Structured Mappings and Conferencing Common Information for Multiple-Access Channels

Mohsen Heidari and S. Sandeep Pradhan, *Senior Member, IEEE*

***Abstract*— In this work, we study two problems: three-user Multiple-Access Channel (MAC) with correlated sources, and MAC with Feedback (MAC-FB) with independent messages. For the first problem, we identify a structure in the joint probability distribution of discrete memoryless sources, and define a new common information called "conferencing common information". We develop a multi-user joint-source channel coding methodology based on structured mappings to encode this common information efficiently and to transmit it over a MAC. We derive a new set of sufficient conditions for this coding strategy using single-letter information quantities for arbitrary sources and channel distributions. Next, we make a fundamental connection between this problem and the problem of communication of independent messages over three-user MAC-FB. In the latter problem, although the messages are independent to begin with, they become progressively correlated given the channel output feedback. Subsequent communication can be modeled as transmission of correlated sources over MAC. Exploiting this connection, we develop a new coding scheme for the problem. We characterize its performance using single-letter information quantities, and derive an inner bound to the capacity region. For both problems, we provide a set of examples where these rate regions are shown to be optimal. Moreover, we analytically prove that this performance is not achievable using random unstructured random mappings/codes.**

***Index Terms*— MAC with feedback (MAC-FB), MAC with correlated sources, joint-source channel coding, structured codes.**

I. Introduction

MANY coding strategies for processing/transmitting sources of information in a distributed fashion harness structures in the statistical description of the sources. Common information/randomness can be viewed as an example of such a structure. Efforts in finding a measure of common information among distributed sources led to several definitions [1]–[4]. A noteworthy definition of common information is due to Gács and Körner [1] and Witsenhausen [2], which is an information-theoretic measure of the amount of common randomness that can be extracted from two sources. Gács-Körner-Witsenhausen (GKW) *common part* between two correlated memoryless sources (S_1, S_2) is defined as a random variable W with the largest entropy, for which there exist functions f, g such that $W = f(S_1) = g(S_2)$ with probability one. The random variable $f(S_1)$ (or equivalently $g(S_2)$) represents the "common randomness" generated from the sources, and the functions (f, g) represent the extraction process applied on the sources.

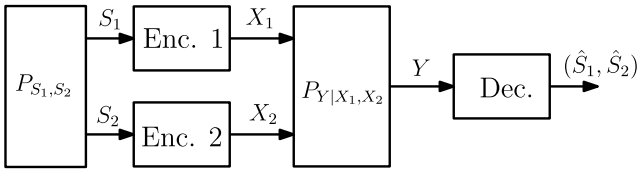

Fig. 1. A schematic of a two-user MAC with correlated sources. In this setup, the source sequences (S_1^n, S_2^n) are observed by the corresponding encoders. The encoders produce (X_1^n, X_2^n) which are channel input sequences. Upon observing the channel output Y^n, the decoder produces an estimate for the sources.

GKW common part has been found useful in many problems such as transmission of distributed sources over channels [5]–[8] and distributed key generation [9]. In MAC with correlated sources, as shown in Figure 1, there are multiple transmitters, each observing a source, and the sources are correlated with each other. The transmitters wish to send their observations in a distributed fashion via a MAC to a central receiver. The receiver reconstructs the sources losslessly. Cover-El Gamal-Salehi (CES) showed that joint source-channel coding outperforms separation-based coding approaches [10], [11]. This was done by introducing a novel transmission scheme [7], which exploits the common information between the sources. In this scheme, GKW common part between the sources is first extracted distributively at the encoders. The encoders can effectively 'fully cooperate' to send this information to the receiver, as it is done in Point-to-Point (PtP) joint source-channel coding problem. The rest of the sources are transmitted using distributed unstructured random mappings. In summary, it employs a two-stage encoding strategy. CES also characterized a set of sufficient conditions, in terms of single-letter information quantities, for transmission of sources over a MAC. The scheme is known to be suboptimal [12] in general. There are a set of necessary conditions developed in [13] and [14]. However, characterizing the optimal necessary and sufficient conditions for transmission of discrete memoryless sources over MAC is still an open problem.

