{n} \log \theta_k$ then recall that $\lambda_{ji}, \lambda_{jk}$ are coset codes of rates S_{ji}, S_{jk} partitioned into $\exp\{nT_{ji}\}, \exp\{nT_{jk}\}$ bins respectively. Observe that cosets λ_{ji} and λ_{ki} are built over the same finite field \mathcal{F}_{θ_i} . To enable contain the range

¹⁷Note that we are demanding the channel to permit user 1 communicate at a rate as though he knew all of the non-linear interference. Moreover, we are employing linear codes to decode the non-linear interference efficiently.

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{X}_j . Codewords of \mathcal{C}_j are partitioned into $\exp\{nL_j\}$ bins.

M_{ji}^U, M_{jk}^U and M_j^X 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 $p_{U_{ji}U_{jk}X_j}$ defined on $\mathcal{U}_{ji} \times \mathcal{U}_{jk} \times \mathcal{X}_j$. The corresponding codeword chosen from \mathcal{C}_j is input on the channel.

Decoder j receives Y_j^n and looks for all triples $(u_{ji}^n, u_{jk}^n, x_j^n)$ of codewords in $\lambda_{ji} \times \lambda_{jk} \times \mathcal{C}_j$ for which there exists a $u_{\oplus}^n \in (\lambda_{ij} \oplus \lambda_{kj})$ such that $(u_{\oplus}^n, u_{ji}^n, u_{jk}^n, x_j^n, Y_j^n)$ are jointly typical with respect to $p_{U_{ij} \oplus U_{kj}, U_{ji}, U_{jk}, X_j, Y_j}$. 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 5. In particular, the codewords of \mathcal{C}_j are chosen independently according to $\prod_{t=1}^n p_{X_j|Q}(\cdot|q^t)$, 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.¹⁸ 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 9: Let $\mathbb{D}_f(\mathcal{T})$ denote the collection of probability mass functions (p_{QUXY}) defined on $\mathcal{Q} \times \mathcal{U} \times \mathcal{X} \times \mathcal{Y}$, where

- 1) \mathcal{Q} is an arbitrary finite set,
- 2) $\mathcal{U}_{ij} = \mathcal{F}_{\theta_j}$ ¹⁹ for each $1 \leq i, j \leq 3$, and $\underline{U} := \mathcal{U}_{12} \times \mathcal{U}_{13} \times \mathcal{U}_{21} \times \mathcal{U}_{23} \times \mathcal{U}_{31} \times \mathcal{U}_{32}$,
- 3) $\underline{U} := (U_{12}, U_{13}, U_{21}, U_{23}, U_{31}, U_{32})$,

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

For $p_{QUXY} \in \mathbb{D}_f(\mathcal{T})$, let $\alpha_f(p_{QUXY})$ 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

$$S_{A_j} - T_{A_j} + K_j > \sum_{a_j \in A_j} \log |\mathcal{U}_{a_j}| + H(X_j|Q) - H(U_{A_j}, X_j|Q), \quad (20)$$

$$S_{A_j} - T_{A_j} > \sum_{a_j \in A_j} \log |\mathcal{U}_{a_j}| - H(U_{A_j}|Q), \quad (21)$$

¹⁸The reader is encouraged to confirm that the distribution induced herein is a simple generalization of that employed in proof of theorem 5.

¹⁹Recall \mathcal{F}_{θ_j} is the finite field of cardinality θ_j .

$$\begin{aligned}
S_{A_j} &< \sum_{a \in A_j} \log |\mathcal{U}_a| - H(U_{A_j}|Q, U_{A_j^c}, U_{ij} \oplus U_{kj}, X_j, Y_j) \\
S_{A_j} + S_{ij} &< \sum_{a \in A_j} \log |\mathcal{U}_a| + \log \theta_j - H(U_{A_j}, U_{ij} \oplus U_{kj}|Q, U_{A_j^c}, X_j, Y_j) \\
S_{A_j} + S_{kj} &< \sum_{a \in A_j} \log |\mathcal{U}_a| + \log \theta_j - H(U_{A_j}, U_{ij} \oplus U_{kj}|Q, U_{A_j^c}, X_j, Y_j) \\
S_{A_j} + K_j + L_j &< \sum_{a \in A_j} \log |\mathcal{U}_a| + H(X_j) - H(U_{A_j}, X_j|Q, U_{A_j^c}, U_{ij} \oplus U_{kj}, Y_j) \\
S_{A_j} + K_j + L_j + S_{ij} &< \sum_{a \in A_j} \log |\mathcal{U}_a| + \log \theta_j + H(X_j) - H(U_{A_j}, X_j, U_{ij} \oplus U_{kj}|Q, U_{A_j^c}, Y_j) \\
S_{A_j} + K_j + L_j + S_{kj} &< \sum_{a \in A_j} \log |\mathcal{U}_a| + \log \theta_j + H(X_j) - H(U_{A_j}, X_j, U_{ij} \oplus U_{kj}|Q, U_{A_j^c}, Y_j),
\end{aligned} \tag{22}$$

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

$$\alpha_f(\underline{\tau}) = \text{cocl} \left(\bigcup_{\substack{p_{QU_{XY}} \in \\ \mathbb{D}_f(\underline{\tau})}} \alpha_f(p_{QU_{XY}}) \right).$$

Theorem 6: For 3-IC $(\underline{\mathcal{X}}, \underline{\mathcal{Y}}, W_{\underline{Y}|\underline{X}}, \kappa)$, $\alpha_f(\underline{\tau})$ is achievable, i.e., $\alpha_f(\underline{\tau}) \subseteq \mathbb{C}(\underline{\tau})$.

Since all the non-trivial elements of this proof are captured in the proof of theorem 5, and is only more involved in notation, we omit the same.

The above coding technique presents an approach to simultaneously manage interference at all of the receivers. It is natural to question whether the use of structured codes to manage interference comes at a cost of respective individual communication. We now provide a simple generalization of example 1 that requires managing interference at two receivers. In contrast to [13], wherein the benefit of interference alignment can be exploited at all receivers, channels equipped with finite alphabets present a fundamental trade-off in managing interference and enabling individual respective communication.

Example 5: Consider a binary additive 3-to-1 IC illustrated in figure 4 with $\mathcal{X}_j = \mathcal{Y}_j = \{0, 1\} : j \in [3]$ with channel transition probabilities $W_{\underline{Y}|\underline{X}}(\underline{y}|\underline{x}) = BSC_{\delta_1}(y_1|x_1 \oplus x_2 \oplus x_3)BSC_{\delta_2}(y_2|x_2 \oplus x_3)BSC_{\delta_3}(y_3|x_3)$. 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.

In order to illustrate the trade-off, let us consider the case $\delta := \delta_2 = \delta_3$ is arbitrarily close to, but greater than $\tau * \delta_1$. For example, one can choose $\delta_1 = 0.01, \tau = \frac{1}{8}$ and $\delta = 0.1326$. If receiver 1 desires communication at $h_b(\delta_1 * \tau) - h_b(\delta_1)$, it needs to decode $X_2 \oplus X_3$. To satisfy user 1's desire, users 2 and 3 have two options. Either employ codes of rates R_2 and R_3 such that $R_2 + R_3 < 1 - h_b(\delta_1 * \tau)$, or employ cosets of the same code with a hope to boost individual rates. In the latter case, user 2 is hampered by the interference caused to it by user 3. While we do not provide a detailed analysis, we encourage the reader to contrast this to the Gaussian IC studied in

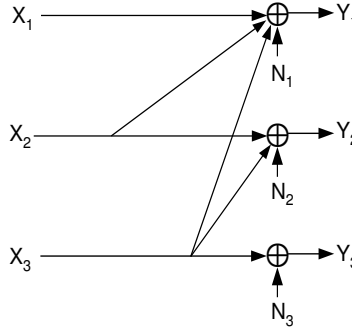


Fig. 4. A binary additive 3-to-1 IC described in example 5.

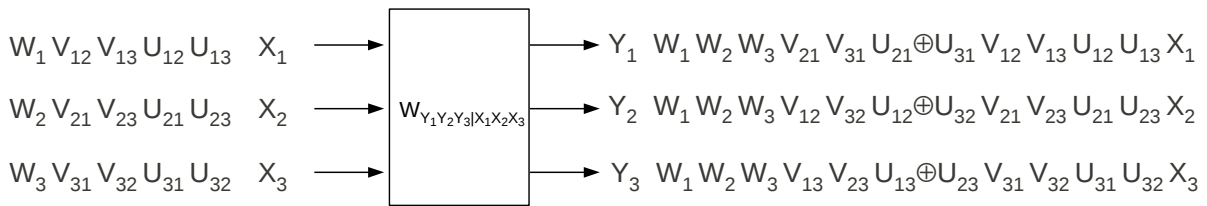


Fig. 5. Collection of random variables associated with coding technique that incorporates unstructured and partitioned coset codes

[13], wherein the richness of the real field enables each receiver to exploit the benefits of alignment. We conjecture an inherent trade-off in the ability to manage interference over finite valued channels using coset codes, and enable individual respective communication.

C. Step III: Enlarging the PCC rate region using unstructured codes

Let us describe a coding technique that unifies both unstructured and partitioned coset codes. We follow the approach of Ahlswede and Han [22, Section VI]. Refer to figure 5 for an illustration of the random variables involved. Each user splits its message into 5 parts. The W -random variable is decoded by all users. In addition, each user decodes a univariate component of the message of the other users. This is represented by the random variable V . Furthermore, it decodes a bivariate interference component denoted using U . Lastly, each decoder decodes all parts of its intended message. As was done by Han and Kobayashi [7], this achievable rate region can be enlarged through the use of a time sharing random variable. Clearly, a description of the above rate region is involved. In the sequel, we illustrate the key elements via a simplified achievable rate region. In particular, we employ PCC and unstructured codes to manage interference seen by only one receiver, say receiver 1 and state the corresponding achievable rate region. We begin with a description of the same.

Definition 10: Consider a 3-IC $(\underline{\mathcal{X}}, \underline{\mathcal{Y}}, W_{\underline{Y}|\underline{X}}, \underline{\kappa})$. Let $\mathbb{D}_{uf}(\underline{\tau})$ denote the collection of distributions $p_{QU_2V_2U_3V_3\underline{X}\underline{Y}}$ defined over $\mathcal{Q} \times \mathcal{U}_2 \times \mathcal{V}_2 \times \mathcal{U}_3 \times \mathcal{V}_3 \times \underline{\mathcal{X}} \times \underline{\mathcal{Y}}$, where $\mathcal{U}_2 = \mathcal{U}_3$ is a finite field and \mathcal{V}_2 and \mathcal{V}_3 are finite sets, such that (i) $p_{\underline{Y}|\underline{X}U_2V_2U_3V_3} = W_{\underline{Y}|\underline{X}}$, (ii) $X_1, (U_2, V_2, X_2)$ and (U_3, V_3, X_3) are conditionally independent given Q , (iii)

$\mathbb{E}\{\kappa_j(X_j)\} \leq \tau_j$ for $j = 1, 2, 3$. For $p_{QU_2V_2U_3V_3XY} \in \mathbb{D}_{uf}(\underline{\tau})$, let $\alpha_{uf}^{3-1}(p_{QU_2V_2U_3V_3XY})$ be defined as the set of rate triples $(R_1, R_2, R_3) \in [0, \infty)^3$ for which $\mathcal{S}_{uf}(p_{QU_2V_2U_3V_3XY}, \underline{R})$ is non-empty, where $\mathcal{S}_{uf}(p_{QU_2V_2U_3V_3XY}, \underline{R})$ is defined as the vectors $(S_{j1}, T_{j1}, S_{j2}, T_{j2}, L_j : j = 2, 3) \in [0, \infty)^{10}$ that satisfy

$$S_{j2} - T_{j2} > \log \theta - H(U_j|V_j, Q), \quad R_j = T_{j1} + T_{j2} + L_j : j = 2, 3 \quad (23)$$

$$L_j + S_{j2} < \log \theta - H(U_j|V_j, Q) + I(U_j, X_j; Y_j|V_j, Q), \quad T_{j1} + L_j < I(U_j; V_j|Q) + I(V_j, X_j; Y_j|U_j, Q) : j = 2, 3, \quad (24)$$

$$L_j < I(X_j; Y_j|U_j, V_j, Q), \quad T_{j1} + S_{j2} + L_j < \log \theta - H(U_j|V_j, Q) + I(U_j, V_j, X_j; Y_j|Q) : j = 2, 3 \quad (25)$$

$$R_1 < I(X_1; Y_1, V_2, V_3, U_2 \oplus U_3|Q), \quad R_1 + S_{j2} < \log \theta - H(U_2 \oplus U_3|Q) + I(X_1, U_2 \oplus U_3; V_2, V_3, Y_1|Q) : j = 2, 3 \quad (26)$$

$$R_1 + T_{j1} < I(X_1, V_j; V_j, U_2 \oplus U_3, Y_1|Q) : j = 2, 3, \quad T_{21} + T_{31} + R_1 < I(V_2, V_3, X_1; U_2 \oplus U_3, Y_1|Q)$$

$$R_1 + T_{j1} + S_{k2} < \log \theta - H(U_2 \oplus U_3|V_j, Q) + I(X_1, V_j, U_2 \oplus U_3; V_j, Y_1|Q) : j = 2, 3 \text{ and } k = 2, 3 \quad (27)$$

$$T_{21} + T_{31} + S_{j2} + R_1 < \log \theta - H(U_2 \oplus U_3|X_1, V_2, V_3, Q) + I(X_1, V_2, V_3, U_2 \oplus U_3; Y_1|Q) \quad (28)$$

where $\theta = |\mathcal{U}_2| = |\mathcal{U}_3|$. Let

$$\alpha_{uf}^{3-1}(\underline{\tau}) = \text{cocl} \left(\bigcup_{\substack{p_{QU_2V_2U_3V_3XY} \in \\ \mathbb{D}_{uf}(\underline{\tau})}} \alpha_{uf}^{3-1}(p_{QU_2V_2U_3V_3XY}) \right).$$

Theorem 7: For 3-IC $(\underline{\mathcal{X}}, \underline{\mathcal{Y}}, W_{Y|X}, \kappa)$, $\alpha_{uf}^{3-1}(\underline{\tau})$ is achievable, i.e., $\alpha_{uf}^{3-1}(\underline{\tau}) \subseteq \mathbb{C}(\underline{\tau})$.

We provide a brief sketch of achievability. For simplicity, user 1 builds an unstructured independent code of rate R_1 over \mathcal{X}_1 by choosing codewords independently and identically according to $p_{X_1}^n$. For $j = 2, 3$, user j builds three random codebooks - one each over $\mathcal{V}_j, \mathcal{U}_j, \mathcal{X}_j$ respectively. An unstructured and independent codebook of rate T_{j1} is built over \mathcal{V}_j by choosing codewords independently and identically according to $p_{V_j}^n$. A random PCC $(n, \frac{nS_{j2}}{\log \theta}, \frac{nT_{j2}}{\log \theta}, G_j, B_j^n, I_j)$, denoted Λ_j , is built over \mathcal{U}_j . As before the PCC's of users 2 and 3 overlap, i.e., if $j_1 \leq j_2$, then $g_{j_2}^T = [g_{j_1}^T, g_{j_2/j_1}^T]$. Consider a codeword in \mathcal{V}_j -codebook and a bin in the PCC. For every such pair, a random unstructured independent codebook is constructed over \mathcal{X}_j .

User j th message is split into three parts - *univariate* part, *bivariate* part and *private* part. The univariate part indexes a codeword, say $V_j^n(M_{jV})$ in \mathcal{V}_j -codebook. The bivariate part indexes a bin in the PCC. A codeword, say $U_j^n(M_{jU})$ is chosen in the indexed bin such that $(V_j^n(M_{jV}), U_j^n(M_{jU}))$ is jointly typical according to the probability distribution $p_{QV_jU_j}$, the marginal of $p_{QU_2V_2U_3V_3XY} \in \mathbb{D}_{uf}(\underline{\tau})$ in question. The codewords of the codebook built over \mathcal{X}_j , corresponding to (M_{jV}, M_{jU}) , are independently and identically distributed according to $p_{X_j|V_jU_j}^n(\cdot|V_j^n(M_{jV}), U_j^n(M_{jU}))$. The private part M_{jX} indexes a codeword in this codebook. This codeword is input on the channel by user j . User 1 inputs the codeword from its \mathcal{X}_1 -codebook that is indexed by its message. It can be verified that the inequality in (23) ensures users 2 and 3 find jointly typical triples of codewords.

Users 2 and 3 employ a simple point-to-point decoding technique. However, note that the codebook over \mathcal{X}_j is conditionally built. Therefore, an error in decoding the correct \mathcal{U}_j - or \mathcal{V}_j -codeword is interpreted as an error even in decoding the \mathcal{X}_j -codeword. It can be verified that (24), (25) ensure the probability of decoding error at receiver j decays exponentially with block length n .

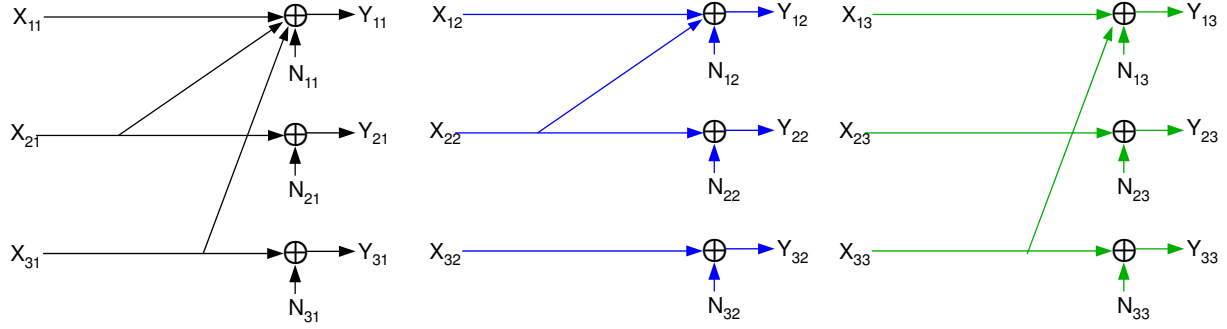


Fig. 6. A 3-IC with univariate and bivariate interference components.

User 1 constructs the sum codebook $\Lambda_2 \oplus \Lambda_3 := \{u_2^n \oplus u_3^n : u_j^n \in \Lambda_j : j = 2, 3\}$ and decodes into $\mathcal{V}_2, \mathcal{V}_3, \Lambda_2 \oplus \Lambda_3, \mathcal{X}_1$ codebooks. In particular it looks for a quadruple of codewords in these codebooks that are jointly typical with the received vector Y_1^n according to $p_{QV_2, V_3, U_2 \oplus U_3 | Y_1}$. It can be verified that (26) - (28) imply the probability of decoding error at receiver 1 decays exponentially with block length.

Example 6: We briefly describe an example wherein the above coding technique can yield larger achievable rate regions than ones based exclusively either on PCC or on unstructured based codes. Consider the 3-IC depicted in figure 6. For each $j = 1, 2, 3$, the input alphabet $\mathcal{X}_j = \times_{k=1}^3 \mathcal{X}_{jk}$ and output alphabet is $\mathcal{Y}_j = \times_{k=1}^3 \mathcal{Y}_{jk}$ where $\mathcal{X}_{jk} = \mathcal{Y}_{jk} = \{0, 1\}$. Essentially, each user can input three binary digits on the channel and each receiver observes three binary digits per channel use. Let $X_{jk} : k = 1, 2, 3$ denote the three binary digits input by transmitter j and $Y_{jk} : k = 1, 2, 3$ denote the three digits observed by receiver j . Figure 6 depicts the input-output relationship. Let us also assume the Bernoulli noise processes $N_{jk} : j = 1, 2, 3, k = 1, 2, 3$ are mutually independent. Users 2 and 3 enjoy complete free point-to-point links for each of the digits. They are only constrained by noise that is modeled by the corresponding Bernoulli noise processes. Receiver 1's digit Y_{11} experiences bivariate interference. Its 2nd and 3rd digits experience univariate interferences.²⁰ The reader will recognize the need for receiver 1 to decode univariate and bivariate parts of user 2 and 3's transmissions. The above coding technique enables the same.