Manuscript received May 9, 2019; revised December 23, 2019; accepted February 21, 2020. Date of publication March 13, 2020; date of current version June 18, 2020. This work was supported by the NSF under Grant CCF 1717299. This article was presented in part at the 2016 IEEE International Symposium on Information Theory and in part at the 2017 IEEE International Symposium on Information Theory. *(Corresponding author: Mohsen Heidari.)*

Mohsen Heidari is with the Department of Computer Sciences, Purdue University, West Lafayette, IN 47907 USA (e-mail: mheidari@purdue.edu).

S. Sandeep Pradhan is with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109 USA (e-mail: pradhanv@umich.edu).

Communicated by N. Merhav, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2020.2980550

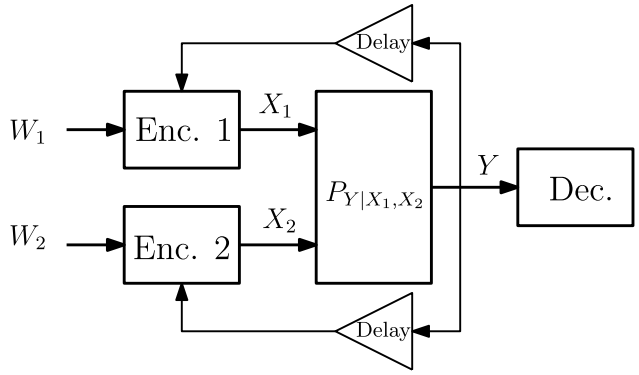

Fig. 2. A schematic of a two-user MAC with feedback setup. The output of the channel is available, with one unit of delay, to the transmitters.

Another fundamental problem in which common information plays a key role is communication of *independent* messages over discrete memoryless MAC-FB. In a MAC-FB setup (see Figure 2), after each channel use, the output of the channel is received at each transmitter noiselessly. This problem has been studied extensively in the literature [15]–[22]. Gaarder and Wolf [15] showed that feedback can expand the capacity region of discrete memoryless MAC. Cover-Leung (CL) [16] studied two-user MAC-FB, developed a coding strategy using unstructured random codes, and characterized an achievable rate region in terms of single-letter information quantities. Later, it was shown by Willems [19] that the CL scheme achieves the feedback capacity for a class of MAC-FB. However, this is not the case for general MAC-FB [22]. There are several improvements over CL achievable region, namely [23] and [18]. A multi-letter characterization of the feedback-capacity of MAC-FB is given by Kramer [17]. However, the characterization is not computable, since it is an infinite-letter characterization. Finding a computable characterization of the capacity region remains an open problem.

The main idea behind CL coding scheme is explained in the following. The scheme operates in two stages. In stage one, the transmitters send the messages with rates that lie outside the no-feedback capacity region (i.e. higher rates than what is achievable without feedback). The transmission rates are taken such that each user can decode the other user's message using feedback. In this stage, the receiver is unable to decode the messages reliably; however, is able to form a list of "highly likely" pairs of messages. The transmitters can also recreate this list. In the second stage, the encoders fully cooperate to send the index of the correct message-pair in the list, and help the receiver decode it.

There is a connection between CES scheme for transmission of correlated sources over MAC and CL scheme for communications over MAC-FB. In a MAC-FB setup, after multiple uses of the channel, conditioned on feedback, the messages become statistically correlated. As explained above, at the end of the first stage in CL scheme, the messages are decoded at the transmitters. Hence, the decoded messages can be viewed as a GKW common part available at the two transmitters after the first stage. This common part is used in the second stage to resolve the uncertainty of the receiver. In connection with CES scheme, the common part is transmitted using identical random unstructured codebooks.