We conclude this section with a discussion, wherein, we employ the notion of common information to argue, more fundamentally, the need to decode bivariate interference components. Let us view the above coding technique from the perspective of common information in the sense of Gacs, Körner and Witsenhausen [23] [24]. Let $K(A; B)$ denote the common information of two random variables A and B . Let \tilde{X}_j denote the collection of random variables decoded at decoder j . The CHK scheme for 2-IC can be interpreted as inducing non-trivial common information between \tilde{X}_1 and \tilde{X}_2 , and $K(\tilde{X}_1; \tilde{X}_2) = H(W_1, W_2)$. The question that comes next is how to extend common

²⁰The IC depicted in figure 6 can be used to model a scenario wherein Tx-Rx pair 1 is assigned frequency bands around carrier frequencies f_1, f_2, f_3 , Tx-Rx pair 2 is assigned frequency bands around carrier frequencies f_1, f_2, f_4 , Tx-Rx pair 3 is assigned frequency bands around carrier frequencies f_1, f_3, f_5 respectively. If the powers transmitted by users 2 and 3 are large, then user 1 does not cause any appreciable interference to users 2 and 3. The interference caused by transmissions of Txs 2 and 3 on each other in frequency band around f_1 has been ignored by this model.

information to 3 random variables? We can consider the following vector as the common information among three random variables A , B and C :

$$[K(A; B; C), K(A; B), K(B; C), K(C; A)],$$

where $K(A; B; C)$ is defined in a natural way. We refer to this as univariate common information as they are characterized using univariate function of the random variables. The \mathcal{USB} -technique induces non-trivial univariate common information among \tilde{X}_1 , \tilde{X}_2 and \tilde{X}_3 , and

$$K(\tilde{X}_1; \tilde{X}_2; \tilde{X}_3) = H(W_1, W_2, W_3), \quad K(\tilde{X}_j; \tilde{X}_k) = H(V_{kj}, V_{jk}).$$

The common information captured via univariate functions can be enhanced with the following components captured via bivariate functions. Define

$$\tilde{K}(A, B; C) := \sup_{h_1, g_3} \inf_{f_1, f_2, g_1, g_2} \left\{ H(V_3 | V_1, V_2) : \begin{array}{l} V_1 = f_1(A) = g_1(C), V_2 = f_2(B) = g_2(C), V_3 = h(A, B) = g_3(C) \text{ where } f_1: A \rightarrow \mathcal{V}, \\ f_2: B \rightarrow \mathcal{V}, g_i: C \rightarrow \mathcal{V}: i=1,2, h: A \times B \rightarrow \mathcal{V} \text{ are maps into a finite set } \mathcal{V} \end{array} \right\}.$$

We define common information among three random variables as a seven-dimensional vector as follows:

$$[K(A; B; C), K(A; B), K(B; C), K(C; A), \tilde{K}(A, B; C), \tilde{K}(B, C; A), \tilde{K}(C, A; B)].$$

We refer to the last three components as bivariate common information. Note that the \mathcal{USB} -technique induces trivial bivariate common information among \tilde{X}_1, \tilde{X}_2 and \tilde{X}_3 . The PCC technique induces non-trivial bivariate common information among them, and $\tilde{K}(\tilde{X}_i, \tilde{X}_j; \tilde{X}_k) = H(U_{ik} \oplus U_{jk})$ for all $i \neq j \neq k$.

VI. STEP IV: ACHIEVABLE RATE REGION USING PCC BUILT OVER ABELIAN GROUPS

In this section, we present PCC scheme using codes built on Abelian groups. The rate region we get can be interpreted as the algebraic extension of that given in theorem 5.

A. Preliminaries about groups

For an Abelian group G , let $\mathcal{P}(G)$ denote the set of all distinct primes which divide $|G|$ and for a prime $p \in \mathcal{P}(G)$ let $S_p(G)$ be the corresponding Sylow subgroup of G . It is known [25, Theorem 3.3.1] that any Abelian group G can be decomposed as a direct sum of its Sylow subgroups in the following manner

$$G = \bigoplus_{p \in \mathcal{P}(G)} S_p(G) \quad (29)$$

Furthermore, each Sylow subgroup $S_p(G)$ can be decomposed into \mathbb{Z}_{p^r} groups as follows:

$$S_p(G) \cong \bigoplus_{r \in \mathcal{R}_p(G)} \mathbb{Z}_{p^r}^{M_{p,r}} \quad (30)$$

where $\mathcal{R}_p(G) \subseteq \mathbb{Z}^+$ and for $r \in \mathcal{R}_p(G)$, $M_{p,r}$ is a positive integer. Note that $\mathbb{Z}_{p^r}^{M_{p,r}}$ is defined as the direct sum of the ring \mathbb{Z}_{p^r} with itself for $M_{p,r}$ times. Combining equations (29) and (30), we can represent any Abelian group as follows:

$$G \cong \bigoplus_{p \in \mathcal{P}(G)} \bigoplus_{r \in \mathcal{R}_p(G)} \mathbb{Z}_{p^r}^{M_{p,r}} = \bigoplus_{p \in \mathcal{P}(G)} \bigoplus_{r \in \mathcal{R}_p(G)} \bigoplus_{m=1}^{M_{p,r}} \mathbb{Z}_{p^r}^{(m)} \quad (31)$$

where $\mathbb{Z}_{p^r}^{(m)}$ is called the m^{th} \mathbb{Z}_{p^r} ring of G or the $(p, r, m)^{\text{th}}$ ring of G . Equivalently, this can be written as follows

$$G \cong \bigoplus_{(p,r,m) \in \mathcal{G}(G)} \mathbb{Z}_{p^r}^{(m)}$$

where $\mathcal{G}(G) \subseteq \mathbb{P} \times \mathbb{Z}^+ \times \mathbb{Z}^+$ is defined as:

$$\mathcal{G}(G) = \{(p, r, m) \in \mathbb{P} \times \mathbb{Z}^+ \times \mathbb{Z}^+ | p \in \mathcal{P}(G), r \in \mathcal{R}_p(G), m \in \{1, 2, \dots, M_{p,r}\}\}$$

This implies that any element a of the Abelian group can be regarded as a vector whose components are indexed by $(p, r, m) \in \mathcal{G}(G)$ and whose $(p, r, m)^{\text{th}}$ component $a_{p,r,m}$ takes values from the ring \mathbb{Z}_{p^r} . With a slight abuse of notation, we represent an element a of G as

$$a = \bigoplus_{(p,r,m) \in \mathcal{G}(G)} a_{p,r,m}$$

For example let $G = \mathbb{Z}_3^5 \oplus \mathbb{Z}_4^3 \oplus \mathbb{Z}_8$. Then we have $\mathcal{P}(G) = \{2, 3\}$, $\mathcal{R}_2(G) = \{2, 3\}$, $\mathcal{R}_3(G) = \{1\}$, $M_{2,2} = 3$, $M_{2,3} = 1$, $M_{3,1} = 5$, and

$$\mathcal{G}(G) = \{(2, 2, 1), (2, 2, 2), (2, 2, 3), (2, 3, 1), (3, 1, 1), (3, 1, 2), (3, 1, 3), (3, 1, 4), (3, 1, 5)\}.$$

For two elements $a, b \in G$, we have

$$a + b = \bigoplus_{(p,r,m) \in \mathcal{G}(G)} a_{p,r,m} +_{p^r} b_{p,r,m}$$

where $+$ denotes the group operation and $+_{p^r}$ denotes addition mod- p^r . Let $[\cdot]_{p,r,m}$ denote the $(p, r, m)^{\text{th}}$ component of it's argument.

Let $\mathbb{I}_{G:p,r,m} \in G$ be a generator for the group which is isomorphic to the $(p, r, m)^{\text{th}}$ ring of G . Then we have

$$a = \sum_{(p,r,m) \in \mathcal{G}(G)}^{(G)} a_{p,r,m} \mathbb{I}_{G:p,r,m} \quad (32)$$

where the summations are done with respect to the group operation and the multiplication $a_{p,r,m} \mathbb{I}_{G:p,r,m}$ is by definition the summation (with respect to the group operation) of $\mathbb{I}_{G:p,r,m}$ to itself for $a_{p,r,m}$ times. In other words, $a_{p,r,m} \mathbb{I}_{G:p,r,m}$ is the short hand notation for

$$a_{p,r,m} \mathbb{I}_{G:p,r,m} = \sum_{i \in \{1, \dots, a_{p,r,m}\}}^{(G)} \mathbb{I}_{G:p,r,m}$$

where the summation is the group operation.

B. The Group Mutual Information

We will need to define information theoretic quantities in relation to groups. Define

$$\mathcal{Q}(G) = \{(p, r) | p \in \mathcal{P}(G), r \in \mathcal{R}_p(G)\} \quad (33)$$

We denote vectors $\hat{\theta}$, w and θ , whose components are indexed by $(p, r) \in \mathcal{Q}(G)$, by $(\hat{\theta}_{p,r})_{(p,r) \in \mathcal{Q}(G)}$, $(w_{p,r})_{(p,r) \in \mathcal{Q}(G)}$ and $(\theta_{p,r})_{(p,r) \in \mathcal{Q}(G)}$ respectively. Let w be a probability distribution on $\mathcal{Q}(G)$.

For $\hat{\theta}$, define a vector

$$\boldsymbol{\theta}(\hat{\theta}) = \left(\min_{\substack{(q,s) \in \mathcal{Q}(G) \\ q=p}} |r-s|^+ + \hat{\theta}_{q,s} \right)_{(p,r) \in \mathcal{Q}(G)}$$

and let

$$\Theta = \left\{ \boldsymbol{\theta}(\hat{\theta}) \mid (\hat{\theta}_{q,s})_{(q,s) \in \mathcal{Q}(G)} : 0 \leq \hat{\theta}_{q,s} \leq s \right\}$$

This set corresponds to a collection of subgroups of G which appear in the expression for the achievable rates. In other words, certain subgroups of the group become important in the achievable rate region when we are confined to use Abelian group codes. For $\theta \in \Theta$, define

$$\omega_\theta = \frac{\sum_{(p,r) \in \mathcal{Q}(G)} \theta_{p,r} w_{p,r} \log p}{\sum_{(p,r) \in \mathcal{Q}(G)} r w_{p,r} \log p}$$

and let H_θ be a subgroup of G defined as

$$H_\theta = \bigoplus_{(p,r,m) \in \mathcal{G}(G)} p^{\theta_{p,r}} \mathbb{Z}_{p^r}^{(m)}. \quad (34)$$

We give an example in the sequel.

Let X and Y be two random variables with X taking values over G and let $[X]_\theta = X + H_\theta$ be the random variable taking values from the cosets of H_θ in G that contains X . Let w be a probability distribution on $\mathcal{Q}(G)$.

We define the source coding group mutual information between X and Y as

$$S_w^G(X; Y) = H(X) - \log |G| + \max_{\substack{\theta \in \Theta \\ \theta \neq \mathbf{0}}} \frac{1}{\omega_\theta} [\log |G : H_\theta| - H([X]_\theta | Y)]$$

where $\mathbf{0}$ is a vector whose components are indexed by $(p, r) \in \mathcal{Q}(G)$ and whose $(p, r)^{\text{th}}$ component is equal to 0, and $G : H_\theta$ is the quotient group. We define the channel coding group mutual information between X and Y as

$$C_w^G(X; Y) = H(X) - \log |G| + \min_{\substack{\theta \in \Theta \\ \theta \neq \mathbf{r}}} \frac{1}{1 - \omega_\theta} [\log |H_\theta| - H(X | [X]_\theta, Y)] \quad (35)$$

where \mathbf{r} is a vector whose components are indexed by $(p, r) \in \mathcal{Q}(G)$ and whose $(p, r)^{\text{th}}$ component is equal to r .

For example, let $G = \mathbb{Z}_2 \times \mathbb{Z}_8 \times \mathbb{Z}_3$. In this case, we have $\mathcal{P}(G) = \{2, 3\}$, $\mathcal{R}_2(G) = \{1, 3\}$, $\mathcal{R}_3(G) = \{1\}$ and $\mathcal{Q}(G) = \{(2, 1), (2, 3), (3, 1)\}$. The vectors w , $\hat{\theta}$ and θ are represented by $w = (w_{2,1}, w_{2,3}, w_{3,1})$, $\hat{\theta} = (\hat{\theta}_{2,1}, \hat{\theta}_{2,3}, \hat{\theta}_{3,1})$ and $\theta = (\theta_{2,1}, \theta_{2,3}, \theta_{3,1})$ and the function $\boldsymbol{\theta}(\cdot)$ is given by

$$\boldsymbol{\theta}(\hat{\theta}) = \left(\min(\hat{\theta}_{2,1}, \hat{\theta}_{2,3}), \min(2 + \hat{\theta}_{2,1}, \hat{\theta}_{2,3}), \hat{\theta}_{3,1} \right)$$

The set Θ turns out to be equal to

$$\Theta = \left\{ (0,0,0), (0,0,1), (0,1,0), (0,1,1), (0,2,0), (0,2,1), (1,1,0), (1,1,1), (1,2,0), (1,2,1), (1,3,0), (1,3,1) \right\}$$

and we have $\mathbf{0} = (0, 0, 0)$ and $\mathbf{r} = (1, 3, 1)$. For $\theta = (1, 1, 0)$, we have $\omega_\theta = \frac{w_{2,1} + w_{2,3}}{w_{2,1} + 3w_{2,3} + w_{3,1} \log 3}$ and $H_\theta = 0 \times 2\mathbb{Z}_8 \times \mathbb{Z}_3$. so that the random variable $[X]_\theta$ takes values from the set of cosets $\left\{ \{0\} \times 2\mathbb{Z}_8 \times \mathbb{Z}_3, \{0\} \times (1 +$

$2\mathbb{Z}_8 \times \mathbb{Z}_3, \{1\} \times 2\mathbb{Z}_8 \times \mathbb{Z}_3, \{1\} \times (1 + 2\mathbb{Z}_8) \times \mathbb{Z}_3$. Furthermore, for this choice of θ , we have $|H_\theta| = 4$ and $|G : H_\theta| = 12$.

When G is cyclic, i.e., $G = \mathbb{Z}_{p^r}$, then $w = 1$ and it can be shown that

$$S_w^G(X; Y) = H(X) - \min_{1 \leq \theta \leq r} \frac{r}{\theta} H([X]_\theta | Y), \quad C_w^G(X; Y) = H(X) - \max_{0 \leq \theta \leq (r-1)} \frac{r}{r-\theta} H(X|[X]_\theta, Y),$$

where $H_\theta = p^\theta \mathbb{Z}_{p^r}$. When G is a primary field, i.e., $G = \mathbb{Z}_p$, then it follows that

$$S_w^G(X; Y) = I(X; Y) = C_w^G(X; Y)$$

C. Managing interference seen by one receiver using PCC built over Abelian groups

In this section, we employ PCC built over Abelian groups to manage interference seen by only receiver 1. As the reader might have guessed, receiver 1 decodes the group sum of codewords chosen by receivers 2 and 3. In the following, we characterize an achievable rate region using codes built over groups.

Definition 11: Let $\mathbb{D}_g(\underline{\tau})$ denote the collection of pairs consisting of a distribution $p_{QU_2U_3\underline{XY}}$ 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 an Abelian group G , and a distribution w on $\mathcal{Q}(G)$ satisfying the following conditions: (i) $p_{\underline{Y}|\underline{X}} = W_{\underline{Y}|\underline{X}}$, (ii) $X_1, (U_2, X_2)$ and (U_3, X_3) are conditionally mutually independent given Q and (iii) $\mathbb{E}\{\kappa_j(X_j)\} \leq \tau_j : j \in [3]$ and (iv) $I(X_j; Y_j | U_j) + C_w^G(U_j; Y_j) - S_w^G(U_j; 0) \geq 0$ for $j = 2, 3$. For $(p_{QU_2U_3\underline{XY}}, w) \in \mathbb{D}_g(\underline{\tau})$, let $\alpha_g^{3-1}(p_{QU_2U_3\underline{XY}}, w)$ be defined as the set of rate triples $(R_1, R_2, R_3) \in [0, \infty)^3$ that satisfy

$$\begin{aligned} R_1 &< I(X_1; Y_1 | QZ) - H(Z|Q) + \min\{H(Z|Q), H(U_j|Q) + C_w^G(Z; Y_1|Q) - S_w^G(U_j; 0|Q) : j = 2, 3\} \\ R_j &< I(X_j; Y_j | QU_j) + C_w^G(U_j; Y_j|Q) \quad : j = 2, 3, \\ R_1 + R_j &< I(X_1; Y_1 | QZ) + C_w^G(Z; Y_1|Q) + H(U_j|Q) - H(Z|Q) + I(X_j; Y_j | QU_j) \\ &\quad + \min\{0, C_w^G(U_j; Y_j|Q) - S_w^G(U_j; 0|Q)\} \quad : j = 2, 3, \end{aligned}$$

where $Z = U_2 \oplus U_3$, and

$$\alpha_g^{3-1}(\underline{\tau}) = \text{cocl} \left(\bigcup_{(p_{QU_2U_3\underline{XY}}, w) \in \mathbb{D}_g(\underline{\tau})} \alpha_g^{3-1}(p_{\underline{XY}}) \right).$$

Theorem 8: For 3-IC $(\underline{\mathcal{X}}, \underline{\mathcal{Y}}, W_{\underline{Y}|\underline{X}}, \underline{\kappa})$, the set $\alpha_g^{3-1}(\underline{\tau})$ is achievable, i.e., $\alpha_g^{3-1}(\underline{\tau}) \subseteq \mathbb{C}(\underline{\tau})$.

We provide an illustration of the main arguments of the proof without giving complete details. In view of our detailed proof of theorem 5, the interested reader can fill in the details. We begin with an alternate characterization of $\alpha_g^{3-1}(p_{\underline{XY}})$ in terms of the parameters of the code.

Definition 12: Consider $(p_{QU_2U_3\underline{XY}}, w) \in \mathbb{D}_g(\underline{\tau})$ and let $G := \mathcal{U}_2 = \mathcal{U}_3$. Let $\tilde{\alpha}_g^{3-1}(p_{QU_2U_3\underline{XY}}, w)$ be defined as the set of rate triples $(R_1, R_2, R_3) \in [0, \infty)^3$ for which $\bigcup_{\delta > 0} \tilde{\mathcal{S}}(\underline{R}, p_{QU_2U_3\underline{XY}}, w, \delta)$ is non-empty, where $\tilde{\mathcal{S}}(\underline{R}, p_{QU_2U_3\underline{XY}}, w, \delta)$ is defined as the collection of vectors $(S_2, T_2, K_2, L_2, S_3, T_3, K_3, L_3, R_g) \in [0, \infty)^9$ that

satisfy

$$\begin{aligned}
R_j &= T_j + L_j, \quad K_j > \delta, \quad (S_j - T_j) > \log |G| - H(U_j|Q) + \delta, \\
(S_j - T_j) + K_j &> \log |G| + H(X_j|Q) - H(U_j, X_j|Q) + \delta, \quad R_g > S_j + \delta \\
S_j &> S_w^G(U_j; 0|Q) + \log |G| - H(U_j|Q) + \delta, \\
T_j > \delta, \quad L_j > \delta, \quad K_j + L_j &< I(X_j; Y_j, U_j|Q) - \delta, \quad S_j < \log |G| + C_w^G(U_j; X_j Y_j|Q) - H(U_j|Q) - \delta, \\
S_j + K_j + L_j &< \log |G| + I(X_j; Y_j U_j|Q) + C_w^G(U_j; Y_j|Q) - H(U_j|Q) - \delta, \quad R_1 < I(X_1; Y_1, Z|Q) - \delta \\
R_1 + R_g &< \log |G| + I(X_1; Y_1|ZQ) + C_w^G(Z; Y_1|Q) - H(Z|Q) - \delta
\end{aligned}$$

for $j = 2, 3$, where $Z = U_2 \oplus U_3$.

Lemma 7: $\tilde{\alpha}_g^{3-1}(p_{QU_2U_3XY}, w) = \alpha_g^{3-1}(p_{QU_2U_3XY}, w)$.

Proof: The proof follows from Fourier-Motzkin elimination. ■

Having obtained the parameters of the codes, we now describe the coding technique. Choose the parameters $(R_1, S_2, T_2, K_2, L_2, S_3, T_3, K_3, L_3, R_g) \in [0, \infty)^{10}$. The coding technique is exactly the same as that considered in the case of finite fields and is given in the proof of Theorem 5, The main exception is that the PCCs are built on the abelian group G . Instead of constructing vector spaces of \mathcal{F}^n , we construct subgroups of G^n . The cloud center codebook λ_j of user j is characterized as follows. Let

$$J_2 = \bigoplus_{(p,r) \in \mathcal{Q}(G)} \mathbb{Z}_{p^r}^{s_2 w_{p,r}}, \quad J_3 = \bigoplus_{(p,r) \in \mathcal{Q}(G)} \mathbb{Z}_{p^r}^{s_3 w_{p,r}},$$

for two positive integers s_2 and s_3 .

$$J = \bigoplus_{(p,r) \in \mathcal{Q}(G)} \mathbb{Z}_{p^r}^{s w_{p,r}}$$

Note that $J_j \leq J$ for $j = 2, 3$. Let ϕ be a homomorphism from J into G^n . Let ϕ_j be the restriction of ϕ to J_j for $j = 2, 3$. It is shown in [26, Equation 11] that ϕ has the following representation

$$\phi(a) = \bigoplus_{(p,r,m) \in \mathcal{G}(G^n)} \bigoplus_{(q,s,l) \in \mathcal{G}(J)}^{(\mathbb{Z}_{p^r})} a_{(q,s,l) \rightarrow (p,r,m)} g_{(q,s,l) \rightarrow (p,r,m)}$$

where $g_{(q,s,l) \rightarrow (p,r,m)} = 0$ for $p \neq q$ and $g_{(q,s,l) \rightarrow (p,r,m)}$ is uniformly distributed over $p^{|r-s|+} \mathbb{Z}_{p^r}$ for $p = q$. The code λ_j is given by $\phi_j(J_j) \oplus b_j^n$, where b_j^n is a bias vector in G^n . For a given choice of w, R_g, S_2 and S_3 , choose s_2, s_3 and s such that

$$s_2 = \frac{nS_2}{\sum_{(p,r) \in \mathcal{Q}(G)} r w_{p,r} \log p}, \quad s_3 = \frac{nS_3}{\sum_{(p,r) \in \mathcal{Q}(G)} r w_{p,r} \log p}, \quad s = \frac{nR_g}{\sum_{(p,r) \in \mathcal{Q}(G)} r w_{p,r} \log p}$$

Note that

$$\frac{1}{n} \log |J| = \frac{s}{n} \sum_{(p,r) \in \mathcal{Q}(G)} r w_{p,r} \log p,$$

and $|J_j|$ can be expressed in terms of s_j similarly. In summary, we have

$$\frac{1}{n} \log J = R_g, \quad \frac{1}{n} \log J_j = S_j : j = 2, 3.$$

The binning functions i_j are defined analogously: $i_j : J_j \rightarrow |G|^{t_j}$, where $t_j = nT_j$, for $j = 2, 3$. The encoding and decoding operations are defined analogously. This implies that $|\mathcal{M}_1| = 2^{nR_1}$, $|\mathcal{M}_{j1}| = |G|^{t_j}$ for $j = 2, 3$. The homomorphism and the bias vectors are chosen independently and with uniform probability over their ranges.

For any $a, \tilde{a} \in J$, and $(q, s, l) \in \mathcal{G}(J)$, let $\hat{\theta}_{q,s,l} \in \{1, 2, \dots, s\}$ be such that

$$\tilde{a}_{q,s,l} - a_{q,s,l} \in q^{\hat{\theta}_{q,s,l}} \mathbb{Z}_{q^s} \setminus q^{\hat{\theta}_{q,s,l}+1} \mathbb{Z}_{q^s}.$$

and any $(p, r) \in \mathcal{Q}(G)$, define

$$\theta_{p,r}(a, \tilde{a}) = \min_{(p,s,l) \in \mathcal{G}(J)} |r - s|^+ + \hat{\theta}_{q,s,l}.$$

Define for any $a \in J$, and any $\theta = (\theta_{p,r})_{(p,r) \in \mathcal{Q}(G)}$,

$$T_{J,\theta}(a) = \{\tilde{a} \in J : \forall (p, r) \in \mathcal{Q}(G), \theta_{p,r}(a, \tilde{a}) = \theta_{p,r}\}.$$

It can be shown that the expected value of the probability of all the error events over the ensemble approach zero as the block length increases if the parameters of the code belong to $\tilde{\alpha}_g^{3-1}(p_{QU_2U_3XY}, w)$. For conciseness, in the following, we give proofs of the elements in this argument that are new as compared to the analysis done in the case of fields.

Upper bound on $P(\epsilon_{l_2})$:- Given a message m_2 that indexes the bin in the cloud center codebook, define

$$\psi_2(m_{21}) = \sum_{a \in J_2} \sum_{u_2 \in T_{2\eta}(U_2)} \mathbb{1}_{\{\phi(a) + B_2 = u_2, i_2(a) = m_2\}}$$

We have

$$\begin{aligned} \mathbb{E} \{\psi_2(m_{21})\} &= \sum_{a \in J_2} \sum_{u_2 \in T_{2\eta}(U_2|Q)} P(\phi(a) + B_2 = u_2, I_2(a) = m_2) \\ &= \sum_{a \in J_2} \sum_{u_2 \in T_{2\eta}(U_2|Q)} \frac{1}{|G|^n} \cdot \frac{1}{|G|^{t_2}} = \frac{|J_2| \cdot |T_{2\eta}(U_2|Q)|}{|G|^n \cdot G^{t_2}} \end{aligned}$$

and

$$\begin{aligned} \mathbb{E} \{\psi_2(m_{21})^2\} &= \sum_{a, \tilde{a} \in J_2} \sum_{u_2, \tilde{u}_2 \in T_{2\eta}(U_2|Q)} P(\phi(a) + B_2 = u_2, \phi(\tilde{a}) + B_2 = \tilde{u}_2, I_2(a) = m_2, I_2(\tilde{a}) = m_2) \\ &= \sum_{\theta \in \Theta} \sum_{a \in J_2} \sum_{\tilde{a} \in T_{J_2, \theta}(a)} \sum_{u_2 \in T_{2\eta}^n(U_2|Q)} \sum_{\substack{\tilde{u}_2 \in T_{2\eta}^n(U_2|Q) \\ \tilde{u}_2 \in u_2 + H_\theta^n}} \frac{1}{|G|^n} \cdot \frac{1}{|H_\theta|^n} \cdot P(I_2(a) = m_2, I_2(\tilde{a}) = m_2) \\ &= \sum_{a \in J_2} \sum_{u_2 \in T_{2\eta}^n(U_2|Q)} \frac{1}{|G|^n} \cdot \frac{1}{G^{t_2}} \\ &+ \sum_{\substack{\theta \in \Theta \\ \theta \neq \mathbf{r}}} \sum_{a \in J_2} \sum_{\tilde{a} \in T_{J_2, \theta}(a)} \sum_{u_2 \in T_{2\eta}^n(U_2|Q)} \sum_{\substack{\tilde{u}_2 \in T_{2\eta}^n(U_2|Q) \\ \tilde{u}_2 \in u_2 + H_\theta^n}} \frac{1}{|G|^n} \cdot \frac{1}{|H_\theta|^n} \cdot \frac{1}{G^{t_2}} \\ &\leq \frac{|J_2| \cdot |T_{2\eta}^n(U_2|Q)|}{|G|^n \cdot G^{t_2}} + \sum_{\substack{\theta \in \Theta \\ \theta \neq \mathbf{r}}} \sum_{a \in J_2} \frac{|T_{J_2, \theta}(a)| \cdot |T_{2\eta}^n(U_2|Q)| \cdot |T_{2\eta}^n(U_2|Q) \cap (u_2 + H_\theta^n)|}{|G|^n \cdot |H_\theta|^n \cdot G^{t_2}} \end{aligned}$$

where, \mathbf{r} is a vector whose components are indexed by $(p, r) \in \mathcal{Q}(G)$ and whose $(p, r)^{\text{th}}$ component is equal to r . Using [26, Lemma IX.2], we get

$$\text{Var} \{ \psi_2(m_{21})^2 \} \leq \frac{|J_2| \cdot |T_{2\eta_2}^n(U_2|Q)|}{|G|^n \cdot G^{t_2}} + \sum_{\substack{\theta \in \Theta \\ \theta \neq \mathbf{0}, \theta \neq \mathbf{r}}} \sum_{a \in J_2} \frac{|T_{J_2, \theta}(a)| \cdot 2^{n[H(U_2|Q)+\eta]} 2^{n[H(U_2|[U_2]_\theta Q)+\eta]}}{|G|^n \cdot |H_\theta|^n \cdot G^{t_2}}$$

Here, $\mathbf{0}$ is a vector whose components are indexed by $(p, r) \in \mathcal{Q}(G)$ and whose $(p, r)^{\text{th}}$ component is equal to 0.

We have

$$\frac{\text{Var} \{ \psi_2(m_{21})^2 \}}{\mathbb{E}^2(\psi_2(m_{21}))} \leq \frac{|G|^n \cdot G^{t_2}}{|J_2| \cdot 2^{n[H(U_2|Q)-\eta]}} + \sum_{\substack{\theta \in \Theta \\ \theta \neq \mathbf{0}, \theta \neq \mathbf{r}}} \sum_{a \in J_2} \frac{|G|^n \cdot |T_{J_2, \theta}(a)| \cdot 2^{n[H(U_2|[U_2]_\theta Q)+\eta]}}{|H_\theta|^n \cdot 2^{n[H(U_2|Q)-\eta]}}$$

Note that $|J_2| = 2^{n(S_2)}$, $\frac{|G|^n}{|H_\theta|^n} = |G : H_\theta|^n$, and using [26, Lemma IX.2] we have $|T_{J_2, \theta}(a)| \leq 2^{n(1-w_\theta)(S_2 \log |G| + \eta_3)}$.

In order for the probability of error to go to zero, we require

$$(S_2 - T_2) > \log |G| - H(U_2|Q)$$

$$S_2 > \max_{\substack{\theta \in \Theta \\ \theta \neq \mathbf{0}}} \frac{1}{\omega_\theta} [\log |G : H_\theta| - H([U_2]_\theta|Q)],$$

which is equivalent to $(S_2 - T_2) > \log |G| - H(U_2|Q)$, and $S_2 > S_w^G(U_2; 0|Q)$.

Upper bound on $P((\epsilon_{11} \cup \epsilon_{12} \cup \epsilon_{13} \cup \epsilon_2 \cup \epsilon_3)^c \cap \epsilon_{41})$: This probability can be decomposed into two parts: (i) the first, P_1 , is the probability of the event that X_1^n and $U_2^n + U_3^n$ are both decoded incorrectly and (ii) the second, P_2 , is the probability of the event that X_1^n is decoded incorrectly but $U_2^n + U_3^n$ is decoded correctly. A vanishing upper bound on the second part can be obtained in a way that is analogous to the case of fields if $R_1 < I(X_1; Y_1|Z)$. In the following we provide an upper bound only on the first part. For simplicity, let us assume that Q is trivial.

$$P_1 \leq \frac{1}{|\mathcal{M}_1|} \sum_{m_1} \sum_{x_1 \in T_{2\eta_2}(X_1)} \mathbb{1}_{\{X_1^n(m_1)=x_1\}} \frac{1}{|G|^{t_2}} \sum_{m_{21}} \frac{1}{|G|^{t_3}} \sum_{m_{31}} \sum_{u_2 \in T_{2\eta_2}^n(U_2)} \frac{2}{\mathbb{E}\{\psi_2(m_{21})\}} \sum_{u_3 \in T_{2\eta_2}^n(U_3)} \frac{2}{\mathbb{E}\{\psi_3(m_{31})\}}$$

$$\sum_{y_1 \in \mathcal{Y}_1^n} p_{Y_1^n|X_1, U_2, U_3}^n(y_1|x_1, u_2, u_3) 2^{-2n\eta_4} \sum_{\tilde{m}_1 \neq m_1} \sum_{a \in J_2, b \in J_3} \mathbb{1}_{\{\phi(a)+B_2=u_2, \phi(b)+B_3=u_3, i_2(a)=m_{21}, i_3(b)=m_{31}\}}$$

$$\sum_{\substack{(\tilde{x}_1, \tilde{z}) \in T_{4\eta_4}(X_1, Z|y_1) \\ \tilde{z} \neq z}} \mathbb{1}_{\{\tilde{x}_1 = X_1^n(\tilde{m}_1)\}} \mathbb{1}_{\{\exists \tilde{a}, \tilde{b} \in J : \phi(\tilde{a}) + B_2 + \phi(\tilde{b}) + B_3 = \tilde{z}\}}$$

Using the condition $z \neq \tilde{z}$, note that the event $\{\exists \tilde{a}, \tilde{b} \in J : \phi(\tilde{a}) + B_2 + \phi(\tilde{b}) + B_3 = \tilde{z}\}$ is equal to the event $\{\exists \tilde{c} \in J : c \neq (a+b), \phi(\tilde{c}) + B_2 + B_3 = \tilde{z}\}$. Therefore, using the union bound, and decomposing the set $J \setminus (a+b)$

into smaller parts, we get

$$\begin{aligned}
\mathbb{E}\{P_1\} &\leq \frac{1}{|\mathcal{M}_1|} \sum_{m_1} \sum_{x_1 \in T_{2\eta_2}(X_1)} \frac{2^{-2n\eta_4}}{|T_{2\eta_2}(X_1)| |G|^{t_2} |G|^{t_3}} \sum_{m_2, m_3} \sum_{\substack{u_2 \in T_{2\eta_2}^n(U_2) \\ u_3 \in T_{2\eta_2}^n(U_3)}} \frac{4}{\mathbb{E}(\psi_2(m_{21})) \mathbb{E}(\psi_3(m_{31}))} \\
&\quad \sum_{y_1 \in \mathcal{Y}_1^n} p_{Y_1|X_1, U_2, U_3}^n(y_1|x_1, u_2, u_3) \sum_{\tilde{m}_1 \neq m_1} \sum_{a \in J_2, b \in J_3} \sum_{\substack{\theta \in \Theta \\ \theta \neq \mathbf{r}}} \sum_{\substack{(\tilde{x}_1, \tilde{z}) \in T_{2\eta_4}(X_1, Z|y_1) \\ \tilde{z} \neq z}} \sum_{\tilde{c} \in T_\theta(a+b)} \frac{1}{|G|^n |G|^{t_2} |G|^{t_3}} \\
&\quad \cdot \frac{1}{|T_{2\eta_2}(X_1)|} \cdot P(\phi(a+b) + B_2 + B_3 = u_2 + u_3, \phi(\tilde{c}) + B_2 + B_3 = \tilde{z}) \\
&\leq \sum_{\substack{\theta \in \Theta \\ \theta \neq \mathbf{r}}} \sum_{a \in J_2} \sum_{b \in J_3} \frac{2^{-2\eta_4} 2^{nR_1} \cdot 2^{n[H(X_1|Z, Y_1) + \eta]} \cdot 2^{n[H(Z|[Z]_\theta Y_1) + \eta]} \cdot |T_{J, \theta}(a+b)|}{|J_2| |J_3| 2^{n[H(X_1) - \eta]} \cdot |H_\theta^n|}
\end{aligned}$$

Using [26, Lemma IX.2], note that $|T_{J, \theta}(a+b)| \leq 2^{n(1-\omega_\theta)R_g}$. Therefore, in order for the probability of error to go to zero, it suffices to have

$$R_1 + (1 - \omega_\theta)R_g < I(X_1; Y_1|Z) + \log |H_\theta| - H(Z|[Z]_\theta Y_1)$$

for $\theta \neq \mathbf{r}$. For optimum weights $\{w_{p,r}\}_{(p,r) \in \mathcal{Q}(G)}$, the condition $R_1 + R_g < I(X_1; Y_1|Z) + C_w^G(Z; Y_1) + \log |G| - H(Z)$ implies

$$\begin{aligned}
R_g &< (I(X_1; Y_1|Z) - R_1) + \min_{\substack{\theta \in \Theta \\ \theta \neq \mathbf{r}}} \frac{1}{1 - \omega_\theta} [\log |H_\theta| - H(Z|[Z]_\theta Y_1)] \\
&\stackrel{(a)}{=} \min_{\substack{\theta \in \Theta \\ \theta \neq \mathbf{r}}} \frac{1}{1 - \omega_\theta} [I(X_1; Y_1|Z) - R_1] + \min_{\substack{\theta \in \Theta \\ \theta \neq \mathbf{r}}} \frac{1}{1 - \omega_\theta} [\log |H_\theta| - H(Z|[Z]_\theta Y_1)] \\
&\leq \min_{\substack{\theta \in \Theta \\ \theta \neq \mathbf{r}}} \frac{1}{1 - \omega_\theta} [I(X_1; Y_1|Z) - R_1 + \log |H_\theta| - H(Z|[Z]_\theta Y_1)]
\end{aligned}$$

which is the desired condition. In the above equations, (a) follows since the maximum of $1 - \omega_\theta$ is attained for $\theta = \mathbf{0}$ and is equal to 1.

Similarly, one can show that the probability of decoding error at decoder 2 and 3 can be made to go to zero for all sufficiently large n if the code parameters are chosen accordingly.

We have thus proved the bounds provided in definition (12) suffice to drive the probability of incorrect decoding exponentially down to 0. We now illustrate the need to build codes over appropriate algebraic objects to enable interference management. In other words, we provide an example where codes built over groups outperform unstructured codes as well as codes built over finite fields.²¹

Example 7: Consider a quaternary 3-to-1 IC with input and output alphabets $\mathcal{X}_j = \mathcal{Y}_j = \mathbb{Z}_4 = \{0, 1, 2, 3\}$ being the Abelian group of cardinality 4. Let \oplus_4 denote the group operation, i.e., addition mod-4 in \mathbb{Z}_4 . The channel transition probabilities are described through the relation $Y_1 = X_1 \oplus_4 X_2 \oplus_4 X_3 \oplus_4 N_1$, $Y_j = X_j \oplus_4 N_j$

²¹While, we do not provide a proof of the statement that codes built over groups outperform PCC built over finite fields, this can be recognized through standard arguments.

for $j = 2, 3$ such that (i) N_1, N_2, N_3 are independent random variables taking values in \mathbb{Z}_4 with

$$P(N_j = n_j) = \begin{cases} 1 - \delta_j & \text{if } n_j = 0 \\ \frac{\delta_j}{3} & \text{otherwise} \end{cases} \quad \text{for } j = 1, 2, 3.$$

It can be verified that the channel transition probabilities are given by $W_{\underline{Y}|\underline{X}}(y|\underline{x}) = QSC_{\delta_1}(y_1|x_1 \oplus_4 x_2 \oplus_4 x_3)QSC_{\delta_2}(y_2|x_2)QSC_{\delta_3}(y_3|x_3)$, where for any $\eta \in [0, 1]$,

$$QSC_{\eta}(a|b) := \begin{cases} 1 - \eta & \text{if } a \ominus_4 b = 0 \\ \frac{\eta}{3} & \text{otherwise} \end{cases}$$

Inputs X_2, X_3 of users 2 and 3 are not costed, i.e., $\kappa_j(x_j) = 0$ for $j = 2, 3$ and any $x_j \in \mathcal{X}_j$, whereas $\kappa_1(x_1) = 1$ if $x_1 \in \{1, 2, 3\}$ and $\kappa_1(0) = 0$. User 1's input is constrained to a average cost of τ per symbol.

The reader will recognize that the 3-to-1 IC described in example 7 is analogous to that in example 1 with the binary field replaced by Abelian group \mathbb{Z}_4 . Our objective here is to illustrate the utility of building codes over Abelian groups. In particular, we prove codes over Abelian groups outperform unstructured codes and coset codes over finite fields.

For simplicity, let us henceforth assume $\delta_2 = \delta_3 = \delta$. Since users 2 and 3 enjoy interference free point-to-point links, they could communicate at their respective capacities even using PCC built on \mathbb{Z}_4 . This can be seen by choosing $U_j = X_j$ and putting a uniform distribution on X_j for $j = 2, 3$, and evaluating the corresponding group capacity as follows:

$$\begin{aligned} C_w^G(X_j; Y_j) &= \min\{I(X_j; Y_j), 2I(X_j; Y_j|[X_j]_1)\} \\ &= \min\{2 - h_b(\delta) - \delta \log_2(3), 2 + 2h_b(2\delta/3) - 2h_b(\delta) - 2\delta \log_2(3)\} = 2 - h_b(\delta) - \delta \log_2(3), \end{aligned} \quad (36)$$

where the last equality follows from the concavity of entropy. In the sequel, we constrain users 2 and 3 to achieve their respective PTP capacities. In particular, we let X_j to be uniformly distributed over \mathcal{X}_j for $j = 2, 3$.

Clearly, user 1 can achieve a rate not greater than $C^* := \sup_{p_{X_1}: p_{X_1}(1) \leq \tau} I(X_1; Y_1|X_2 \oplus_4 X_3)$. In the sequel, we state the conditions under which $R_1 < C^*$. Our approach is similar to that in section IV.

Lemma 8: Consider the 3-to-1 IC described in example 7 with $\delta_2 = \delta_3 = \delta \in (0, \frac{1}{4})$, $\delta_1 \in (0, \frac{1}{4})$ and $\tau < \frac{3}{4}$. If δ_1, τ and δ are such that

$$C^* + 2(2 - h_b(\delta) - \delta \log_2 3) > 2 - h_b(\delta_1) - \delta_1 \log_2 3, \quad (37)$$

then the rate triple $(C^*, 2 - h_b(\delta) - \delta \log_2 3, 2 - h_b(\delta) - \delta \log_2 3) \notin \alpha_u(\tau, 0, 0)$.

Proof: We first note that for any $p_{QU_2U_3XY} \in \mathbb{D}_u(\tau, 0, 0)$ with $H(X_j|Q, U_j) = 0$ for $j = 2, 3$, we have $R_1 + R_2 + R_3 < I(\underline{X}; Y_1) \leq \sup_{p_{X_1} p_{X_2} p_{X_3}} I(\underline{X}; Y_1)$. This follows from substituting the corresponding quantities in (2). It can be easily verified that $\sup_{p_{X_1} p_{X_2} p_{X_3}} I(\underline{X}; Y_1) = 2 - h_b(\delta_1) - \delta_1 \log_2 3$ which is achieved for all those distributions $p_{X_1} p_{X_2} p_{X_3}$ that ensure Y_1 is uniformly distributed. Condition (37) therefore implies $(C^*, 2 - h_b(\delta) - \delta \log_2 3, 2 - h_b(\delta) - \delta \log_2 3) \notin \alpha_u(p_{QU_2U_3XY})$ if $H(X_j|Q, U_j) = 0$ for $j = 2, 3$. Hence either $H(X_2|Q, U_2) > 0$ or $H(X_3|Q, U_3) > 0$. Assume $H(X_j|Q, U_j) > 0$ and $\{j, \hat{j}\} = \{2, 3\}$. By the conditional independence of (U_2, X_2)

and (U_3, X_3) given Q , we have $0 < H(X_j|Q, U_j) = H(X_j|Q, U_j, U_{\dot{j}}, X_{\dot{j}}) = H(X_j \oplus_4 X_{\dot{j}}|Q, U_j, U_{\dot{j}}, X_{\dot{j}}) = H(X_2 \oplus_4 X_3|Q, U_2, U_3, X_{\dot{j}}) \leq H(X_2 \oplus_4 X_3|Q, U_2, U_3)$.

We only need to prove $H(X_2 \oplus_4 X_3|Q, U_2, U_3) > 0$ implies $I(X_1; Y_1|Q, U_2, U_3) < C^*$. For this, we allude to the proof of fifth claim in appendix F. Therein, we have proved an analogous statement for example 4. The statement herein can be proved through an analogous sequence of steps and we let the reader fill in these details. \blacksquare

We now show that user 1 can achieve rate equal to C^* exploiting the fact that user 2 and 3 use group codes. We also derive the condition (37) in terms of parameters δ_1, τ, δ . Note that the channel between $X_2 \oplus_4 X_3$ and Y_1 is additive with noise given by $X_1 \oplus_4 N_1$. Let us choose $p_{X_1}(x_1) = \frac{\tau}{3}$ for $x_1 \in \{1, 2, 3\}$. The resulting distribution of $X_1 \oplus_4 N_1$ is given by $p_{X_1 \oplus_4 N_1}(a) = \beta/3$ for $a \in \{1, 2, 3\}$, where $\beta = \delta_1 + \tau - \frac{4\delta_1\tau}{3}$. Using concavity of entropy once again, we get

$$C_w^G(X_2 \oplus_4 X_3; Y) = \min\{2 - h_b(\beta) - \beta \log_2(3), 2 + 2h_b(2\beta/3) - 2h_b(\beta) - 2\beta \log_2(3)\} = 2 - h_b(\beta) - \beta \log_2(3). \quad (38)$$

Note that for $\delta_1 \in (0, \frac{1}{4})$ and $\tau < \frac{3}{4}$, using the fact that X_1 and N_1 are independent, we get $\beta \in (0, \frac{3}{4})$. Note also that $2 - h_b(\beta) - \beta \log_2(3)$ is monotone decreasing for $\beta \in (0, 3/4)$. Hence if $\beta \geq \delta$, the signal $X_2 \oplus_4 X_3$ can be decoded at decoder 1, and user 1 can communicate at the rate C^* . A simple calculation yields $C^* = h_b(\beta) + \beta \log_2 3 - h_b(\delta_1) - \delta_1 \log_2 3$. In summary, under the following conditions: (i) $\delta, \delta_1 \in (0, \frac{1}{4})$, (ii) $\tau < \frac{3}{4}$, (iii) equation (37), and (iv) $\beta \geq \delta$, all three users can achieve their respective capacities using PCCs built on \mathbb{Z}_4 , but the corresponding rates cannot be achieved using unstructured codes. It can be shown that there exists a non-empty set of parameters (δ, δ_1, τ) that satisfy these conditions. An example is given by $\delta = \frac{1}{8}, \delta_1 = \tau = \frac{3}{4} - \frac{\sqrt{30}}{8}$.