In this work, we study three-user MAC with correlated sources, and three-user MAC-FB with independent messages. Motivated by the notion of common information and its imperative role in these problems, we start by identifying common information among a triplet of sources (say S_1, S_2, S_3). One can extend GKW common part to define a (mutual) common part for (S_1, S_2, S_3) in a straightforward way. In addition, one can define the pairwise GKW common parts between any pair (S_i, S_j) as a part of the common information. The mutual common part together with the pairwise common parts characterize a vector of four components of common information which we refer to as *univariate common parts*.

We make the following contributions in this work. We, first, identify a new additional structure in the joint probability distribution of the sources, called "conferencing common part". This common part can be viewed as the GKW common part between a source (say S_1) and a pair of sources (say S_2, S_3). More explicitly, it is defined as the random variable T with the largest entropy for which there exist a function $f(\cdot)$ and a bivariate function $g(\cdot,\cdot)$ such that $T = f(S_1) = g(S_2, S_3)$ with probability one. Therefore, for the triplet (S_1, S_2, S_3), there are three conferencing common parts, one between each source and the other pair. We also refer to these as bivariate common parts. Hence, in total, we identify the common parts among a triplet of the sources as a vector of seven components, including four univariate and three conferencing (bivariate) common parts.

Next, we develop a new coding strategy to exploit a particular form of the conferencing common parts among the sources, one given by additive functions. Efficient encoding of conferencing common parts is a more challenging task as compared to the univariate ones — which is done using identical random unstructured mappings/codebooks. This is because conferencing common parts are not available at any one transmitter— rather a conference among a subset of the users is needed to extract these common parts. We develop a multiuser joint-source channel coding methodology based on structured mappings to encode these common parts efficiently to be transmitted over a MAC.

In particular, we design coding strategies based on random structured mappings for three-user MAC with correlated sources and MAC-FB. For the former problem, our coding strategy exploits the univariate and the conferencing common information among the sources. We derive a new set of sufficient conditions for this coding strategy using single-letter information quantities for arbitrary sources and channel distributions. For the latter problem, based on our notion for common information, we develop a new coding scheme for communications over three-user MAC-FB with independent messages. We characterize its performance using single-letter information quantities and derive an inner bound to the capacity region. For both problems we provide a set of examples, where these rate regions are shown to be optimal. Moreover, we analytically prove that this performance is not achievable

using random unstructured mappings/codes. The main results of this paper are given in Proposition 2 and Theorem 1-4.

Prior works on structured codes for multiuser problems: Structured codes have been used in many problems involving either source coding or channel coding. For example, they have been used in distributed source coding [24]–[27], computation over MAC [28]–[31], MAC with side information [25], [32]–[35], interference channels [36]–[41], and broadcast channels [42].

Notations: In this paper, random variables are denoted using capital letters such as X, Y, and their realizations are shown using lower case letters such as x, y, respectively. Vectors are shown using lowercase bold letters such as $\mathbf{x}, \mathbf{y}$. Calligraphic letters are used to denote sets such as $\mathcal{X}, \mathcal{Y}$. For any set $\mathcal{A}$, let $S_{\mathcal{A}} = \{S_a\}_{a\in\mathcal{A}}$. If $\mathcal{A} = \emptyset$, then $S_{\mathcal{A}} = \emptyset$. As a shorthand, we sometimes denote a triple (s_1, s_2, s_3) by $\underline{s}$. We also denote a triple of sequences $(\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3)$ by $\underline{\mathbf{s}}$. Binary entropy function is denoted by $h_b(\cdot)$. By $\mathbb{F}_q$, we denote the field of integers modulo-q, where q is a prime number. Modulo-q addition is denoted by $\oplus_q$, and, when it is clear from the context, the subscript q is removed. For any mapping $\Phi : \mathcal{A} \mapsto \mathcal{B}$ and any integer n, define the mapping $\Phi^n : \mathcal{A}^n \mapsto \mathcal{B}^n$ such that $\Phi^n(a^n) \triangleq (\Phi(a_1), \Phi(a_2), ..., \Phi(a_n))$ for all $a^n \in \mathcal{A}^n$. Given a probability distribution P_X on a finite alphabet $\mathcal{X}$, let $A_\epsilon^{(n)}(X)$ denote the set of strongly ϵ-typical sequences of length n. We follow the definition of typical sequences as given in [43], [44].