APPENDIX A

UPPER BOUND ON $P(\epsilon_{l_j})$

Recall

$$\phi_j(q^n, M_j) := \sum_{a^{sj} \in \mathcal{U}^{sj}} \sum_{b_{jX} \in c_{jX}} 1_{\{I_j(a^{sj}) = M_{j1}, (q^n, U_j^n(a^{sj}), X_j^n(M_{jX}, b_{jX})) \in T_{2n}(Q, U_j, X_j)\}}, \quad \mathcal{L}_j(n) := \frac{1}{2} \mathbb{E} \{ \phi_j(q^n, M_j) \}$$

and $\epsilon_{l_j} = \{ \phi_j(q^n, M_j) < \mathcal{L}_j(n) \}$. Employing Cheybshev's inequality, we have

$$P(\epsilon_{l_j}) = P(\phi_j(q^n, M_j) < \mathcal{L}_j(n)) \leq P(|\phi_j(q^n, M_j) - \mathbb{E}\{\phi_j(q^n, M_j)\}| \geq \frac{1}{2} \mathbb{E}\{\phi_j(q^n, M_j)\}) \leq \frac{4\text{Var}\{\phi_j(q^n, M_j)\}}{(\mathbb{E}\{\phi_j(q^n, M_j)\})^2}.$$

Note that $\text{Var}\{\phi_j(q^n, M_j)\} = \mathcal{T}_0 + \mathcal{T}_1 + \mathcal{T}_2 + \mathcal{T}_3 - \mathcal{T}_0^2$, where

$$\begin{aligned}
\mathcal{T}_0 &= \sum_{a^{sj} \in \mathcal{U}^{sj}} \sum_{b_{jX} \in c_{jX}} \sum_{\substack{(u_j^n, x_j^n) \in \\ T_{2\eta}(U_j, X_j | q^n)}} P \left(\begin{matrix} I_j(a^{sj})=M_{j1}, U_j^n(a^{sj})=u_j^n \\ X_j^n(M_{jX}, b_{jX})=x_j^n \end{matrix} \right) = \mathbb{E}\{\phi_j(q^n, M_j)\}, \\
\mathcal{T}_1 &= \sum_{a^{sj} \in \mathcal{U}^{sj}} \sum_{\substack{b_{jX}, \bar{b}_{jX} \in c_{jX} \\ b_{jX} \neq \bar{b}_{jX}}} \sum_{\substack{(u_j^n, x_j^n), (u_j^n, \bar{x}_j^n) \in \\ T_{2\eta}(U_j, X_j | q^n)}} P \left(\begin{matrix} I_j(a^{sj})=M_{j1}, X_j^n(M_{jX}, b_{jX})=x_j^n, \\ U_j^n(a^{sj})=u_j^n, X_j^n(M_{jX}, \bar{b}_{jX})=\bar{x}_j^n \end{matrix} \right), \\
\mathcal{T}_2 &= \sum_{\substack{a^{sj}, \bar{a}^{sj} \in \mathcal{U}^{sj} \\ a^{sj} \neq \bar{a}^{sj}}} \sum_{b_{jX} \in c_{jX}} \sum_{\substack{(u_j^n, x_j^n), (\bar{u}_j^n, \bar{x}_j^n) \in \\ T_{2\eta}(U_j, X_j | q^n)}} P \left(\begin{matrix} I_j(a^{sj})=M_{j1}, I_j(\bar{a}^{sj})=M_{j1}, U_j^n(a^{sj})=u_j^n, \\ X_j^n(M_{jX}, b_{jX})=x_j^n, U_j^n(\bar{a}^{sj})=\bar{u}_j^n \end{matrix} \right), \\
\mathcal{T}_3 &= \sum_{\substack{a^{sj}, \bar{a}^{sj} \in \mathcal{U}^{sj} \\ a^{sj} \neq \bar{a}^{sj}}} \sum_{\substack{b_{jX}, \bar{b}_{jX} \in c_{jX} \\ b_{jX} \neq \bar{b}_{jX}}} \sum_{\substack{(u_j^n, x_j^n), (\bar{u}_j^n, \bar{x}_j^n) \in \\ T_{2\eta}(U_j, X_j | q^n)}} P \left(\begin{matrix} I_j(a^{sj})=M_{j1}, X_j^n(M_{jX}, b_{jX})=x_j^n, U_j^n(a^{sj})=u_j^n, \\ I_j(\bar{a}^{sj})=M_{j1}, X_j^n(M_{jX}, \bar{b}_{jX})=\bar{x}_j^n, U_j^n(\bar{a}^{sj})=\bar{u}_j^n \end{matrix} \right).
\end{aligned} \tag{39}$$

The codewords of PCC Λ_j are pairwise independent [27, Theorem 6.2.1], and therefore

$$P \left(\begin{matrix} I_j(a^{sj})=M_{j1}, X_j^n(M_{jX}, b_{jX})=x_j^n, U_j^n(a^{sj})=u_j^n, \\ I_j(\bar{a}^{sj})=M_{j1}, X_j^n(M_{jX}, \bar{b}_{jX})=\bar{x}_j^n, U_j^n(\bar{a}^{sj})=\bar{u}_j^n \end{matrix} \right) = P \left(\begin{matrix} I_j(a^{sj})=M_{j1}, U_j^n(a^{sj})=u_j^n \\ X_j^n(M_{jX}, b_{jX})=x_j^n \end{matrix} \right) P \left(\begin{matrix} I_j(\bar{a}^{sj})=M_{j1}, U_j^n(\bar{a}^{sj})=\bar{u}_j^n, \\ X_j^n(M_{jX}, \bar{b}_{jX})=\bar{x}_j^n \end{matrix} \right).$$

It can be verified that $\mathcal{T}_3 \leq \mathcal{T}_0^2$, and therefore, $P(\epsilon_{1j}) \leq 4 \frac{\mathcal{T}_0 + \mathcal{T}_1 + \mathcal{T}_2}{\mathcal{T}_0^2}$. For sufficiently large n , we employ upper bounds on conditional probability and the number of conditional typical sequences to conclude

$$\begin{aligned}
\mathcal{T}_0 &\geq \frac{\exp\{-nH(X_j|Q) - 4n\eta\} |c_{jX}| |T_{2\eta}(U_j, X_j | q^n)|}{\theta^{t_j + n - s_j}} \\
\mathcal{T}_1 &\leq \frac{\exp\{-2nH(X_j|Q) + 8n\eta + nH(X_j|U_j, Q) + 8n\eta\} |c_{jX}| (|c_{jX}| - 1) |T_{2\eta}(U_j, X_j | q^n)|}{\theta^{t_j + n - s_j}} \\
\mathcal{T}_2 &\leq \frac{\exp\{-nH(X_j|Q) + 4n\eta + nH(U_j|X_j, Q) + 8n\eta\} |c_{jX}| |T_{2\eta}(U_j, X_j | q^n)|}{\theta^{2(t_j + n - s_j)}}.
\end{aligned} \tag{40}$$

For sufficiently large n , $\exp\{-4n\eta\} \leq \exp\{-nH(U_j, X_j|Q)\} |T_{2\eta}(U_j, X_j | q^n)| \leq \exp\{4n\eta\}$. Substituting $S_j = \frac{s_j \log \theta}{n}$, $T_j = \frac{t_j \log \theta}{n}$ and $|c_{jX}| = \exp\{nK_j\}$, it may be verified that, for sufficiently large n ,

$$\begin{aligned}
P(\epsilon_{1j}) &\leq 4 \exp\{-n[S_j - T_j + K_j - (\log \theta + H(X_j|Q) - H(U_j, X_j|Q)) - 8\eta]\} + \\
&\quad 4 \exp\{-n[S_j - T_j - (\log \theta - H(U_j|Q)) - 28\eta]\} + 4 \exp\{-n[K_j - 32\eta]\}.
\end{aligned}$$

Using the bounds on S_j, T_j and K_j as given in definition 12 in terms of δ , we have

$$P(\epsilon_{1j}) \leq 12 \exp\{-n(\delta - 32\eta)\} \tag{41}$$

for sufficiently large n . 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 (39), (40), we have for sufficiently large n ,

$$\begin{aligned}
\mathcal{L}_j(n) &= \frac{1}{2} \mathbb{E}\{\phi_j(q^n, M_j)\} = \frac{\mathcal{T}_0}{2} \geq \frac{\exp\{-nH(X_j|Q) - 4n\eta\} |c_{jX}| |T_{2\eta}(U_j, X_j | q^n)|}{2\theta^{t_j + n - s_j}} \\
&\geq \frac{1}{2} \exp\{n[S_j - T_j + K_j - (\log \theta + H(X_j|Q) - H(U_j, X_j|Q)) - 8\eta]\} \geq \frac{1}{2} \exp\{n[\delta - 8\eta]\} \tag{42}
\end{aligned}$$

where, as before, we have employed $S_j = \frac{s_j \log \theta}{n}$, $T_j = \frac{t_j \log \theta}{n}$ and $|c_{jX}| = \exp\{nK_j\}$, the lower bounds on $|T_{2\eta}(U_j, X_j | q^n)|$ and the definition of δ .

APPENDIX B

UPPER BOUNDS ON $P(\tilde{\epsilon}_1^c \cap \epsilon_2)$, $P((\tilde{\epsilon}_1 \cup \epsilon_2)^c \cap \epsilon_3)$

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

$$\epsilon_2 = \{(q^n, U_2^n(A^{s_2}), U_3^n(A^{s_3}), X_1^n(M_1), X_2^n(M_{2X}, B_{2X}), X_3^n(M_{3X}, B_{3X})) \notin T_{\eta_1}(Q, U_2, U_3, \underline{X})\}. \quad (43)$$

was defined in (8). 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

$$\Theta(q^n) := \left\{ \begin{array}{l} (u_2^n, u_3^n, \underline{x}^n) \in \mathcal{U}_2^n \times \mathcal{U}_3^n \times \mathcal{X}^n : (q^n, u_j^n, x_j^n) \in T_{2\eta}(Q, U_j, X_j) : j = 2, 3 \\ (q^n, x_1^n) \in T_{2\eta}(Q, X_1), (q^n, u_2^n, u_3^n, \underline{x}^n) \notin T_{\eta_1}(Q, U_2, U_3, \underline{X}) \end{array} \right\}.$$

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(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 \right. \\ &\quad \left. \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 \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_{b_{2X} \in \mathcal{C}_{2X}} \bigcup_{c_{3X} \in \mathcal{C}_{3X}} \left\{ 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} \right. \right. \\ &\quad \left. \left. \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 \right\} \right) \\ &\leq \sum_{\substack{(u_2^n, u_3^n, \underline{x}^n) \\ \in \Theta(q^n)}} \sum_{a^{s_2} \in \mathcal{U}_2^{s_2}} \sum_{a^{s_3} \in \mathcal{U}_3^{s_3}} \sum_{b_{2X} \in \mathcal{C}_{2X}} \sum_{c_{3X} \in \mathcal{C}_{3X}} P \left(\begin{array}{l} 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{array} \right) P \left(\begin{array}{l} A^{s_j}=a^{s_j} \\ B_{jX}=b_{jX} \\ : j=2,3 \end{array} \middle| \begin{array}{l} 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{array} \right) \\ &\leq \sum_{\substack{(u_2^n, u_3^n, \underline{x}^n) \\ \in \Theta(q^n)}} \sum_{a^{s_2} \in \mathcal{U}_2^{s_2}} \sum_{a^{s_3} \in \mathcal{U}_3^{s_3}} \sum_{b_{2X} \in \mathcal{C}_{2X}} \sum_{c_{3X} \in \mathcal{C}_{3X}} P \left(\begin{array}{l} 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{array} \right) \prod_{j=2}^3 P \left(\begin{array}{l} A^{s_j}=a^{s_j} \\ B_{jX}=b_{jX} \end{array} \middle| \begin{array}{l} 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{array} \right). \quad (44) \end{aligned}$$

Let us now evaluate a generic term in the above sum (44). Since the codebooks $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3, \Lambda_2, \Lambda_3$ are mutually independent, the probability of the event in question factors as

$$P \left(\begin{array}{l} U_j^n(a^{s_j})=u_j^n, X_j^n(M_{jX}, b_{jX})=x_j^n \\ I_j(a^{s_j})=M_{j1} : j=2,3, X_1^n(M_1)=x_1^n \end{array} \right) = P(X_1^n(M_1) = x_1^n) P \left(U_j^n(a^{s_j})=u_j^n, : j = 2, 3 \right) \prod_{j=2}^3 P(X_j^n(M_{jX}, b_{jX}) = x_j^n)$$

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

$$P \left(\begin{array}{l} U_j^n(a^{s_j})=u_j^n, X_j^n(M_{jX}, b_{jX})=x_j^n \\ I_j(a^{s_j})=M_{j1} : j=2,3, X_1^n(M_1)=x_1^n \end{array} \right) = P(U_j^n(a^{s_j}) = u_j^n : j = 2, 3) \frac{\prod_{j=1}^3 \prod_{t=1}^n p_{X_j|Q}(x_{jt}|q_t)}{\theta^{t_2+t_3}} \quad (45)$$

The following simple lemma enables us characterize $P(U_j^n(a^{s_j}) = u_j^n : j = 2, 3)$.

Lemma 9: Let $s_2, s_3, n \in \mathbb{N}$ be such that $s_2 \leq s_3$. Let $G_3^T := [G_2^T \ G_{3/2}^T] \in \mathcal{F}_\theta^{s_3 \times n}$ be a random matrix such that $G_2 \in \mathcal{F}_\theta^{s_2 \times n}$ and $B_2^n, B_3^n \in \mathcal{F}_\theta^n$ 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 $a^{s_j} \in \mathcal{F}_\theta^{s_j}$, let $U(a^{s_j}) := a^{s_j}G_j \oplus B_j^n$ be a random vector in the corresponding coset. Then $P(U_j^n(a^{s_j}) = u_j^n : j = 2, 3) = \frac{1}{\theta^{2n}}$.

Proof: The proof follows from a simple counting argument. It maybe verified that for every $g_3 \in \mathcal{F}_\theta^{s_3 \times n}$, there exists a unique pair of vectors $b_2^n, b_3^n \in \mathcal{F}_\theta^n$ such that $a^{s_j}g_j \oplus b_j^n = u_j^n$ for $j = 2, 3$. Therefore

$$|\{(g_3, b_2^n, b_3^n) \in \mathcal{F}_\theta^{s_3 \times n} \times \mathcal{F}_\theta^n \times \mathcal{F}_\theta^n : a^{s_j}g_j \oplus b_j^n = u_j^n \text{ for } j = 2, 3\}| = \theta^{ns_3}.$$

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

$$\frac{|\{(g_3, b_2^n, b_3^n) \in \mathcal{F}_\theta^{s_3 \times n} \times \mathcal{F}_\theta^n \times \mathcal{F}_\theta^n : a^{s_j}g_j \oplus b_j^n = u_j^n \text{ for } j = 2, 3\}|}{|\{(g_3, b_2^n, b_3^n) \in \mathcal{F}_\theta^{s_3 \times n} \times \mathcal{F}_\theta^n \times \mathcal{F}_\theta^n\}|} = \frac{\theta^{ns_3}}{\theta^{ns_3+2n}} = \frac{1}{\theta^{2n}}.$$

We therefore have

$$P\left(\begin{array}{l} U_j^n(a^{s_j})=u_j^n, X_j^n(M_{jX}, b_{jX})=x_j^n \\ I_j(a^{s_j})=M_{j1}:j=2,3, X_1^n(M_1)=x_1^n \end{array}\right) = \frac{\prod_{j=1}^3 \prod_{t=1}^n p_{X_j|Q}(x_{jt}|q_t)}{\theta^{2n+t_2+t_3}} \leq \frac{\prod_{t=1}^n p_{X_1|Q}(x_{1t}|q_t) \exp\{-nH(X_2|Q)\}}{\exp\{-8n\eta + nH(X_3|Q)\} \theta^{2n+t_2+t_3}} \quad (46)$$

Encoders 2 and 3 choose one among the jointly typical pairs uniformly at random. Hence,

$$\prod_{j=2}^3 P\left(\begin{array}{l} A^{s_j}=a^{s_j} \\ B_{jX}=b_{jX} \end{array} \middle| \begin{array}{l} I_j(a^{s_j})=M_{j1} \\ \phi_j(q^n, M_j) \geq \frac{1}{2} \mathbb{E}\{\phi_j(q^n, M_j)\} \end{array}\right) \leq \frac{4}{\mathbb{E}\{\phi_2(q^n, M_2)\} \mathbb{E}\{\phi_3(q^n, M_3)\}}. \quad (47)$$

It maybe verified from (39) that

$$2\mathcal{L}_j(n) = \mathbb{E}\{\phi_j(q^n, M_j)\} \geq \theta^{s_j-t_j-n} |c_{jX}| \exp\{-n(H(X_j|Q) + 4\eta)\} |T_{2\eta}(U_j, X_j|q^n)|. \quad (48)$$

Substituting (48), (47) and (46) in (44), we have

$$\begin{aligned} P(\tilde{\epsilon}_1^c \cap \epsilon_2) &\leq \sum_{\substack{(u_2^n, u_3^n, \underline{x}^n) \\ \in \Theta(q^n)}} \frac{\exp\{n16\eta\} \prod_{t=1}^n p_{X_1|Q}(x_{1t}|q_t)}{|T_{2\eta}(U_2, X_2|q^n)| |T_{2\eta}(U_3, X_3|q^n)|} \\ &\leq \sum_{\substack{(u_2^n, u_3^n, \underline{x}^n) \\ \in \Theta(q^n)}} \prod_{t=1}^n p_{X_1|Q}(x_{1t}|q_t) \frac{\exp\{24n\eta - nH(U_3, X_3|Q)\}}{\exp\{nH(U_2, X_2|Q)\}} \end{aligned} \quad (49)$$

where the last inequality follows from lower bound on size of the conditional typical set. We now employ the lower bound for conditional probability of jointly typical vectors. In particular,

$$\exp\{-nH(U_j, X_j|Q) - 4n\eta\} \leq \prod_{t=1}^n p_{U_j, X_j|Q}(u_{jt}, x_{jt}|q_t) \leq \exp\{-nH(U_j, X_j|Q) + 4n\eta\} \quad (50)$$

for any $(u_2^n, u_3^n, \underline{x}^n) \in \Theta(q^n)$. Substituting lower bound (50) in (49), for n sufficiently large, we have

$$\begin{aligned} P(\tilde{\epsilon}_1^c \cap \epsilon_2) &\leq \left[\sum_{\substack{(u_2^n, u_3^n, \underline{x}^n) \\ \in \Theta(q^n)}} \prod_{t=1}^n p_{X_1|Q}(x_{1t}|q_t) \prod_{j=2}^3 \prod_{t=1}^n p_{U_j, X_j|Q}(u_{jt}, x_{jt}|q_t) \right] \exp\{32n\eta\} \\ &\leq \left[\sum_{\substack{(u_2^n, u_3^n, \underline{x}^n) \\ \in \Theta(q^n)}} \prod_{t=1}^n p_{X_1 U_2 X_2 U_3 X_3|Q}(x_{1t}, u_{2t}, x_{2t}, u_{3t}, x_{3t}|q_t) \right] \exp\{32n\eta\}, \end{aligned} \quad (51)$$

where (51) 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 due to Hoeffding [28], Sanov [29]. Under the condition $\eta_1 \geq 4\eta$, a ‘conditional version’ of Sanov’s lemma [29] guarantees

$$\sum_{\substack{(u_2^n, u_3^n, \underline{x}^n) \\ \in \bar{\Theta}(q^n)}} \prod_{t=1}^n p_{X_1 U_2 X_2 U_3 X_3 | Q}(x_{1t}, u_{2t}, x_{2t}, u_{3t}, x_{3t} | q_t) \leq 2 \exp\{-n^3 \mu \eta_1^2\} \quad (52)$$

for sufficiently large n , to enable us conclude

$$P(\tilde{\epsilon}_1^c \cap \epsilon_2) \leq 2 \exp\{-n(n^2 \mu \eta_1^2 - 32\eta)\} \quad (53)$$

for such an n .

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

$$\epsilon_3 = \{(q^n, U_2^n(A^{s_2}), U_3^n(A^{s_3}), X_1^n(M_1), X_2^n(M_{2X}, B_{2X}), X_3^n(M_{3X}, B_{3X}), \underline{Y}^n) \notin T_{2\eta_1}(Q, X_1, U_2, U_3, \underline{X}, \underline{Y})\} \quad (54)$$

was defined in (9). 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 $P(\epsilon_2^c \cap \epsilon_3) \leq \frac{\eta}{32}$ for sufficiently large n .