The rest of the paper is organized as follows: Section II contains problem formulation and known results for MAC with correlated sources. We present our contributions for this problem in Section III. Similarly, we present the problem formulation and known results for MAC-FB in Section IV, and provide our contributions for this problem in V. Lastly, Section VI concludes the paper.

II. Transmission of Sources Over MAC: Preliminaries

A. Problem Formulation

As depicted in Figure 1, the problem of MAC with correlated sources consists of multiple transmitters, each observing a source sequence statistically correlated to others. The source sequences are sent by the encoders via a MAC to a central decoder. The objective of the receiver is to reconstruct the source sequences losslessly. It is assumed that the channel is a discrete memoryless MAC and the source sequences are discrete and generated IID according to a known joint PMF. In what follows, we formulate this problem more precisely.

Definition 1: A discrete memoryless MAC with 3 users is defined by input alphabet $\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3$, output alphabet $\mathcal{Y}$, and a transition probability matrix $P_{Y|X_1,X_2,X_3}$. The input and output alphabets are assumed to be finite sets. The MAC is denoted by the triple $(\underline{\mathcal{X}}, \mathcal{Y}, P_{Y|\underline{X}})$.

We assume that the channel is memoryless, stationary and used without feedback, and, hence, the transition probability of the n-length channel output vector given the n-length channel input vectors is given by

$$\prod_{i=1}^{n} P_{Y|X_1X_2X_3}(y_i|x_{1i}, x_{2i}, x_{3i}),$$

for all $\underline{\mathbf{x}} \in \underline{\mathcal{X}}^n$ and $\mathbf{y} \in \mathcal{Y}^n$.

Definition 2: A discrete memoreless stationary source (S_1, S_2, S_3) is defined by alphabet $\mathcal{S}_1 \times \mathcal{S}_2 \times \mathcal{S}_3$ and a distribution P_{S_1,S_2,S_3}. The source is denoted by the pair $(\underline{S}, P_{\underline{S}})$

The distribution of n-length source sequences is given by

$$\prod_{i=1}^{n} P_{S_1S_2S_3}(s_{1i}, s_{2i}, s_{3i}),$$

for all $\underline{\mathbf{s}} \in \underline{\mathcal{S}}^n$.

In this paper, the bandwidth expansion factor is assumed to be unity, i.e., the channel is used n times for transmission of n samples of the sources.

Definition 3: A coding scheme (without bandwidth expansion) with parameter n for transmission of a source $(\underline{S}, P_{\underline{S}})$ over a MAC $(\underline{\mathcal{X}}, \mathcal{Y}, P_{Y|\underline{X}})$ consists of encoding functions $e_i : \mathcal{S}_i^n \to \mathcal{X}_i^n, i = 1, 2, 3,$ and a decoding function $d : \mathcal{Y}^n \to \mathcal{S}_1^n \times \mathcal{S}_2^n \times \mathcal{S}_3^n$. The parameter n is called block-length.

Definition 4: A source $(\underline{S}, P_{\underline{S}})$ is said to be transmissible over a MAC $(\underline{\mathcal{X}}, \mathcal{Y}, P_{Y|\underline{X}})$, if for all $\epsilon > 0$ and for all sufficiently large n, there exists a coding scheme with parameter n such that

$$\sum_{\underline{\mathbf{s}}\in\underline{\mathcal{S}}^n} P_{\underline{S}}^n(\underline{\mathbf{s}}) \sum_{\mathbf{y}:d(\mathbf{y})\neq\underline{\mathbf{s}}} P_{Y|\underline{X}}^n\Big(\mathbf{y} \mid \mathbf{x}_i = e_i(\mathbf{s}_i),\ i = 1, 2, 3\Big) \leq \epsilon.$$

B. CES Sufficient Conditions: Two-User Case

The two-user version of MAC with correlated sources was investigated in [7] and CES scheme was proposed based on unstructured random mappings. Further, a sufficient condition for transmissibility is derived in terms of single-letter information quantities. In this scheme the notion of GKW *common part* plays an important role. The formal definition of such common part and the CES sufficient conditions are given below.

Definition 5 (GKW Common part): A *common part* between random variables (S_1, S_2) is a random variable W_{12} with the largest entropy for which there exist functions f, g such that $W_{12} = f(S_1)$, and $W_{12} = g(S_2)$ with probability one. In this work, such a random variable W_{12} is called a univariate common part.

Fact 1 (CES sufficient conditions): A source $(\mathcal{S}_1, \mathcal{S}_2, P_{S_1S_2})$ is transmissible over a MAC $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}, P_{Y|X_1X_2})$, if there exist distributions $P_{U_{12}}, P_{X_1|S_1,U_{12}}$ and $P_{X_2|S_2,U_{12}}$ such that,

$$\begin{aligned} H(S_1|S_2) &\leq I(X_1; Y|X_2, S_2, U_{12}), \\ H(S_2|S_1) &\leq I(X_2; Y|X_1, S_1, U_{12}), \\ H(S_1, S_2|W_{12}) &\leq I(X_1X_2; Y|W_{12}, U_{12}), \\ H(S_1, S_2) &\leq I(X_1\ X_2; Y), \end{aligned}$$

where, U_{12} is an auxiliary random variable with a finite alphabet $\mathcal{U}_{12}$, and the joint distribution of all the random variables factors as

$$P_{S_1,S_2,U_{12},X_1,X_2,Y} = P_{S_1,S_2}P_{U_{12}}P_{X_1|S_1,U_{12}}P_{X_2|S_2,U_{12}}P_{Y|X_1,X_2}.$$

C. A Sufficient Condition Based on Unstructured Mappings: Three-User Case

One can extend CES sufficient conditions for three-user case based on unstructured random codes. For that, first we need to generalize the definition of GKW common part for more than two random variables.

Definition 6: The common part among random variables (S_1, S_2, S_3) is the random variable W_{123} with the largest entropy for which there exist functions $f_i, i = 1, 2, 3$ such that $W_{123} = f_i(S_i)$ holds with probability one.

It is worth noting that for the triple (S_1, S_2, S_3) there are four common parts namely $(W_{12}, W_{13}, W_{23}, W_{123})$. For the case of multiple sources, say (S_1, S_2, S_3), a similar idea as in CES can be used to encode the univariate common parts. In what follows, we provide an extension of CES scheme to the three-user case based on unstructured random mappings.

Definition 7: Given a source $(\underline{\mathcal{S}}, P_{\underline{S}})$ and a MAC $(\underline{\mathcal{X}}, \mathcal{Y}, P_{Y|X_1X_2X_3})$, let $\mathscr{P}_{CES}$ be the set of conditional distributions $P_{\underline{U},\underline{X}|\underline{S}}$ defined on $\underline{\mathcal{U}} \times \underline{\mathcal{X}}$ which factors as

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

where, with a slight abuse of notation, $\underline{U} \triangleq (U_{123}, U_{12}, U_{13}, U_{23})$ and its alphabet is a finite set denoted by $\underline{\mathcal{U}}$.