If

$$\bar{\Theta}(q^n) := \left\{ (u_2^n, u_3^n, \underline{x}^n, \underline{y}^n) \in \mathcal{U}_2^n \times \mathcal{U}_3^n \times \mathcal{X}^n \times \mathcal{Y}^n : \begin{array}{l} (u_2^n, u_3^n, \underline{x}^n) \in T_{\eta_1}(U_2, U_3, \underline{X} | q^n), \\ (u_2^n, u_3^n, \underline{x}^n, \underline{y}^n) \notin T_{2\eta_1}(U_2, U_3, \underline{X}, \underline{Y} | q^n) \end{array} \right\},$$

then

$$\begin{aligned} P(\epsilon_2^c \cap \epsilon_3) &= \sum_{\substack{(u_2^n, u_3^n, \underline{x}^n, \underline{y}^n) \\ \in \bar{\Theta}(q^n)}} P(U_j^n(A^{s_j}) = u_j^n, X_j^n(M_{jX}, B_{jX}) = x_j^n : j = 2, 3, X_1^n(M_1) = x_1^n, \underline{Y}^n = \underline{y}^n) \\ &= \sum_{\substack{(u_2^n, u_3^n, \underline{x}^n, \underline{y}^n) \\ \in \bar{\Theta}(q^n)}} P\left(\begin{array}{l} U_j^n(A^{s_j}) = u_j^n, X_1^n(M_1) = x_1^n \\ X_j^n(M_{jX}, B_{jX}) = x_j^n : j = 2, 3, \end{array}\right) P(\underline{Y}^n = \underline{y}^n | \begin{array}{l} U_j^n(A^{s_j}) = u_j^n, X_1^n(M_1) = x_1^n \\ X_j^n(M_{jX}, B_{jX}) = x_j^n : j = 2, 3, \end{array}) \\ &= \sum_{\substack{(u_2^n, u_3^n, \underline{x}^n, \underline{y}^n) \\ \in \bar{\Theta}(q^n)}} P\left(\begin{array}{l} U_j^n(A^{s_j}) = u_j^n, X_1^n(M_1) = x_1^n \\ X_j^n(M_{jX}, B_{jX}) = x_j^n : j = 2, 3, \end{array}\right) \prod_{t=1}^n W_{Y|X}(y_t | \underline{x}_t) \\ &= \sum_{\substack{(u_2^n, u_3^n, \underline{x}^n, \underline{y}^n) \\ \in \bar{\Theta}(q^n)}} P\left(\begin{array}{l} U_j^n(A^{s_j}) = u_j^n, X_1^n(M_1) = x_1^n \\ X_j^n(M_{jX}, B_{jX}) = x_j^n : j = 2, 3, \end{array}\right) \prod_{t=1}^n p_{Y|X U_2 U_3}(y_t | \underline{x}_t, u_{2t}, u_{3t}) \end{aligned} \quad (55)$$

$$\leq \sum_{\substack{(u_2^n, u_3^n, \underline{x}^n) \in \\ T_{\eta_1}(U_2, U_3, \underline{X} | q^n)}} P\left(\begin{array}{l} U_j^n(A^{s_j}) = u_j^n, X_1^n(M_1) = x_1^n \\ X_j^n(M_{jX}, B_{jX}) = x_j^n : j = 2, 3, \end{array}\right) \sum_{\substack{y^n : y^n \notin \\ T_{2\eta_1}(Y | u_2^n, u_3^n, \underline{x}^n)}} \prod_{t=1}^n p_{Y|X U_2 U_3}(y_t | \underline{x}_t, u_{2t}, u_{3t}), \quad (56)$$

where (55) follows from the Markov chain $(Q, U_2, U_3) - \underline{X} - \underline{Y}$. Once again, the upper bound on the probability of conditional typical set enables us to conclude

$$\sum_{\substack{y^n \in \\ T_{2\eta_1}(Y | u_2^n, u_3^n, \underline{x}^n)}} \prod_{t=1}^n p_{Y|X U_2 U_3}(y_t | \underline{x}_t, u_{2t}, u_{3t}) \leq \frac{\eta}{32}$$

and therefore $P(\epsilon_2^c \cap \epsilon_3) \leq \frac{\eta}{32}$ for sufficiently large n .

APPENDIX C

AN UPPER BOUND ON $P((\tilde{\epsilon}_1 \cup \epsilon_2 \cup \epsilon_3)^c \cap \epsilon_{41})$

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 $\tilde{\epsilon}_1 = \epsilon_1 \cup \epsilon_l$,

$$(\epsilon_1 \cup \epsilon_2 \cup \epsilon_3)^c \cap \epsilon_{41} = \bigcup_{a^{s_3} \in \mathcal{U}_3^{s_3}} \bigcup_{\hat{m}_1 \neq M_1} \left\{ \left(\begin{array}{l} U_j^n(A^{s_j}):j=2,3, X_1^n(M_1), \\ X_j^n(M_{jX}, B_{jX}):j=2,3, Y_1^n \end{array} \right) \in \hat{T}(q^n), \left(\begin{array}{l} U_{\oplus}^n(a^{s_3}), Y_1^n \\ X_1^n(\hat{m}_1) \end{array} \right) \in T_{4\eta_1}(U_2 \oplus U_3, Y_1, X_1 | q^n) \right\}.$$

where

$$\hat{T}(q^n) := \left\{ \begin{array}{l} (u_2^n, u_3^n, \underline{x}^n, y_1^n) \in \\ \mathcal{U}_2^n \times \mathcal{U}_3^n \times \mathcal{X}^n \times \mathcal{Y}_1^n \end{array} : \begin{array}{l} (u_2^n, u_3^n, \underline{x}^n, y_1^n) \in T_{2\eta_1}(U_2, U_3, \underline{X}, Y_1 | q^n), \\ (u_2^n, u_3^n, \underline{x}^n) \in T_{\eta_1}(U_2, U_3, \underline{X} | q^n) \\ (u_j^n, x_j^n) \in T_{2\eta}(U_j, X_j | q^n):j=2,3, x_1^n \in T_{2\eta}(X_1 | q^n) \end{array} \right\}.$$

Employing the union bound, we have

$$P((\tilde{\epsilon}_1 \cup \epsilon_2 \cup \epsilon_3)^c \cap \epsilon_{41}) \leq \sum_{\substack{\hat{a}^{s_3} \in \\ \mathcal{U}_3^{s_3}}} \sum_{\substack{m_1, \hat{m}_1 \\ \hat{m}_1 \neq m_1}} \sum_{\substack{(u_2^n, u_3^n, \underline{x}^n, y_1^n) \in \\ \hat{T}(q^n)}} \sum_{\substack{(\hat{u}^n, \hat{x}_1^n) \in \\ T_{4\eta_1}(U_2 \oplus U_3, X_1 | y_1^n, q^n)}} P \left(\left\{ \begin{array}{l} X_j^n(M_{jX}, B_{jX}) = x_j^n, U_j^n(A^{s_j}) = u_j^n \\ I_j(A^{s_j}) = M_{j1}, X_1^n(M_1) = x_1^n, U_{\oplus}(\hat{a}^{s_3}) = \hat{u}^n \\ X_1^n(\hat{m}_1) = \hat{x}_1^n, Y_1^n = y_1^n, M_1 = m_1: j=2,3 \end{array} \right\} \cap \epsilon_l^c \right). \quad (57)$$

We evaluate a generic term in the above sum. Defining $\mathcal{S}(\hat{a}^{s_3}) := \{(a^{s_2}, a^{s_3}) \in \mathcal{U}_2^{s_2} \times \mathcal{U}_3^{s_3} : a^{s_2} 0^{s_+} \oplus a^{s_3} \neq \hat{a}^{s_3}\}$,

where $s_+ := s_3 - s_2$, $\mathcal{S}^c(\hat{a}^{s_3}) := (\mathcal{U}_2^{s_2} \times \mathcal{U}_3^{s_3}) \setminus \mathcal{S}(\hat{a}^{s_3})$, and

$$E := \left\{ \begin{array}{l} X_j^n(m_{jX}, b_{jX}) = x_j^n, U_j^n(a^{s_j}) = u_j^n, M_j = m_j \\ I_j(a^{s_j}) = m_{j1}, X_1^n(M_1) = x_1^n, U_{\oplus}(\hat{a}^{s_3}) = \hat{u}^n, \\ X_1^n(\hat{m}_1) = \hat{x}_1^n, M_1 = m_1: j=2,3, \end{array} \right\}$$

we have

$$P \left(\left\{ \begin{array}{l} X_j^n(M_{jX}, B_{jX}) = x_j^n, U_j^n(A^{s_j}) = u_j^n \\ I_j(A^{s_j}) = M_{j1}, X_1^n(M_1) = x_1^n, U_{\oplus}(\hat{a}^{s_3}) = \hat{u}^n \\ X_1^n(\hat{m}_1) = \hat{x}_1^n, Y_1^n = y_1^n, M_1 = m_1: j=2,3 \end{array} \right\} \cap \epsilon_l^c \right) = \sum_{m_2, m_3} \sum_{b_{2X}, b_{3X}} \sum_{\substack{(a^{s_2}, a^{s_3}) \\ \in \mathcal{S}(\hat{a}^{s_3})}} P \left(E \cap \epsilon_l^c \cap \left\{ \begin{array}{l} Y_1^n = y_1^n, A^{s_j} = a^{s_j} \\ B_{jX} = b_{jX}: j=2,3 \end{array} \right\} \right) \\ + \sum_{m_2, m_3} \sum_{b_{2X}, b_{3X}} \sum_{\substack{(a^{s_2}, a^{s_3}) \\ \in \mathcal{S}^c(\hat{a}^{s_3})}} P \left(E \cap \epsilon_l^c \cap \left\{ \begin{array}{l} Y_1^n = y_1^n, A^{s_j} = a^{s_j} \\ B_{jX} = b_{jX}: j=2,3 \end{array} \right\} \right) \quad (58)$$

Note that

$$P \left(Y_1^n = y_1^n \mid E \cap \epsilon_l^c \cap \left\{ \begin{array}{l} A^{s_j} = a^{s_j} \\ B_{jX} = b_{jX}: j=2,3 \end{array} \right\} \right) = W_{Y_1 | \underline{X}}^n(y_1^n | \underline{x}^n), \quad (59)$$

$$P \left(E \cap \epsilon_l^c \cap \left\{ \begin{array}{l} A^{s_j} = a^{s_j} \\ B_{jX} = b_{jX}: j=2,3 \end{array} \right\} \right) = P(E) P \left(\begin{array}{l} A^{s_j} = a^{s_j} \\ B_{jX} = b_{jX}: j=2,3 \end{array} \mid E \cap \epsilon_l^c \right) = P(E) \frac{1}{\mathcal{L}_2(n) \mathcal{L}_3(n)} \quad (60)$$

Moreover, for $(u_2^n, u_3^n, x_1^n, x_2^n, x_3^n, y_1^n) \in \hat{T}(q^n)$, $(\hat{u}^n, \hat{x}_1^n) \in T_{4\eta_1}(U_2 \oplus U_3, X_1 | y_1^n, q^n)$, we have

$$P(E) \leq \begin{cases} \frac{P(M_j = m_j: j=2,3, M_1 = m_1)}{\theta^{3n+t_2+t_3} \exp\{n(H(X_1|Q) + \sum_{j=1}^3 H(X_j|Q) - 20\eta_1)\}} & \text{if } (a^{s_2}, a^{s_3}) \in \mathcal{S}(\hat{a}^{s_3}), \\ \frac{P(M_{jX} = m_{jX}: j=2,3, M_1 = m_1) W_{Y_1 | \underline{X}}^n(y_1^n | \underline{x}^n) 1_{\{\hat{u}^n = u_2^n \oplus u_3^n\}}}{\theta^{2n+t_2+t_3} \exp\{n(H(X_1|Q) + \sum_{j=1}^3 H(X_j|Q) - 20\eta_1)\}} & \text{if } (a^{s_2}, a^{s_3}) \in \mathcal{S}^c(\hat{a}^{s_3}) \end{cases} \quad (61)$$

In deriving the above upper bounds, we have used the upper bound on conditional probability of jointly typical sequences. We have also employed independence of (triple in the former and pair in the latter) codewords in the coset code. Substituting (59), (60) and (61), in (58), we have

$$P \left(\left\{ \begin{array}{l} X_j^n(M_{jX}, B_{jX}) = x_j^n, U_j^n(A^{s_j}) = u_j^n \\ I_j(A^{s_j}) = M_{j1}, X_1^n(M_1) = x_1^n, U_{\oplus}(\hat{a}^{s_3}) = \hat{u}^n \\ X_1^n(\hat{m}_1) = \hat{x}_1^n, Y_1^n = y_1^n, M_1 = m_1: j=2,3 \end{array} \right\} \cap \epsilon_l^c \right) \leq \frac{\theta^{s_2-t_2} P(M_1 = m_1) W_{Y_1 | \underline{X}}^n(y_1^n | \underline{x}^n) |c_{2X}| |c_{3X}| \left[\frac{\theta^{s_3}}{\theta^n} + 1_{\{\hat{u}^n = u_2^n \oplus u_3^n\}} \right]}{\theta^{2n+t_3} \exp\{n(H(X_1|Q) + \sum_{j=1}^3 H(X_j|Q) - 20\eta_1)\}} \mathcal{L}_2(n) \mathcal{L}_3(n). \quad (62)$$

Our next step is to substitute (62) in (57). Let us restate (57) below as (63) for ease of reference.

$$P((\tilde{\epsilon}_1 \cup \epsilon_2 \cup \epsilon_3)^c \cap \epsilon_{41}) \leq \sum_{\substack{\hat{a}^{s_3} \in \\ \mathcal{U}_3^{s_3}}} \sum_{\substack{m_1, \hat{m}_1 \\ \hat{m}_1 \neq m_1}} \sum_{\substack{(u_2^n, u_3^n, \underline{x}^n, y_1^n) \in \\ \hat{T}(q^n)}} \sum_{\substack{(\hat{u}^n, \hat{x}_1^n) \in \\ T_{4\eta_1}(U_2 \oplus U_3, X_1 | y_1^n, q^n)}} P \left(\left\{ \begin{array}{l} X_j^n(M_{jX}, B_{jX}) = x_j^n, U_j^n(A^{s_j}) = u_j^n \\ I_j(A^{s_j}) = M_{j1}, X_1^n(M_1) = x_1^n, U_{\oplus}(\hat{a}^{s_3}) = \hat{u}^n \\ X_1^n(\hat{m}_1) = \hat{x}_1^n, Y_1^n = y_1^n, M_1 = m_1: j=2,3 \end{array} \right\} \cap \epsilon_l^c \right). \quad (63)$$

We do some spade work before we substitute (62) in (63). (62) is a sum of two terms. The first term is not dependent on the arguments of the innermost summation in (63). By conditional frequency typicality lemma, for sufficiently large n we have $|T_{4\eta_1}(U_2 \oplus U_3, X_1|y_1^n, q^n)| \leq \exp\{n(H(U_2 \oplus U_3, X_1|Y_1, Q)) + 8\eta_1\}$. Substituting this upper bound, the summation in (63) corresponding to the first term in (62) is upper bounded by

$$\mathcal{F}_1 := \sum_{\hat{a}^{s_3}} \sum_{\substack{m_1, \hat{m}_1 \\ \hat{m}_1 \neq m_1}} \sum_{\substack{(u_2^n, u_3^n, \underline{x}^n, y_1^n) \in \\ \hat{T}(q^n)}} \frac{W_{Y_1|X}^n(y_1^n|\underline{x}^n) \theta^{s_2+s_3} |c_{2X}| |c_{3X}| P(M_1 = m_1) \exp\{n(H(U_2 \oplus U_3, X_1|Y_1, Q))\}}{\mathcal{L}_2(n) \mathcal{L}_3(n) \theta^{3n+t_2+t_3} \exp\left\{n(H(X_1|Q)) + \sum_{j=1}^3 H(X_j|Q) - 28\eta_1\right\}}.$$

The indicator in the second term of (62) restricts the outermost summation in (63) to $\hat{x}_1^n \in T_{4\eta_1}(X_1|u_2^n \oplus u_3^n, y_1^n, q^n)$. As earlier, note that the second term is independent of \hat{x}_1^n . Once again, employing the conditional frequency typicality lemma, for sufficiently large n , $|T_{4\eta_1}(X_1|u_2^n \oplus u_3^n, y_1^n, q^n)| \leq \exp\{n(H(X_1|U_2 \oplus U_3, Y_1, Q)) + 8\eta_1\}$. Substituting this upper bound, the summation in (63) corresponding to the second term in (62) is upper bounded by

$$\mathcal{F}_2 := \sum_{\hat{a}^{s_3}} \sum_{\substack{m_1, \hat{m}_1 \\ \hat{m}_1 \neq m_1}} \sum_{\substack{(u_2^n, u_3^n, \underline{x}^n, y_1^n) \in \\ \hat{T}(q^n)}} \frac{W_{Y_1|X}^n(y_1^n|\underline{x}^n) \theta^{s_2} |c_{2X}| |c_{3X}| P(M_1 = m_1) \exp\{n(H(X_1|U_2 \oplus U_3, Y_1, Q))\}}{\mathcal{L}_2(n) \mathcal{L}_3(n) \theta^{2n+t_2+t_3} \exp\left\{n(H(X_1|Q)) + \sum_{j=1}^3 H(X_j|Q) - 28\eta_1\right\}}.$$

It can be verified that

$$\sum_{\substack{(u_2^n, u_3^n, \underline{x}^n, y_1^n) \in \\ \hat{T}(q^n)}} W_{Y_1|X}^n(y_1^n|\underline{x}^n) \leq \min\{|T_{2\eta}(U_2, X_2|q^n)| |T_{2\eta}(U_3, X_3|q^n)| |T_{2\eta}(X_1|q^n)|, |T_{\eta_1}(U_2, U_3, \underline{X}|q^n)|\}. \quad (64)$$

Using (64) and lower bounds $\mathcal{L}_j(n) : j = 2, 3$ from (48), we have

$$\mathcal{F}_1 \leq 2 \frac{\theta^{s_3} \exp\{-n(2H(X_1|Q)) - 8\eta - R_1\} |T_{2\eta}(X_1|q^n)|}{\theta^n \exp\{-n(H(U_2 \oplus U_3, X_1|Y_1, Q)) + 28\eta_1\}} \leq 2 \frac{\theta^{s_3} \exp\{-n(H(X_1|Q)) - 12\eta - R_1\}}{\theta^n \exp\{-n(H(U_2 \oplus U_3, X_1|Y_1, Q)) + 28\eta_1\}},$$

where the last inequality above follows from upper bound on $|T_{2\eta}(X_1|q^n)|$. An identical sequence of steps yields

$$\mathcal{F}_2 \leq 2 \frac{\exp\{-n(H(X_1|Q)) - 28\eta_1 - R_1\}}{\exp\{-n(H(X_1|U_2 \oplus U_3, Y_1, Q)) + 12\eta\}}.$$

for sufficiently large n . Substituting $\frac{s_3 \log \theta}{n} = S_3$, we have

$$P((\tilde{\epsilon}_1 \cup \epsilon_2 \cup \epsilon_3)^c \cap \epsilon_{41}) \leq 2 \exp\{n(28\eta_1 + 12\eta + S_3 + R_1 - \log \theta - H(X_1|Q)) + H(X_1, U_2 \oplus U_3|Y_1, Q)\} \\ + 2 \exp\{n(28\eta_1 + 12\eta + R_1 - I(X_1; U_2 \oplus U_3, Y_1|Q))\}.$$

Employing the definition of δ , we have

$$P((\tilde{\epsilon}_1 \cup \epsilon_2 \cup \epsilon_3)^c \cap \epsilon_{41}) \leq 4 \exp\{-n[\delta - 28\eta_1 - 12\eta]\}. \quad (65)$$

for sufficiently large n .