Proposition 1: A source $(\underline{\mathcal{S}}, P_{S_1S_2S_3})$ is transmissible over a $(\underline{\mathcal{X}}, \mathcal{Y}, P_{Y|X_1X_2X_3})$, if there exists a conditional distribution $P_{\underline{U},\underline{X}|\underline{S}} \in \mathscr{P}_{CES}$ such that for any distinct $i, j, k \in \{1, 2, 3\}$ and any $\mathcal{B} \subseteq \{12, 13, 23\}$ the following inequalities hold

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

where we have identified $U_{ij} = U_{ji}$ and $W_{ij} = W_{ji}$.

The three-user extension of CES involves three layers of coding. In the first layer W_{123} is encoded at each transmitter to U_{123}. Next, based on the output of the first layer, W_{ij}'s are encoded to U_{ij}. Finally, based on the output of the first and the second layers, S_1, S_2 and S_3 are encoded. Figure 3 shows the random variables involved in the extension of CES.

Outline of the proof: Fix a conditional distribution $P_{\underline{U},\underline{X}|\underline{S}} \in \mathscr{P}_{CES}$. Let the sequence $\mathbf{s}_i \in \mathcal{S}_i^n$ be a realization of the ith source, where $i = 1, 2, 3$.

Codebook Generation: The construction of the codebooks at each transmitter is given below:

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

Fig. 3. The random variables involved in the three-user extension of CES.

1) For each realization $\mathbf{w}_{123}$ of the mutual common part, a sequence $\mathbf{U}_{123}$ is generated randomly according to $\prod_{l\in[1,n]} P_{U_{123}}$. Such a sequence is indexed by $\mathbf{U}_{123}(\mathbf{w}_{123})$.
2) Given $b \in \{12, 13, 23\}$, and for each $\mathbf{u}_{123}$ and $\mathbf{w}_b$, a sequence $\mathbf{U}_b$ is generated randomly according to $\prod_{l\in[1,n]} P_{U_b|W_bU_{123}}$. Such a sequence is indexed by $\mathbf{U}_b(\mathbf{w}_b, \mathbf{u}_{123})$.
3) Given distinct elements $i, j, k \in \{1, 2, 3\}$, any realization $\mathbf{s}_i$ of the source, the common parts $(\mathbf{w}_{123}, \mathbf{w}_{ij}, \mathbf{w}_{ik})$, and the corresponding sequences $\mathbf{U}_{123}(\mathbf{w}_{123}), \mathbf{U}_{ij}(\mathbf{w}_{ij}, \mathbf{U}_{123})$ and $\mathbf{U}_{ik}(\mathbf{w}_{ik}, \mathbf{U}_{123})$, a sequence $\mathbf{X}_i$ is generated randomly according to $\prod_{l\in[1,n]} P_{X_i|S_iU_{123}U_{ij}U_{ik}}$. For shorthand, such a sequence is denoted by $\mathbf{X}_i(\mathbf{s}_i, \mathbf{U}_{123}, \mathbf{U}_{ij}, \mathbf{U}_{ik})$.

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

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

and sends $\mathbf{X}_i(\mathbf{s}_i, \mathbf{U}_{123}, \mathbf{U}_{ij}, \mathbf{U}_{ik})$ over the channel.

Decoding: Upon receiving the channel output sequence $\mathbf{y}$, the decoder finds a unique triple $(\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2, \tilde{\mathbf{s}}_3)$ such that

$$\begin{aligned}
&(\underline{\tilde{\mathbf{s}}}, \tilde{\mathbf{U}}_{123}, \tilde{\mathbf{U}}_{12}, \tilde{\mathbf{U}}_{13}, \tilde{\mathbf{U}}_{23}, \tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2, \tilde{\mathbf{X}}_3, \mathbf{y})\\
&\qquad \in A_\epsilon^{(n)}(\underline{S}, U_{123}, U_{12}, U_{13}, U_{23}, X_1, X_2, X_3, Y),
\end{aligned}$$

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

A decoding error occurs, if no unique $(\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2, \tilde{\mathbf{s}}_3)$ is found. Using a standard argument as in [7], it can be shown that the probability of error can be made sufficiently small for large enough n, if the conditions in Proposition 1 are satisfied.

■

III. Transmission of Sources Over MAC: Structured Mappings

In this section, we provide a new sufficient condition characterized using single-letter information quantities for transmissibility of the sources over MAC using structured mappings. The main results of this section are given in Proposition 2, Theorem 1 and 2.

A. Conferencing Common Information

The joint distribution of triple (S_1, S_2, S_3) also has an additional structure which is not captured by the univariate

common parts defined previously. This will be addressed by defining a new common part as follows.

Definition 8: The *conferencing* common part of a triple of random variables (S_1, S_2, S_3) is the triple of random variables (T_1, T_2, T_3) with the largest joint entropy, for which there exist functions $f_i, g_i, i \in \{1,2,3\}$ such that $T_i = f_i(S_i) = g_i(S_j, S_k)$ hold with probability one for all distinct $i, j, k \in \{1,2,3\}$[1].

From definitions 5 and 8, the common parts among the three random variables (S_1, S_2, S_3) are $(W_{12}, W_{13}, W_{23}, W_{123}, T_1, T_2, T_3)$, where W_{ij} is the pairwise common part between (S_i, S_j), W_{123} is the mutual common part (all in the sense of Definition 5), and (T_1, T_2, T_3) are conferencing common parts (as in Definition 8) among (S_1, S_2, S_3). In this work, we focus on a special class of conferencing common part which is defined as follows.

Definition 9: The additive common part of a triple of random variables (S_1, S_2, S_3) is the triple of random variables (T_1, T_2, T_3) with the largest entropy for which there exist a finite field $\mathbb{F}_q$ and functions $f_i : \mathcal{S}_i \mapsto \mathbb{F}_q, i = 1,2,3$ such that $T_i = f_i(S_i)$ and $T_1 \oplus_q T_2 \oplus_q T_3 = 0$.

The following example provides a triplet of binary sources with additive common part where the associated finite field is $\mathbb{F}_2$.

Example 1: Let S_1, S_2 and S_3 be three Bernoulli random variables. Suppose S_1 and S_2 are independent, with biases p_1 and p_2, respectively, and $S_3 = S_1 \oplus_2 S_2$ with probability one. It is not difficult to show that univariate common parts (the pairwise as well as the mutual) are trivial, i.e., $(W_{12}, W_{13}, W_{23}, W_{123})$ is a constant. As for the conferencing common parts, set $T_i = S_i, i = 1,2,3$. Then (T_1, T_2, T_3) satisfies the conditions in Definition 9 for $q = 2$. Therefore, (T_1, T_2, T_3) is the additive common part of (S_1, S_2, S_3).

Unlike univariate common information, conferencing common parts are not available at any terminal. This is due to the fact that conferencing common parts are bivariate functions of the sources. As a result, to exploit conferencing common information, a new coding technique needs to be developed. For this purpose, we use affine maps. The key concepts are described in the following.

We construct three affine maps for encoding of such common parts. Let $\mathbf{G}$ be a n by n matrix with elements in $\mathbb{F}_q$. We, also, select vectors $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3 \in \mathbb{F}_q^n$ such that $\mathbf{b}_1 \oplus \mathbf{b}_2 \oplus \mathbf{b}_3 = \mathbf{0}$. The additive common parts are encoded as $\mathbf{V}_i^n = \mathbf{T}_i^n \mathbf{G} \oplus \mathbf{b}_i$, for $i = 1,2,3$, and hence, the equality $\mathbf{V}_1^n \oplus \mathbf{V}_2^n \oplus \mathbf{V}_3^n = \mathbf{0}$ holds with probability one. One may adopt a randomized affine map to encode the additive common parts. For that, we can select the matrix $\mathbf{G}$ and the vectors $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$ randomly and uniformly from the set of all matrices and vectors with elements in $\mathbb{F}_q$.