APPENDIX D

AN UPPER BOUND ON $P((\tilde{\epsilon}_1 \cup \epsilon_2 \cup \epsilon_3)^c \cap \epsilon_{4j})$

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

$$\begin{aligned} \epsilon_{4j}^1 &:= \bigcup_{\hat{m}_{j1} \neq M_{j1}} \bigcup_{\hat{a}^{sj} \in \mathcal{U}_j^{sj}} \bigcup_{\hat{b}_{jX} \in \mathcal{C}_{jX}} \left\{ (q^n, U_j(\hat{a}^{sj}), X_j(M_{jX}, \hat{b}_{jX}), Y_j^n) \in T_{4\eta_1}(Q, U_j, V_j, Y_j), (q^n, U_j(A^{sj}), X_j^n(M_{jX}, B_{jX})) \in \right. \\ &\quad \left. T_{2\eta}(Q, U_j, X_j), I_j(\hat{a}^{sj}) = \hat{m}_{j1}, (q^n, U_j^n(A^{sj}), X_j^n(M_{jX}, B_{jX}), Y_j^n) \in T_{2\eta_1}(Q, U_j, X_j, Y_j) \right\}, \\ \epsilon_{4j}^2 &:= \bigcup_{\hat{m}_{jX} \neq M_{jX}} \bigcup_{a^{sj} \in \mathcal{U}_j^{sj}} \bigcup_{b_{jX} \in \mathcal{C}_{jX}} \left\{ (q^n, U_j(a^{sj}), X_j(\hat{m}_{jX}, b_{jX}), Y_j^n) \in T_{4\eta_1}(Q, U_j, V_j, Y_j), (q^n, U_j(A^{sj}), X_j^n(M_{jX}, B_{jX})) \in \right. \\ &\quad \left. T_{2\eta}(Q, U_j, X_j), I_j(a^{sj}) = M_{j1}, (q^n, U_j^n(A^{sj}), X_j^n(M_{jX}, B_{jX}), Y_j^n) \in T_{2\eta_1}(Q, U_j, X_j, Y_j) \right\}, \\ \epsilon_{4j}^3 &:= \bigcup_{\substack{\hat{m}_{j1} \neq \hat{m}_{jX} \\ M_{j1} \neq M_{jX}}} \bigcup_{a^{sj} \in \mathcal{U}_j^{sj}} \bigcup_{b_{jX} \in \mathcal{C}_{jX}} \left\{ (q^n, U_j(a^{sj}), X_j(\hat{m}_{jX}, b_{jX}), Y_j^n) \in T_{4\eta_1}(Q, U_j, V_j, Y_j), (q^n, U_j(A^{sj}), X_j^n(M_{jX}, B_{jX})) \in \right. \\ &\quad \left. T_{2\eta}(Q, U_j, X_j), I_j(a^{sj}) = \hat{m}_{j1}, (q^n, U_j^n(A^{sj}), X_j^n(M_{jX}, B_{jX}), Y_j^n) \in T_{2\eta_1}(Q, U_j, X_j, Y_j) \right\}. \end{aligned}$$

The event of interest is $\epsilon_{l_j}^c \cap (\epsilon_{4j}^1 \cup \epsilon_{4j}^2 \cup \epsilon_{4j}^3)$. 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,

$$\tilde{T}(q^n) := \{(u_j^n, x_j^n, y_j^n) \in T_{2\eta_1}(U_j, X_j, Y_j | q^n) : (u_j^n, x_j^n) \in T_{2\eta}(U_j, X_j | q^n)\}, \text{ we have (66)}$$

$$\begin{aligned} P(\epsilon_{l_j}^c \cap \epsilon_{4j}^1) &= P \left(\bigcup_{\substack{m_{j1}, \hat{m}_{j1} \in \mathcal{M}_{j1} \\ m_{j1} \neq \hat{m}_{j1}}} \bigcup_{\hat{a}^{sj} \in \mathcal{U}_j^{sj}} \bigcup_{\hat{b}_{jX} \in \mathcal{C}_{jX}} \bigcup_{(u_j^n, x_j^n, y_j^n) \in \tilde{T}(q^n)} \bigcup_{(u_j^n, \hat{x}_j^n) \in T_{4\eta_1}(U_j, X_j | y_j^n, q^n)} \left\{ \begin{array}{l} U_j(A^{sj}) = u_j^n, U_j(\hat{a}^{sj}) = \hat{u}_j^n, M_{j1} = m_{j1} \\ I_j(A^{sj}) = m_{j1}, Y_j^n = y_j^n, I_j(\hat{a}^{sj}) = \hat{m}_{j1}, \\ X_j^n(M_{jX}, B_{jX}) = x_j^n, X_j^n(M_{jX}, \hat{b}_{jX}) = \hat{x}_j^n \end{array} \right\} \cap \epsilon_{l_j}^c \right) \\ &\leq \sum_{\substack{m_{j1}, \hat{m}_{j1} \in \mathcal{M}_{j1} \\ m_{j1} \neq \hat{m}_{j1}}} \sum_{\hat{a}^{sj} \in \mathcal{U}_j^{sj}} \sum_{\hat{b}_{jX} \in \mathcal{C}_{jX}} \sum_{(u_j^n, x_j^n, y_j^n) \in \tilde{T}(q^n)} \sum_{(u_j^n, \hat{x}_j^n) \in T_{4\eta_1}(U_j, X_j | y_j^n, q^n)} P \left(\left\{ \begin{array}{l} U_j(A^{sj}) = u_j^n, U_j(\hat{a}^{sj}) = \hat{u}_j^n, M_{j1} = m_{j1} \\ I_j(A^{sj}) = m_{j1}, Y_j^n = y_j^n, I_j(\hat{a}^{sj}) = \hat{m}_{j1}, \\ X_j^n(M_{jX}, B_{jX}) = x_j^n, X_j^n(M_{jX}, \hat{b}_{jX}) = \hat{x}_j^n \end{array} \right\} \cap \epsilon_{l_j}^c \right). \quad (67) \end{aligned}$$

We now consider two factors of generic term in the above summation. Since $X_1^n(M_1), X_j^n(M_{jX}, B_{jX})$ is independent of the collection $U_j(A^{sj}), U_j(\hat{a}^{sj}), M_{j1}, I_j(A^{sj}), I_j(\hat{a}^{sj}), X_j^n(M_{jX}, B_{jX}), X_j^n(M_{jX}, \hat{b}_{jX})$ for any $(\hat{a}^{sj}, \hat{b}_{jX})$, and $Y_1^n - (X_1^n(M_1), X_j^n(M_{jX}, B_{jX}) : j = 2, 3) - (U_j(A^{sj}), U_j(\hat{a}^{sj}), M_{j1}, I_j(A^{sj}), I_j(\hat{a}^{sj}), X_j^n(M_{jX}, \hat{b}_{jX}))$ is a Markov chain, we have

$$P \left(Y_j^n = y_j^n \left| \begin{array}{l} U_j(A^{sj}) = u_j^n, U_j(\hat{a}^{sj}) = \hat{u}_j^n, M_{j1} = m_{j1} \\ \phi_j(q^n, M_j) \geq \mathcal{L}_j(n), I_j(A^{sj}) = m_{j1}, I_j(\hat{a}^{sj}) = \hat{m}_{j1}, \\ X_j^n(M_{jX}, B_{jX}) = x_j^n, X_j^n(M_{jX}, \hat{b}_{jX}) = \hat{x}_j^n \end{array} \right. \right) = P(Y_j^n = y_j^n | X_j^n(M_{jX}, B_{jX}) = x_j^n) =: \hat{\theta}(y_j^n | x_j^n).$$

By the law of total probability, we have

$$\begin{aligned} P \left(\begin{array}{l} U_j(A^{sj}) = u_j^n, U_j(\hat{a}^{sj}) = \hat{u}_j^n, M_{j1} = m_{j1} \\ \phi_j(q^n, M_j) \geq \mathcal{L}_j(n), I_j(A^{sj}) = m_{j1}, I_j(\hat{a}^{sj}) = \hat{m}_{j1}, \\ X_j^n(M_{jX}, B_{jX}) = x_j^n, X_j^n(M_{jX}, \hat{b}_{jX}) = \hat{x}_j^n \end{array} \right) &= \sum_{m_{jX} \in \mathcal{M}_{jX}} \sum_{a^{sj} \in \mathcal{U}_j^{sj}} P \left(\left\{ \begin{array}{l} U_j(a^{sj}) = u_j^n, U_j(\hat{a}^{sj}) = \hat{u}_j^n, M_j = m_j, B_{jX} = \hat{b}_{jX} \\ A^{sj} = a^{sj}, I_j(a^{sj}) = m_{j1}, I_j(\hat{a}^{sj}) = \hat{m}_{j1}, \\ X_j^n(m_{jX}, \hat{b}_{jX}) = x_j^n, X_j^n(m_{jX}, b_{jX}) = \hat{x}_j^n \end{array} \right\} \cap \epsilon_{l_j}^c \right) + \\ &+ \sum_{m_{jX} \in \mathcal{M}_{jX}} \sum_{a^{sj} \in \mathcal{U}_j^{sj}} \sum_{\substack{b_{jX} \in \mathcal{C}_{jX} \\ b_{jX} \neq \hat{b}_{jX}}} P \left(\left\{ \begin{array}{l} U_j(a^{sj}) = u_j^n, U_j(\hat{a}^{sj}) = \hat{u}_j^n, M_j = m_j, B_{jX} = b_{jX} \\ A^{sj} = a^{sj}, I_j(a^{sj}) = m_{j1}, I_j(\hat{a}^{sj}) = \hat{m}_{j1}, \\ X_j^n(m_{jX}, b_{jX}) = x_j^n, X_j^n(m_{jX}, \hat{b}_{jX}) = \hat{x}_j^n \end{array} \right\} \cap \epsilon_{l_j}^c \right). \end{aligned}$$

Now recognize that a generic term of the sum in (67) 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 (67), we simplify the terms involved in the

second identity (involving the two sums). Denoting

$$E^1 := \left\{ \begin{array}{l} U_j(a^{sj})=u_j^n, U_j(\hat{a}^{sj})=\hat{u}_j^n, M_j=m_j \\ I_j(a^{sj})=m_{j1}, I_j(\hat{a}^{sj})=\hat{m}_{j1}, \\ X_j^n(m_{jX}, b_{jX})=x_j^n, X_j^n(m_{jX}, \hat{b}_{jX})=\hat{x}_j^n \end{array} \right\}, \text{ we have,}$$

$$P \left(\left\{ \begin{array}{l} U_j(a^{sj})=u_j^n, U_j(\hat{a}^{sj})=\hat{u}_j^n, M_j=m_j, B_{jX}=b_{jX} \\ A^{sj}=a^{sj}, I_j(a^{sj})=m_{j1}, I_j(\hat{a}^{sj})=\hat{m}_{j1}, \\ X_j^n(m_{jX}, b_{jX})=x_j^n, X_j^n(m_{jX}, \hat{b}_{jX})=\hat{x}_j^n \end{array} \right\} \cap \epsilon_{I_j}^c \right) \leq P(E^1) P \left(A^{sj}=a^{sj} \middle| E^1 \cap \epsilon_{I_j}^c \right) \text{ where,}$$

$$P(E^1) = P \left(\begin{array}{l} M_j=m_j, I_j(a^{sj})=m_{j1}, I_j(\hat{a}^{sj})=\hat{m}_{j1}, \\ X_j^n(m_{jX}, b_{jX})=x_j^n, X_j^n(m_{jX}, \hat{b}_{jX})=\hat{x}_j^n \end{array} \right) P \left(U_j^n(\hat{a}^{sj})=\hat{u}_j^n \middle| U_j(a^{sj})=u_j^n \right), \quad P \left(A^{sj}=a^{sj} \middle| E^1 \cap \epsilon_{I_j}^c \right) = \frac{1}{\mathcal{L}_j(n)} = \frac{2}{\mathbb{E}\{\phi_j(q^n, M_j)\}} \quad (68)$$

Let us work with $P(E^1)$. If $\hat{m}_{j1} \neq m_{j1}$ and $\hat{a}^{sj} \neq a^{sj}$, then

$$P \left(\begin{array}{l} M_j=m_j, I_j(a^{sj})=m_{j1}, I_j(\hat{a}^{sj})=\hat{m}_{j1}, \\ X_j^n(m_{jX}, b_{jX})=x_j^n, X_j^n(m_{jX}, \hat{b}_{jX})=\hat{x}_j^n \end{array} \right) P \left(U_j^n(\hat{a}^{sj})=\hat{u}_j^n \middle| U_j(a^{sj})=u_j^n \right) \leq \begin{cases} \frac{P(M_j=m_j) \exp\{-n(2H(X_j|Q))\}}{\theta^{2n+2t_j} \exp\{-n4\eta - n8\eta_1\}} & \text{if } \hat{b}_{jX} \neq b_{jX} \\ \frac{P(M_j=m_j) \exp\{-n(H(X_j|Q))\}}{\theta^{2n+2t_j} \exp\{-n4\eta\}} & \text{otherwise.} \end{cases} \quad (69)$$

Substituting the above observations in (67), we have

$$\begin{aligned} P(\epsilon_{I_j}^c \cap \epsilon_{4j}^1) &\leq \sum_{m_j \in \mathcal{M}_j} \sum_{\hat{m}_{j1} \neq m_{j1}} \sum_{\substack{a^{sj}, \hat{a}^{sj} \\ a^{sj} \neq \hat{a}^{sj}}} \sum_{\substack{b_{jX}, \hat{b}_{jX} \\ \hat{b}_{jX} \neq b_{jX}}} \sum_{\substack{(u_j^n, x_j^n, y_j^n) \in \\ \hat{T}(q^n)}} \hat{\theta}(y_j^n | x_j^n) \sum_{\substack{(\hat{u}_j^n, \hat{x}_j^n) \in \\ T_{4\eta_1}(U_j, X_j | y_j^n, q^n)}} \frac{P(M_j = m_j) \exp\{-2nH(X_j|Q)\}}{\theta^{2n+2t_j} \exp\{-n4\eta - n8\eta_1\}} \mathcal{L}_j(n) + \\ &+ \sum_{m_j \in \mathcal{M}_j} \sum_{\hat{m}_{j1} \neq m_{j1}} \sum_{\substack{a^{sj}, \hat{a}^{sj} \\ a^{sj} \neq \hat{a}^{sj}}} \sum_{b_{jX} \in c_{jX}} \sum_{\substack{(u_j^n, x_j^n, y_j^n) \in \\ \hat{T}(q^n)}} \hat{\theta}(y_j^n | x_j^n) \sum_{\substack{\hat{u}_j^n \in \\ T_{4\eta_1}(U_j | x_j^n, y_j^n, q^n)}} \frac{P(M_j = m_j) \exp\{-nH(X_j|Q)\}}{\theta^{2n+2t_j} \exp\{-n4\eta\}} \mathcal{L}_j(n). \end{aligned}$$

We now employ the upper bound on cardinality of the conditional frequency typical sets $T_{4\eta_1}(U_j, X_j | y_j^n, q^n)$ and $T_{4\eta_1}(U_j | x_j^n, y_j^n, q^n)$. For sufficiently large n ,

$$|T_{4\eta_1}(U_j, X_j | y_j^n, q^n)| \leq \exp\{n(H(U_j, X_j | Y_j, Q) + 8\eta_1)\}, \quad |T_{4\eta_1}(U_j | x_j^n, y_j^n, q^n)| \leq \exp\{n(H(U_j | X_j, Y_j, Q) + 8\eta_1)\},$$

for any $(x_j^n, y_j^n, q^n) \in T_{2\eta_1}(X_j, Y_j, Q)$. For such an n , we have

$$\begin{aligned} P(\epsilon_{I_j}^c \cap \epsilon_{4j}^1) &\leq \sum_{m_j \in \mathcal{M}_j} \sum_{\hat{m}_{j1} \neq m_{j1}} \sum_{\substack{a^{sj}, \hat{a}^{sj} \\ a^{sj} \neq \hat{a}^{sj}}} \sum_{\substack{b_{jX}, \hat{b}_{jX} \\ \hat{b}_{jX} \neq b_{jX}}} \sum_{\substack{(u_j^n, x_j^n, y_j^n) \in \\ \hat{T}(q^n)}} \hat{\theta}(y_j^n | x_j^n) \frac{P(M_j = m_j) \exp\{-2nH(X_j|Q) + n16\eta_1\}}{\theta^{2n+2t_j} \exp\{-n4\eta - nH(U_j, X_j | Y_j, Q)\}} \mathcal{L}_j(n) + \\ &+ \sum_{m_j \in \mathcal{M}_j} \sum_{\hat{m}_{j1} \neq m_{j1}} \sum_{\substack{a^{sj}, \hat{a}^{sj} \\ a^{sj} \neq \hat{a}^{sj}}} \sum_{b_{jX} \in c_{jX}} \sum_{\substack{(u_j^n, x_j^n, y_j^n) \in \\ \hat{T}(q^n)}} \hat{\theta}(y_j^n | x_j^n) \frac{P(M_j = m_j) \exp\{-nH(X_j|Q) + 8n\eta_1\}}{\theta^{2n+2t_j} \exp\{-n4\eta - nH(U_j | X_j, Y_j, Q)\}} \mathcal{L}_j(n) \\ &\leq \sum_{m_j \in \mathcal{M}_j} \sum_{\hat{m}_{j1} \neq m_{j1}} \sum_{\substack{a^{sj}, \hat{a}^{sj} \\ a^{sj} \neq \hat{a}^{sj}}} \sum_{\substack{b_{jX}, \hat{b}_{jX} \\ \hat{b}_{jX} \neq b_{jX}}} \sum_{\substack{(u_j^n, x_j^n) \in \\ T_{2\eta}(U_j, X_j | q^n)}} \frac{P(M_j = m_j) \exp\{-2nH(X_j|Q) + n16\eta_1\}}{\theta^{2n+2t_j} \exp\{-n4\eta - nH(U_j, X_j | Y_j, Q)\}} \mathcal{L}_j(n) + \\ &+ \sum_{m_j \in \mathcal{M}_j} \sum_{\hat{m}_{j1} \neq m_{j1}} \sum_{\substack{a^{sj}, \hat{a}^{sj} \\ a^{sj} \neq \hat{a}^{sj}}} \sum_{b_{jX} \in c_{jX}} \sum_{\substack{(u_j^n, x_j^n) \in \\ T_{2\eta}(U_j, X_j | q^n)}} \frac{P(M_j = m_j) \exp\{-nH(X_j|Q) + 8n\eta_1\}}{\theta^{2n+2t_j} \exp\{-n4\eta - nH(U_j | X_j, Y_j, Q)\}} \mathcal{L}_j(n). \end{aligned}$$

Substituting the lower bound for $\mathcal{L}_j(n)$ from (48) and noting that the terms in the summation do not depend on the arguments of the sum, for $n \geq N_{11}(\eta_1)$, it can be verified that

$$P(\epsilon_{I_j}^c \cap \epsilon_{4j}^1) \leq 2 \frac{\theta^{s_j} \exp\{-nH(X_j|Q) + 8n\eta_1 + 4n\eta\}}{\theta^n \exp\{-nH(U_j | X_j, Y_j, Q)\}} \left(\frac{\exp\{-nH(X_j|Q) + 8n\eta_1\}}{\exp\{-nH(X_j | Y_j, Q) - nK_j\}} + 1 \right).$$

Finally, substituting $\frac{s_j \log \theta}{n} = S_j, \delta$, we have

$$\begin{aligned} P(\epsilon_{l_j}^c \cap \epsilon_{4j}^1) &\leq 2 \exp\{-n[(\log \theta - H(U_j|X_j, Y_j, Q)) - S_j - (8\eta_1 + 4\eta)]\} + \\ &+ 2 \exp\{-n[(\log \theta + H(X_j|Q) - H(U_j, X_j|Y_j, Q)) - (S_j + K_j) - (16\eta_1 + 4\eta)]\} \\ &\leq 4 \exp\{-n[\delta - (16\eta_1 + 8\eta)]\} \end{aligned} \quad (70)$$

for sufficiently large n .

We follow a similar sequence of steps to derive an upper bound on $P(\epsilon_{4j}^2)$. Defining $\tilde{T}(q^n)$ as in (66), we have

$$\begin{aligned} P(\epsilon_{l_j}^c \cap \epsilon_{4j}^2) &= P \left(\bigcup_{\substack{m_{jX}, \hat{m}_{jX} \in \mathcal{M}_{jX} \\ \hat{m}_{jX} \neq m_{jX}}} \bigcup_{\substack{\hat{a}^{sj} \\ \in U_j^{sj}}} \bigcup_{\substack{\hat{b}_{jX} \\ \in c_{jX}}} \bigcup_{\substack{(u_j^n, x_j^n, y_j^n) \in \\ \tilde{T}(q^n)}} \bigcup_{(\hat{u}_j^n, \hat{x}_j^n) \in T_{4\eta_1}(U_j, X_j|y_j^n, q^n)} \left\{ \begin{array}{l} X_j^n(\hat{m}_{jX}, \hat{b}_{jX}) = \hat{x}_j^n, U_j(\hat{a}^{sj}) = \hat{u}_j^n, Y_j^n = y_j^n \\ I_j(A^{sj}) = I_j(\hat{a}^{sj}) = M_{j1}, M_{jX} = m_{jX}, \\ X_j^n(M_{jX}, B_{jX}) = x_j^n, U_j(A^{sj}) = u_j^n \end{array} \right\} \cap \epsilon_{l_j}^c \right) \\ &\leq \sum_{\substack{m_{jX}, \hat{m}_{jX} \in \mathcal{M}_{jX} \\ \hat{m}_{jX} \neq m_{jX}}} \sum_{\substack{\hat{a}^{sj} \\ \in U_j^{sj}}} \sum_{\substack{\hat{b}_{jX} \\ \in c_{jX}}} \sum_{\substack{(u_j^n, x_j^n, y_j^n) \in \\ \tilde{T}(q^n)}} \sum_{(\hat{u}_j^n, \hat{x}_j^n) \in T_{4\eta_1}(U_j, X_j|y_j^n, q^n)} P \left(\left\{ \begin{array}{l} X_j^n(\hat{m}_{jX}, \hat{b}_{jX}) = \hat{x}_j^n, U_j(\hat{a}^{sj}) = \hat{u}_j^n, Y_j^n = y_j^n \\ I_j(A^{sj}) = I_j(\hat{a}^{sj}) = M_{j1}, M_{jX} = m_{jX}, \\ X_j^n(M_{jX}, B_{jX}) = x_j^n, U_j(A^{sj}) = u_j^n \end{array} \right\} \cap \epsilon_{l_j}^c \right) \end{aligned} \quad (71)$$

We now consider two factors of a generic term in the above sum. Since $X_1^n(M_1), X_j^n(M_{jX}, B_{jX})$ is independent of the collection $X_j^n(\hat{m}_{jX}, \hat{b}_{jX}), U_j(\hat{a}^{sj}), I_j(A^{sj}), I_j(\hat{a}^{sj}), M_{jX}, X_j^n(M_{jX}, B_{jX}), U_j(A^{sj})$ for any $(\hat{a}^{sj}, \hat{b}_{jX})$ as long as $\hat{m}_{jX} \neq M_{jX}$, and $Y_1^n - (X_1^n(M_1), X_j^n(M_{jX}, B_{jX}) : j = 2, 3) - (X_j^n(\hat{m}_{jX}, \hat{b}_{jX}), U_j(\hat{a}^{sj}), I_j(A^{sj}), I_j(\hat{a}^{sj}), M_{jX}, X_j^n(M_{jX}, B_{jX}), U_j(A^{sj}))$ is a Markov chain, we have

$$P \left(Y_j^n = y_j^n \left| \left\{ \begin{array}{l} X_j^n(\hat{m}_{jX}, \hat{b}_{jX}) = \hat{x}_j^n, U_j(\hat{a}^{sj}) = \hat{u}_j^n \\ I_j(A^{sj}) = I_j(\hat{a}^{sj}) = M_{j1}, M_{jX} = m_{jX}, \\ X_j^n(M_{jX}, B_{jX}) = x_j^n, U_j(A^{sj}) = u_j^n \end{array} \right\} \cap \epsilon_{l_j}^c \right. \right) = P(Y_j^n = y_j^n | X_j^n(M_{jX}, B_{jX}) = x_j^n) =: \hat{\theta}(y_j^n | x_j^n).$$

By the law of total probability, we have

$$\begin{aligned} P \left(\left\{ \begin{array}{l} X_j^n(\hat{m}_{jX}, \hat{b}_{jX}) = \hat{x}_j^n, U_j(\hat{a}^{sj}) = \hat{u}_j^n \\ I_j(A^{sj}) = I_j(\hat{a}^{sj}) = M_{j1}, M_{jX} = m_{jX}, \\ X_j^n(M_{jX}, B_{jX}) = x_j^n, U_j(A^{sj}) = u_j^n \end{array} \right\} \cap \epsilon_{l_j}^c \right) &= \sum_{m_{j1} \in \mathcal{M}_{j1}} \sum_{b_{jX} \in c_{jX}} P \left(\left\{ \begin{array}{l} X_j^n(\hat{m}_{jX}, \hat{b}_{jX}) = \hat{x}_j^n, U_j(\hat{a}^{sj}) = \hat{u}_j^n, A^{sj} = \hat{a}^{sj} \\ I_j(\hat{a}^{sj}) = M_{j1}, M_j = m_j, B_{jX} = b_{jX} \\ X_j^n(m_{jX}, b_{jX}) = x_j^n, U_j(\hat{a}^{sj}) = u_j^n \end{array} \right\} \cap \epsilon_{l_j}^c \right) \\ &+ \sum_{m_{j1} \in \mathcal{M}_{j1}} \sum_{b_{jX} \in c_{jX}} \sum_{\substack{a^{sj} \in U_j^{sj} \\ a^{sj} \neq \hat{a}^{sj}}} P \left(\left\{ \begin{array}{l} X_j^n(\hat{m}_{jX}, \hat{b}_{jX}) = \hat{x}_j^n, U_j(\hat{a}^{sj}) = \hat{u}_j^n, A^{sj} = a^{sj} \\ I_j(a^{sj}) = I_j(\hat{a}^{sj}) = M_{j1}, M_j = m_j, B_{jX} = b_{jX} \\ X_j^n(m_{jX}, b_{jX}) = x_j^n, U_j(a^{sj}) = u_j^n \end{array} \right\} \cap \epsilon_{l_j}^c \right). \end{aligned}$$

Now recognize that a generic term of the sum in (71) 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 (71), we simplify the terms involved in the second identity (involving the two sums). Denoting

$$E^2 := \left\{ \begin{array}{l} X_j^n(\hat{m}_{jX}, \hat{b}_{jX}) = \hat{x}_j^n, U_j(\hat{a}^{sj}) = \hat{u}_j^n, \\ I_j(a^{sj}) = I_j(\hat{a}^{sj}) = m_{j1}, M_j = m_j \\ X_j^n(m_{jX}, b_{jX}) = x_j^n, U_j(a^{sj}) = u_j^n \end{array} \right\}, \text{ we have,}$$

$$P \left(\left\{ \begin{array}{l} X_j^n(\hat{m}_{jX}, \hat{b}_{jX}) = \hat{x}_j^n, U_j(\hat{a}^{sj}) = \hat{u}_j^n, A^{sj} = a^{sj} \\ I_j(a^{sj}) = I_j(\hat{a}^{sj}) = m_{j1}, M_j = m_j, B_{jX} = b_{jX} \\ X_j^n(m_{jX}, b_{jX}) = x_j^n, U_j(a^{sj}) = u_j^n \end{array} \right\} \cap \epsilon_{l_j}^c \right) \leq P(E^2) P \left(A^{sj} = a^{sj} \left| B_{jX} = b_{jX} \right. E^2 \cap \epsilon_{l_j}^c \right),$$

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 $\hat{m}_{jX} \neq m_{jX}$, we have

$$\begin{aligned} P\left(\begin{array}{l} X_j^n(\hat{m}_{jX}, \hat{b}_{jX}) = \hat{x}_j^n, U_j(\hat{a}^{s_j}) = \hat{u}_j^n, \\ I_j(a^{s_j}) = I_j(\hat{a}^{s_j}) = m_{j1}, M_j = m_j \\ X_j^n(m_{jX}, b_{jX}) = x_j^n, U_j(a^{s_j}) = u_j^n \end{array}\right) &= P(M_j = m_j) P\left(\begin{array}{l} X_j^n(\hat{m}_{jX}, \hat{b}_{jX}) = \hat{x}_j^n \\ X_j^n(m_{jX}, b_{jX}) = x_j^n \end{array}\right) P\left(\begin{array}{l} U_j(a^{s_j}) = u_j^n \\ U_j(\hat{a}^{s_j}) = \hat{u}_j^n \end{array}\right) P\left(\begin{array}{l} I_j(a^{s_j}) = m_{j1} \\ I_j(\hat{a}^{s_j}) = m_{j1} \end{array}\right) \\ &= \begin{cases} \frac{P(M_j = m_j) \exp\{-2nH(X_j|Q)\}}{\theta^{n+t_j} \exp\{-4n\eta - 8n\eta_1\}} & \text{if } a^{s_j} = \hat{a}^{s_j}, u_j^n = \hat{u}_j^n \\ \frac{P(M_j = m_j) \exp\{-2nH(X_j|Q)\}}{\theta^{2n+2t_j} \exp\{-4n\eta - 8n\eta_1\}} & \text{if } a^{s_j} \neq \hat{a}^{s_j} \end{cases}. \end{aligned}$$

Substituting the above observations in (71), we have

$$\begin{aligned} P(\epsilon_{l_j}^c \cap \epsilon_{4j}^2) &\leq \sum_{m_j \in \mathcal{M}_j} \sum_{\hat{m}_{jX} \neq m_{jX}} \sum_{\substack{b_{jX}, \hat{b}_{jX} \\ \in c_{jX}}} \sum_{\substack{a^{s_j} \in \mathcal{U}_j^{s_j} \\ a^{s_j} \neq \hat{a}^{s_j}}} \sum_{\substack{(u_j^n, x_j^n, y_j^n) \in \\ \hat{T}(q^n)}} \hat{\theta}(y_j^n | x_j^n) \sum_{\substack{\hat{x}_j^n \in \\ T_{4\eta_1}(X_j | u_j^n, y_j^n, q^n)}} \frac{P(M_j = m_j) \exp\{-2nH(X_j|Q)\}}{\theta^{n+t_j} \exp\{-4n\eta - n8\eta_1\}} \mathcal{L}_j(n) + \\ &+ \sum_{m_j \in \mathcal{M}_j} \sum_{\hat{m}_{jX} \neq m_{jX}} \sum_{\substack{b_{jX}, \hat{b}_{jX} \\ \in c_{jX}}} \sum_{\substack{a^{s_j}, \hat{a}^{s_j} \in \mathcal{U}_j^{s_j} \\ a^{s_j} \neq \hat{a}^{s_j}}} \sum_{\substack{(u_j^n, x_j^n, y_j^n) \in \\ \hat{T}(q^n)}} \hat{\theta}(y_j^n | x_j^n) \sum_{\substack{(\hat{u}_j^n, \hat{x}_j^n) \in \\ T_{4\eta_1}(U_j, X_j | y_j^n, q^n)}} \frac{P(M_j = m_j) \exp\{-2nH(X_j|Q)\}}{\theta^{2n+2t_j} \exp\{-4n\eta - n8\eta_1\}} \mathcal{L}_j(n). \end{aligned}$$

We now employ the upper bounds on $|T_{4\eta_1}(X_j | u_j^n, y_j^n, q^n)|$ and $|T_{4\eta_1}(U_j, X_j | y_j^n, q^n)|$. For sufficiently large n , $|T_{4\eta_1}(X_j | u_j^n, y_j^n, q^n)| \leq \exp\{n(H(X_j | U_j, Y_j, Q) + 8\eta_1)\}$ and $|T_{4\eta_1}(U_j, X_j | y_j^n, q^n)| \leq \exp\{n(H(U_j, X_j | Y_j, Q) + 8\eta_1)\}$ for all $(u_j^n, y_j^n, q^n) \in T_{2\eta_1}(U_j, Y_j, Q)$. For such an n , we have

$$\begin{aligned} P(\epsilon_{l_j}^c \cap \epsilon_{4j}^2) &\leq \sum_{m_j \in \mathcal{M}_j} \sum_{\hat{m}_{jX} \neq m_{jX}} \sum_{\substack{b_{jX}, \hat{b}_{jX} \\ \in c_{jX}}} \sum_{\substack{a^{s_j} \in \mathcal{U}_j^{s_j} \\ a^{s_j} \neq \hat{a}^{s_j}}} \sum_{\hat{T}(q^n)} \hat{\theta}(y_j^n | x_j^n) \frac{\theta^{-n-t_j} P(M_j = m_j) \exp\{-2nH(X_j|Q)\}}{\exp\{-4n\eta - n16\eta_1 - nH(X_j | U_j, Y_j, Q)\}} \mathcal{L}_j(n) + \\ &+ \sum_{m_j \in \mathcal{M}_j} \sum_{\hat{m}_{jX} \neq m_{jX}} \sum_{\substack{b_{jX}, \hat{b}_{jX} \\ \in c_{jX}}} \sum_{\substack{a^{s_j}, \hat{a}^{s_j} \in \mathcal{U}_j^{s_j} \\ a^{s_j} \neq \hat{a}^{s_j}}} \sum_{\hat{T}(q^n)} \hat{\theta}(y_j^n | x_j^n) \frac{\theta^{-2n-2t_j} P(M_j = m_j) \exp\{-2nH(X_j|Q)\}}{\exp\{-4n\eta - n16\eta_1 - nH(X_j, U_j | Y_j, Q)\}} \mathcal{L}_j(n) \\ &\leq \sum_{m_j \in \mathcal{M}_j} \sum_{\hat{m}_{jX} \neq m_{jX}} \sum_{\substack{b_{jX}, \hat{b}_{jX} \\ \in c_{jX}}} \sum_{\mathcal{U}_j^{s_j}} \sum_{T_{2\eta}(U_j, X_j | q^n)} \frac{\theta^{-n-t_j} P(M_j = m_j) \exp\{-2nH(X_j|Q)\}}{\exp\{-4n\eta - n16\eta_1 - nH(X_j | U_j, Y_j, Q)\}} \mathcal{L}_j(n) + \\ &+ \sum_{m_j \in \mathcal{M}_j} \sum_{\hat{m}_{jX} \neq m_{jX}} \sum_{\substack{b_{jX}, \hat{b}_{jX} \\ \in c_{jX}}} \sum_{\substack{a^{s_j}, \hat{a}^{s_j} \in \mathcal{U}_j^{s_j} \\ a^{s_j} \neq \hat{a}^{s_j}}} \sum_{T_{2\eta}(U_j, X_j | q^n)} \frac{\theta^{-2n-2t_j} P(M_j = m_j) \exp\{-2nH(X_j|Q)\}}{\exp\{-4n\eta - n16\eta_1 - nH(X_j, U_j | Y_j, Q)\}} \mathcal{L}_j(n). \end{aligned}$$

Substituting the lower bound for $\mathcal{L}_j(n)$ from (48), we have

$$\begin{aligned}
P(\epsilon_{l_j}^c \cap \epsilon_{4_j}^2) &\leq 2 \sum_{m_j \in \mathcal{M}_j} \sum_{\hat{m}_{jX} \neq m_{jX}} \sum_{\substack{b_{jX}, \hat{b}_{jX} \\ \in c_{jX}}} \sum_{\substack{a^{sj} \in \\ \mathcal{U}_j^{sj}}} \frac{P(M_j = m_j) \exp\{-nH(X_j|Q) + n16\eta_1\}}{\theta^{sj} \exp\{-n8\eta - nH(X_j|U_j, Y_j, Q)\} |c_{jX}|} + \\
&+ 2 \sum_{m_j \in \mathcal{M}_j} \sum_{\hat{m}_{jX} \neq m_{jX}} \sum_{\substack{b_{jX}, \hat{b}_{jX} \\ \in c_{jX}}} \sum_{\substack{a^{sj}, \hat{a}^{sj} \in \mathcal{U}_j^{sj} \\ a^{sj} \neq \hat{a}^{sj}}} \frac{P(M_j = m_j) \theta^{-sj} \exp\{-nH(X_j|Q) + n16\eta_1\}}{\theta^{n+t_j} \exp\{-n8\eta - nH(X_j, U_j|Y_j, Q)\} |c_{jX}|} \\
&\leq 2 \sum_{m_j \in \mathcal{M}_j} \sum_{\hat{m}_{jX} \neq m_{jX}} \sum_{\substack{b_{jX}, \hat{b}_{jX} \\ \in c_{jX}}} \frac{P(M_j = m_j) \exp\{-nH(X_j|Q) + n16\eta_1\}}{\exp\{-n8\eta - nH(X_j|U_j, Y_j, Q)\} |c_{jX}|} + \\
&+ 2 \sum_{m_j \in \mathcal{M}_j} \sum_{\hat{m}_{jX} \neq m_{jX}} \sum_{\substack{b_{jX}, \hat{b}_{jX} \\ \in c_{jX}}} \frac{P(M_j = m_j) \theta^{sj} \exp\{-nH(X_j|Q) + n16\eta_1\}}{\theta^{n+t_j} \exp\{-n8\eta - nH(X_j, U_j|Y_j, Q)\} |c_{jX}|} \\
&\leq 2 \sum_{m_j \in \mathcal{M}_j} \sum_{\hat{m}_{jX} \neq m_{jX}} \frac{P(M_j = m_j) \exp\{-nH(X_j|Q) + n16\eta_1\}}{\exp\{-n8\eta - nH(X_j|U_j, Y_j, Q) - nK_j\}} + \\
&+ 2 \sum_{m_j \in \mathcal{M}_j} \sum_{\hat{m}_{jX} \neq m_{jX}} \frac{P(M_j = m_j) \theta^{sj} \exp\{-nH(X_j|Q) + n16\eta_1\}}{\theta^{n+t_j} \exp\{-n8\eta - nH(X_j, U_j|Y_j, Q) - nK_j\}} \\
&\leq 2 \frac{\exp\{-nH(X_j|Q) + nL_j + n\eta_1 + n16\eta_1\}}{\exp\{-n8\eta - nH(X_j|U_j, Y_j, Q) - nK_j\}} \left[1 + \frac{\exp\{nH(U_j|Y_j, Q)\}}{\theta^{n+t_j-s_j}} \right]
\end{aligned}$$

We have for sufficiently large n

$$\begin{aligned}
P(\epsilon_{l_j}^c \cap \epsilon_{4_j}^2) &\leq 2 \exp\{-n(I(X_j; U_j, Y_j|Q) - K_j - L_j - [9\eta_1 + 16\eta_1])\} \\
&+ 2 \exp\left\{-n \left[\left(\frac{\log \theta + H(X_j|Q)}{H(X_j, U_j|Y_j, Q)} \right) - \left(\frac{K_j + L_j +}{(s_j - T_j) \log \theta} \right) - [(9 + 16\eta_1)] \right]\right\} \\
&\leq 4 \exp\{-n(\delta - (9\eta + 16\eta_1))\}.
\end{aligned} \tag{72}$$

We are left to study $P(\epsilon_{4_j}^3)$. Defining $\tilde{T}(q^n)$ as in (66), and

$$E^3 := \left\{ \begin{array}{l} X_j^n(\hat{m}_{jX}, \hat{b}_{jX}) = \hat{x}_j^n, U_j(\hat{a}^{sj}) = \hat{u}_j^n \\ I_j(a^{sj}) = m_{j1}, I_j(\hat{a}^{sj}) = \hat{m}_{j1} \\ X_j^n(m_{jX}, b_{jX}) = x_j^n, U_j(a^{sj}) = u_j^n, M_j = m_j \end{array} \right\} \tag{73}$$

the union bound yields

$$P(\epsilon_{l_j}^c \cap \epsilon_{4_j}^3) \leq \sum_{\substack{m_{j1}, \hat{m}_{j1} \\ m_{j1} \neq \hat{m}_{j1}}} \sum_{\substack{m_{jX}, \hat{m}_{jX} \\ m_{jX} \neq \hat{m}_{jX}}} \sum_{\substack{a^{sj}, \hat{a}^{sj} \\ \hat{a}^{sj} \neq a^{sj}}} \sum_{\substack{b_{jX}, \hat{b}_{jX}}} \sum_{\substack{(u_j^n, x_j^n, y_j^n) \in \\ \tilde{T}(q^n)}} \sum_{\substack{(\hat{u}_j^n, \hat{x}_j^n) \in \\ T_{4\eta_1}(U_j, X_j|y_j^n, q^n)}} P\left(\left\{Y_j^n = y_j^n, B_{jX} = b_{jX}\right\} \cap E^3 \cap \epsilon_{l_j}^c\right) \tag{74}$$

As earlier, we consider a generic term in the above sum and simplify the same. Observe that

$$\begin{aligned}
P\left(Y_j^n = y_j^n \mid \left\{ \begin{array}{l} A^{sj} = a^{sj} \\ B_{jX} = b_{jX} \end{array} \right\} \cap E^3 \cap \epsilon_{l_j}^c\right) &= P\left(Y_j^n = y_j^n \mid X_j^n(M_{jX}, B_{jX}) = x_j^n\right) =: \hat{\theta}(y_j^n | x_j^n), \\
P\left(\left\{ \begin{array}{l} A^{sj} = a^{sj} \\ B_{jX} = b_{jX} \end{array} \right\} \cap E^3 \cap \epsilon_{l_j}^c\right) &\leq P(E^3) P\left(\left\{ \begin{array}{l} A^{sj} = a^{sj} \\ B_{jX} = b_{jX} \end{array} \right\} \mid E^3 \cap \epsilon_{l_j}^c\right) \\
&\leq \frac{P(M_j = m_j) \exp\{-2nH(X_j|Q)\}}{\theta^{2n+2t_j} \exp\{-4n\eta - 8n\eta_1\}} \frac{1}{\mathcal{L}_j(n)}.
\end{aligned}$$

Substituting the above observations in (74), we have

$$P(\epsilon_{l_j}^c \cap \epsilon_{4_j}^3) \leq \sum_{\substack{m_{j1}, \hat{m}_{j1} \\ m_{j1} \neq \hat{m}_{j1}}} \sum_{\substack{m_{jX}, \hat{m}_{jX} \\ m_{jX} \neq \hat{m}_{jX}}} \sum_{\substack{a^{sj}, \hat{a}^{sj} \\ \hat{a}^{sj} \neq a^{sj}}} \sum_{\substack{b_{jX}, \hat{b}_{jX}}} \sum_{\substack{(u_j^n, x_j^n, y_j^n) \in \\ \tilde{T}(q^n)}} \hat{\theta}(y_j^n | x_j^n) \sum_{\substack{(\hat{u}_j^n, \hat{x}_j^n) \in \\ T_{4\eta_1}(U_j, X_j|y_j^n, q^n)}} \frac{P(M_j = m_j) \exp\{-2nH(X_j|Q)\}}{\theta^{2n+2t_j} \exp\{-4n\eta - 8n\eta_1\}} \mathcal{L}_j(n).$$

There exists $N_{15}(\eta_1) \in \mathbb{N}$ such that for all $n \geq \max\{N_{12}(\eta), N_{15}(\eta_1)\}$, we have

$$|T_{4\eta_1}(U_j, X_j | y_j^n, q^n)| \leq \exp\{n(H(U_j, X_j | Y_j, Q) + 8\eta_1)\} \text{ for all } (y_j^n, q^n) \in T_{2\eta_1}(Y_j, Q)$$

and hence

$$\begin{aligned} P(\epsilon_{l_j}^c \cap \epsilon_{4j}^3) &\leq \sum_{\substack{m_{j1}, \hat{m}_{j1} \\ m_{j1} \neq \hat{m}_{j1}}} \sum_{\substack{m_{jX}, \hat{m}_{jX} \\ m_{jX} \neq \hat{m}_{jX}}} \sum_{\substack{a^{sj}, \hat{a}^{sj} \\ \hat{a}^{sj} \neq a^{sj}}} \sum_{\substack{b_{jX}, \hat{b}_{jX} \\ b_{jX} \neq \hat{b}_{jX}}} \sum_{\substack{(u_j^n, x_j^n) \in \\ T_{2\eta}(U_j, X_j | q^n)}} \frac{\theta^{-2n-2t_j} P(M_j = m_j) \exp\{-2nH(X_j | Q)\}}{\exp\{-n4\eta - n16\eta_1 - nH(X_j, U_j | Y_j, Q)\} \mathcal{L}_j(n)} \\ &\leq 2 \sum_{\substack{m_{j1}, \hat{m}_{j1} \\ m_{j1} \neq \hat{m}_{j1}}} \sum_{\substack{m_{jX}, \hat{m}_{jX} \\ m_{jX} \neq \hat{m}_{jX}}} \frac{\theta^{s_j} P(M_j = m_j) \exp\{-nH(X_j | Q) + n16\eta_1\}}{\theta^{n+t_j} \exp\{-n8\eta - nH(X_j, U_j | Y_j, Q) - nK_j\}} \\ &\leq 2 \frac{\theta^{s_j} P(M_j = m_j) \exp\{-nH(X_j | Q) + n16\eta_1 + nL_j\}}{\theta^n \exp\{-n8\eta - nH(X_j, U_j | Y_j, Q) - nK_j - n\eta_1\}} \\ &\leq 2 \exp\left\{-n \left[\binom{\log \theta + H(X_j | Q) -}{H(X_j, U_j | Y_j, Q)} - \binom{K_j + L_j +}{S_j \log \theta} - \binom{9\eta_1 + 16\eta_1}{+\log \theta \eta_1} \right]\right\} \\ &\leq 2 \exp\{-n(\delta - (9\eta + 16\eta_1))\}. \end{aligned} \quad (75)$$

We now collect all the upper bounds derived in (70), (72) and (75). For $n \geq \max\{N_{14}(\eta), N_{16}(\eta)\}$, we have

$$P((\tilde{\epsilon}_1 \cup \epsilon_3)^c \cap \epsilon_{4j}) \leq 10 \exp\{-n(\delta - (9\eta + 16\eta_1))\} \quad (76)$$

APPENDIX E

PROOF OF LEMMA 3

We prove this by contradiction. Suppose $(h_b(\tau_1 * \delta_1) - h_b(\delta_1), h_b(\tau * \delta) - h_b(\delta), h_b(\tau * \delta) - h_b(\delta)) \in \text{coel}(\alpha_f^{3-1}(p_{QU_2U_3XY}))$ for some $p_{QU_2U_3XY} \in \mathbb{D}_{3-1}(\tau_1, \tau, \tau)$. In the sequel, we characterize such a $p_{QU_2U_3XY}$ and employ the same to derive a contradiction. 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 (1) we have

$$\begin{aligned} R_j &\leq I(U_j X_j; Y_j | Q) = H(Y_j | Q) - H(Y_j | X_j U_j Q) = H(Y_j | Q) - h_b(\delta) = \sum_{q \in \mathcal{Q}} p_Q(q) H(Y_j | Q = q) - h_b(\delta) \\ &= \sum_{q \in \mathcal{Q}} p_Q(q) H(X_j \oplus N_j | Q = q) - h_b(\delta) \text{ for } j = 2, 3. \end{aligned} \quad (77)$$

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 (77), we have

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

from Jensen's inequality. Since $p_{X_j}(1) \leq \tau < \frac{1}{2}$, we have $p_{X_j}(1)(1-2\delta) + \delta \leq \tau(1-2\delta) + \delta < \frac{1}{2}(1-2\delta) + \delta = \frac{1}{2}$.²²

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 $h_b(\tau * \delta)$.²³ Moreover,

²²Here we have used the positivity of $(1-2\delta)$, or equivalently δ being in the range $(0, \frac{1}{2})$.

²³This is consequence of $p_{X_j}(1) \leq \tau$.

the 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 (1). As in proof of theorem 4, 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 and the proof of the first claim presented above. Note that

$$\begin{aligned}
R_1 &\leq I(X_1; Y_1 | \tilde{Q}) = H(Y_1 | \tilde{Q}) - H(Y_1 | \tilde{Q}, X_1) \\
&= \sum_{\tilde{q}} p_{\tilde{Q}}(\tilde{q}) H(X_1 \oplus N_1 \oplus (X_2 \vee X_3) | \tilde{Q} = \tilde{q}) - \sum_{x_1, \tilde{q}} p_{X_1 \tilde{Q}}(x_1, \tilde{q}) H(N_1 \oplus (X_2 \vee X_3) | \tilde{Q} = \tilde{q}) \quad (78) \\
&= \sum_{\tilde{q}} p_{\tilde{Q}}(\tilde{q}) H(X_1 \oplus N_1 \oplus (X_2 \vee X_3) | \tilde{Q} = \tilde{q}) - \sum_{\tilde{q}} p_{\tilde{Q}}(\tilde{q}) H(N_1 \oplus (X_2 \vee X_3) | \tilde{Q} = \tilde{q}) \\
&\leq \sum_{\tilde{q}} p_{\tilde{Q}}(\tilde{q}) H(X_1 \oplus N_1 | \tilde{Q} = \tilde{q}) - \sum_{\tilde{q}} p_{\tilde{Q}}(\tilde{q}) H(N_1 | \tilde{Q} = \tilde{q}) = \sum_q p_{\tilde{Q}}(\tilde{q}) H(X_1 \oplus N_1 | \tilde{Q} = \tilde{q}) - h_b(\delta_1) \quad (79) \\
&= \sum_{\tilde{q}} p_{\tilde{Q}}(\tilde{q}) h_b(\tau_{1\tilde{q}} * \delta_1) - h_b(\delta_1) \leq h_b(\mathbb{E}_{\tilde{Q}}[\tau_{1\tilde{q}} * \delta_1]) - h_b(\delta_1) = h_b(p_{X_1}(1) * \delta_1) - h_b(\delta_1), \quad (80)
\end{aligned}$$

where (i) (79) 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 1, (iii) the first inequality in (80) 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 increasing²⁴ in $p_{X_1}(1)$ and is bounded above by $h_b(\tau_1 * \delta_1)$.²⁵ 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.²⁶

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., $H(X_2 | Q, U_2) = H(X_3 | Q, U_3) = 0$. In this case, the upper bound on $R_1 + R_2 + R_3$ in (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.²⁷ The hypothesis (14) therefore precludes $(h_b(\tau_1 * \delta_1) - h_b(\delta_1), h_b(\tau * \delta) - h_b(\delta), h_b(\tau * \delta) - h_b(\delta)) \in \alpha_f^{3-1}(p_{QU_2U_3XY})$ if $H(X_2 | Q, U_2) = H(X_3 | Q, U_3) = 0$. This proves our third claim.

²⁴This also employs the positivity of $1 - 2\delta_1$, or equivalently δ_1 being in the range $(0, \frac{1}{2})$.

²⁵This is consequence of $p_{X_1}(1) \leq \tau_1$.

²⁶We 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 .

²⁷ $\beta := (1 - \tau)^2\delta_1 + (2\tau - \tau^2)(1 - \delta_1)$ is as defined in the statement of the lemma.

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 $H(X_2 | Q, U_2) > 0$. 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 $p_{X_2 | U_2 Q}(x_2 | u_2^*, q^*) \notin \{0, 1\}$ for each $x_2 \in \{0, 1\}$.

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

$$0 < 1 - \tau = p_{X_3 | Q}(0 | q^*) = \sum_{u_3 \in \mathcal{U}_3} p_{X_3 U_3 | Q}(0, u_3 | q^*).$$

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

We have therefore identified $(q^*, u_2^*, u_3^*) \in \mathcal{Q} \times \mathcal{U}_2 \times \mathcal{U}_3$ such that $p_Q(q^*) > 0$, $p_{U_2 | Q}(u_2^* | q^*) > 0$, $p_{U_3 | Q}(u_3^* | q^*) > 0$, $p_{X_2 | U_2 Q}(x_2 | u_2^*, q^*) \notin \{0, 1\}$ for each $x_2 \in \{0, 1\}$ and $1 \geq p_{X_3 | U_3 Q}(0 | u_3^*, q^*) > 0$. By conditional independence of the pairs (X_2, U_2) and (X_3, U_3) given Q , we also have $p_{X_2 | U_2 U_3 Q}(x_2 | u_2^*, u_3^*, q^*) \notin \{0, 1\}$ for each $x_2 \in \{0, 1\}$ and $1 \geq p_{X_3 | U_2 U_3 Q}(0 | u_2^*, u_3^*, q^*) > 0$. The reader may now verify $p_{X_2 \vee X_3 | U_2 U_3 Q}(x | u_2^*, u_3^*, q^*) \notin \{0, 1\}$ for each $x \in \{0, 1\}$. Since $p_{Q U_2 U_3}(q^*, u_2^*, u_3^*) = p_Q(q^*) p_{U_2 | Q}(u_2^* | q^*) p_{U_3 | Q}(u_3^* | q^*) > 0$, 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 herein, or in the proof of theorem 4. 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

$$\begin{aligned} R_1 &\leq I(X_1; Y_1 | \tilde{Q}) = \sum_{\tilde{q}} p_{\tilde{Q}}(\tilde{q}) H(X_1 \oplus N_1 \oplus (X_2 \vee X_3) | \tilde{Q} = \tilde{q}) - \sum_{\tilde{q}} p_{\tilde{Q}}(\tilde{q}) H(N_1 \oplus (X_2 \vee X_3) | \tilde{Q} = \tilde{q}) \\ &< \sum_{\tilde{q}} p_{\tilde{Q}}(\tilde{q}) H(X_1 \oplus N_1 | \tilde{Q} = \tilde{q}) - \sum_{\tilde{q}} p_{\tilde{Q}}(\tilde{q}) H(N_1 | \tilde{Q} = \tilde{q}) = \sum_{\tilde{q}} p_{\tilde{Q}}(\tilde{q}) H(X_1 \oplus N_1 | \tilde{Q} = \tilde{q}) - h_b(\delta_1) \quad (81) \\ &= \sum_{\tilde{q}} p_{\tilde{Q}}(\tilde{q}) h_b(\tau_{1\tilde{q}} * \delta_1) - h_b(\delta_1) \leq h_b(\mathbb{E}_{\tilde{Q}} \{\tau_{1\tilde{q}} * \delta_1\}) - h_b(\delta_1) = h_b(p_{X_1}(1) * \delta_1) - h_b(\delta_1), \quad (82) \end{aligned}$$

where (i) (81) 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 1, (iii) the first inequality in (82) 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 increasing²⁸ in $p_{X_1}(1)$ and is bounded above by $h_b(\tau_1 * \delta_1)$. We therefore have $R_1 < h_b(\tau_1 * \delta_1) - h_b(\delta_1)$.

APPENDIX F

PROOF OF LEMMA 5

The proof here closely mimics proof of lemma 3. In fact, we allude to appendix E to avoid restating certain elements.

²⁸This also employs the positivity of $1 - 2\delta_1$, or equivalently δ_1 being in the range $(0, \frac{1}{2})$.

We assume $\underline{C}^* \in \alpha_u(\underline{\tau})$, and derive a contradiction. Suppose $\underline{C}^* \in \text{cocl}(\alpha_u(p_{QU_2U_3XY}))$ for some $p_{QU_2U_3XY} \in \mathbb{D}_u(\underline{\tau})$ ²⁹. In the sequel, we characterize such a $p_{QU_2U_3XY}$ and employ the same to derive a contradiction. Our first claim, as in appendix E, is $p_{X_j|Q}(1|q) = \tau$ for $j = 2, 3$ and every $q \in \mathcal{Q}$. Since the corresponding arguments in appendix E hold verbatim, we allude to the same for a proof of this claim. We conclude the triplet $(Q, X_1), X_2, X_3$ to be mutually independent, and in particular X_1, X_2, X_3 to be mutually independent. This enables us conclude that for any $p_{QU_2U_3XY} \in \mathbb{D}_u(\underline{\tau})$ for which $\underline{C}^* \in \text{cocl}(\alpha_u(p_{QU_2U_3XY}))$, we have it's corresponding marginal $p_{XY} \in \mathcal{D}(\underline{\tau})$.

Our second claim is $p_{X_1|Q}(1|q) = p_{X_1}^*(1)$ for every $q \in \mathcal{Q}$ for which $p_Q(q) > 0$. We begin with the upper bound on R_1 in (1). Denoting

$I(p_{A|C}(\cdot|c); p_{B|A,C}(\cdot|c)) := I(A; B|C = c)$ for any random variables A, B, C , we have,

$$I(X_1; Y_1|Q, U_2, U_3) \leq I(X_1; Y_1|Q, X_2 \vee X_3) \quad (83)$$

$$\begin{aligned} &= \sum_s p_{X_2 \vee X_3}(s) \sum_q p_{Q|X_2 \vee X_3}(q|s) I(p_{X_1|Q, X_2 \vee X_3}(\cdot|q, s); p_{Y_1|X_1, Q, X_2 \vee X_3}(\cdot|q, s)) \\ &= \sum_s p_{X_2 \vee X_3}(s) \sum_q p_{Q|X_2 \vee X_3}(q|s) I(p_{X_1|Q, X_2 \vee X_3}(\cdot|q, s); p_{Y_1|X_1, X_2 \vee X_3}(\cdot|q, s)) \end{aligned} \quad (84)$$

$$\leq \sum_s p_{X_2 \vee X_3}(s) I\left(\sum_q p_{Q|X_2 \vee X_3}(q|s) p_{X_1|Q, X_2 \vee X_3}(\cdot|q, s); p_{Y_1|X_1, X_2 \vee X_3}(\cdot|q, s)\right) \quad (85)$$

$$= \sum_s p_{X_2 \vee X_3}(s) I(p_{X_1|X_2 \vee X_3}(\cdot|s); p_{Y_1|X_1, X_2 \vee X_3}(\cdot|s)) = I(X_1; Y_1|X_2 \vee X_3) \leq C_1 \quad (86)$$

where (i) (83) follows from the Markov chains $(U_2, U_3) - (X_2 \vee X_3) - Y_1$ and $(U_2, U_3) - (X_1, X_2 \vee X_3) - Y_1$, (ii) (84) follows from the Markov chain $Q - X_1, X_2 \vee X_3 - Y_1$ resulting from the nature of the channel from the inputs to Y_1 , (iii) (85) follows from Jensen's inequality, and (iv) (86) follows from $p_{XY} \in \mathcal{D}(\underline{\tau})$ and definition of C_1 . The strict concavity of $I(p_{A|C}(\cdot|c); p_{B|A,C}(\cdot|c))$ in $p_{A|C}(\cdot|c)$ implies equality holds in (85) if and only if $p_{X_1|Q, X_2 \vee X_3}(1|q, s) = p_{X_1|Q}(1|q)$ is invariant with q for every $q \in \mathcal{Q}$ for which $p_{Q|X_2 \vee X_3}(q|s) = p_Q(q) > 0$.³⁰ By the uniqueness of p_{XY}^* , and in particular $p_{X_1}^*$, we conclude $p_{X_1|Q}(1|q) = p_{X_1}^*(1)$ for every $q \in \mathcal{Q}$ for which $p_Q(q) > 0$.

Our first and second claims imply that if $\underline{C}^* \in \text{cocl}(\alpha_u(p_{QU_2U_3XY}))$ for some $p_{QU_2U_3XY} \in \mathbb{D}_u(\underline{\tau})$, then $\sum_{q, u_2, u_3} p_{QU_2U_3XY}(q, u_2, u_3, \underline{x}, \underline{y}) = p_{XY}^*(\underline{x}, \underline{y}) \in \mathcal{D}(\underline{\tau})$, and furthermore, Q is independent of \underline{X} . We therefore reiterate that any entropy or mutual information terms involving random variables in $\underline{X}, \underline{Y}$, stated in the sequel, is evaluated with respect to p_{XY}^* .

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., $H(X_2|Q, U_2) = H(X_3|Q, U_3) = 0$. In this case, the upper bound on $R_1 + R_2 + R_3 = C_1 + 2(h_b(\tau * \delta) - h_b(\delta))$ in (2) is

$$R_1 + R_2 + R_3 = C_1 + 2(h_b(\tau * \delta) - h_b(\delta)) \leq I(X_2, X_3, X_1; Y_1|Q) = I(\underline{X}; Y_1) \quad (87)$$

²⁹Recall $\underline{\tau} := (\tau_1, \tau, \tau)$.

³⁰We have proved in our first claim Q and (X_2, X_3) are independent.

where the last equality follows from independence of Q and \underline{X} and thereby implying independence of Q and $(\underline{X}, \underline{Y})$. (87) contradicts the hypothesis (19) of the lemma.

Our fourth claim is $H(X_2 \vee X_3 | Q, U_2, U_3) > 0$. The proof of this claim is identical to the proof of the corresponding claim in appendix E and the reader is alluded to the same. As a consequence of $H(X_2 \vee X_3 | \tilde{Q}) > 0$, where $\tilde{Q} := (Q, U_2, U_3)$, there exists $\tilde{q}^* := (q^*, u_2^*, u_3^*) \in \tilde{\mathcal{Q}} := \mathcal{Q} \times \mathcal{U}_2 \times \mathcal{U}_3$ for which $p_{\tilde{Q}}(\tilde{q}^*) > 0$ and $H(X_2 \vee X_3 | \tilde{Q} = \tilde{q}^*) > 0$.

Our fifth claim and final claim is that $H(X_2 \vee X_3 | Q, U_2, U_3) > 0$ implies $C_1 < I(X_1; Y_1 | X_2 \vee X_3)$ thereby contradicting the definition of C_1 (17). The reader will recognize that our proof for the fifth claim in appendix E *cannot* be employed here. We employ a more powerful technique that we will have opportunity to use in our study of example 7. The upper bound (1) on R_1 implies

$$\begin{aligned} C_1 = R_1 &\leq I(X_1; Y_1 | \tilde{Q}) = \sum_{\tilde{q}} p_{\tilde{Q}}(\tilde{q}) I(p_{X_1 | \tilde{Q}}(\cdot | \tilde{q}); p_{Y_1 | X_1 \tilde{Q}}(\cdot | \cdot, \tilde{q})) \\ &= \sum_{\tilde{q}} p_{\tilde{Q}}(\tilde{q}) I \left(p_{X_1 | \tilde{Q}}(\cdot | \tilde{q}); \sum_s p_{Y_1 | X_1, X_2 \vee X_3, \tilde{Q}}(\cdot | \cdot, s, \tilde{q}) p_{X_2 \vee X_3 | \tilde{Q}}(s | \tilde{q}) \right) \\ &< \sum_{\tilde{q}} p_{\tilde{Q}}(\tilde{q}) \sum_s p_{X_2 \vee X_3 | \tilde{Q}}(s | \tilde{q}) I \left(p_{X_1 | \tilde{Q}}(\cdot | \tilde{q}); p_{Y_1 | X_1, X_2 \vee X_3, \tilde{Q}}(\cdot | \cdot, s, \tilde{q}) \right) \end{aligned} \quad (88)$$

$$= \sum_{s, \tilde{q}} p_{\tilde{Q}, X_2 \vee X_3}(\tilde{q}, s) I(p_{X_1}(\cdot); p_{Y_1 | X_1, X_2 \vee X_3}(\cdot | \cdot, s)) \quad (89)$$

$$= \sum_{s, \tilde{q}} p_{\tilde{Q}, X_2 \vee X_3}(\tilde{q}, s) I(p_{X_1 | X_2 \vee X_3}(\cdot | s); p_{Y_1 | X_1, X_2 \vee X_3}(\cdot | \cdot, s)) = I(X_1; Y_1 | X_2 \vee X_3) \leq C_1, \quad (90)$$

where (i) (88) follows from strict convexity of the mutual information in the conditional distribution (channel transition probabilities), the presence of $\tilde{q}^* \in \tilde{\mathcal{Q}}$ for which $p_{X_2 \vee X_3 | \tilde{Q}}(\cdot | \tilde{q}^*)$ is non-degenerate and $p_{Y_1 | X_1, X_2 \vee X_3, \tilde{Q}}(\cdot | \cdot, s, \tilde{q}^*)$ distinct, (ii) (89) follows from conditional independence of X_1 and (U_2, U_3) given Q , the second claim above, and the Markov chain $\tilde{Q} - X_1, X_2 \vee X_3 - Y_1$ induced by the nature of the channel, and (iii) (90) follows from X_1, X_2, X_3 being mutually independent, $p_{\underline{X}\underline{Y}} \in \mathcal{D}(\underline{\mathcal{T}})$ and the definition of C_1 . We have thus derived a contradiction $C_1 < C_1$.

REFERENCES

- [1] C. E. Shannon, "Two-way communication channels," Proc. 4th Berkeley Symp. Mathematical Statistics and Probability, vol. 1, pp. 611–644, 1961.
- [2] 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.
- [3] H. Sato, "Two-user communication channels," Information Theory, IEEE Transactions on, vol. 23, no. 3, pp. 295–304, 1977.
- [4] A. Carleial, "A case where interference does not reduce capacity (corresp.)," Information Theory, IEEE Transactions on, vol. 21, no. 5, pp. 569–570, 1975.
- [5] T. M. Cover, "Broadcast channels," IEEE Trans. Inform. Theory, vol. IT-18, no. 1, pp. 2–14, Jan. 1972.
- [6] 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.
- [7] 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.

- [8] 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.
- [9] 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.
- [10] 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.
- [11] A. Jafarian and S. Vishwanath, "Achievable rates for k-user gaussian interference channels," submitted to IEEE Trans. of Information theory 2011, available at <http://arxiv.org/abs/1109.5336>.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] 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>.
- [16] V. R. Cadambe and S. A. Jafar, "Interference alignment and a noisy interference regime for many-to-one interference channels," available at <http://arxiv.org/abs/0912.3029>.
- [17] 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.
- [18] H.-F. Chong, M. Motani, and H. K. Garg, "A comparison of two achievable rate regions for the interference channel," San Diego, Feb. 2006.
- [19] G. Kramer, "Review of rate regions for interference channels," in Communications, 2006 International Zurich Seminar on, 2006, pp. 162–165.
- [20] K. Kobayashi and T. Han, "A further consideration on the HK and the CMG regions for the interference channel," San Diego, Feb. 2007.
- [21] B. Bandemer and A. El Gamal, "Interference decoding for deterministic channels," Information Theory, IEEE Transactions on, vol. 57, no. 5, pp. 2966–2975, 2011.
- [22] 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.
- [23] 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.
- [24] H. S. Witsenhausen, "On sequences of pairs of dependent random variables," SIAM Journal of Applied Mathematics, vol. 28, no. 1, pp. 100–113, January 1975.
- [25] M. Hall, The theory of groups. New York: Macmillan, 1959.
- [26] 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>.
- [27] R. G. Gallager, Information Theory and Reliable Communication. New York: John Wiley & Sons, 1968.
- [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.