B. Sub-Optimality of Unstructured Mappings

In what follows, we show that applications of affine maps for transmission of additive common parts improves upon the scheme based on unstructured random mappings given in the previous section.

Example 2: Suppose (S_1, S_2, S_3) are as in Example 1. The sources are to be transmitted via a MAC with binary inputs $\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3$, binary outputs $\mathcal{Y}_1 \times \mathcal{Y}_2$, and a conditional probability distribution that satisfies

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

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

As explained in Example 1, the univariate common parts are trivial, and the 2-additive common parts are $T_i = S_i, i = 1,2,3$. For such a setup, we use random affine maps explained above. The following lemma provides a necessary and sufficient condition for reliable transmission of (S_1, S_2, S_3). The achievability is obtained using the above approach.

Proposition 2: Consider the source given in Example 1 with $p_1 = p_2 = p$. Such a source is transmissible over the MAC given in Example 2, if and only if $h_b(p) \leq 1 - h_b(\delta), i = 1,2$. Moreover, the source with parameter $p = h_b^{-1}(1 - h_b(\delta))$ does not satisfy the sufficient condition in Proposition 1.

Proof: The proof for the direct part follows using random affine maps. For that, set $X_i^n = S_i^n \mathbf{G} \oplus \mathbf{B}_i, i = 1,2,3$, where $\mathbf{G}, \mathbf{B}_1, \mathbf{B}_2, \mathbf{B}_3$ are selected randomly, and uniformly with elements from $\mathbb{F}_q$ and satisfying $\mathbf{B}_1 \oplus \mathbf{B}_2 \oplus \mathbf{B}_3 = \mathbf{0}$. In this case, $X_3^n = X_1^n \oplus X_2^n$ which implies that $Y_1^n = X_1^n \oplus N_\delta^n$ and $Y_2^n = X_2^n \oplus N_{\delta'}^n$. Hence, from the properties of random linear maps for the point-to-point joint source-channel setting, (S_1, S_2) can be decoded with arbitrary small error probability, if $h_b(p_i) \leq 1 - h_b(\delta), i = 1,2$.

For the converse part, suppose (S_1, S_2, S_3) are transmissible. Therefore, for any $\epsilon > 0$ there exists a coding scheme with error probability at most ϵ. Suppose (e_1, e_2, e_3) are the encoders and d is the decoder of such a scheme. Then, from Fano's inequality,

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

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

Next, we prove the last statement of the proposition by contradiction. Suppose the sources with parameter $p_1 = p_2 = h_b^{-1}(1 - h_b(\delta))$ satisfy the conditions in Proposition 1.

[1] Note that the conferencing common part random variables are unique upto a relabeling.

Then, from the fourth inequality in Proposition 1,

$$2-2h_b(\delta) \le \max_{P_{\underline{U},\underline{X}|\underline{S}} \in \mathscr{P}_{CES}} I(X_1,X_2,X_3;Y|\underline{U}) = \max_{P_{\underline{U}} P_{\underline{X}|\underline{US}}} I(X_1X_2\ X_3;Y|\underline{U}),$$

where $P_{\underline{X}|\underline{US}} = \prod_{i=1}^{3} P_{X_i|S_i,\underline{U}}$. The equality holds as there is no univariate common part, and hence, $\underline{U}$ is independent of the sources. Since, $\underline{U}$ appears in the conditioning in the mutual information term, the above inequality is equivalent to

$$2-2h_b(\delta) \le \max_{P_{X_1|S_1} P_{X_2|S_2} P_{X_3|S_3}} I(X_1X_2X_3;Y). \quad (3)$$

One can verify that $I(X_1,X_2,X_3;Y) \le 2-2h_b(\delta)$, with equality, if and only if, $X_3 = X_1 \oplus X_2$ with probability one, and X_1 and X_2 are uniform over $\{0,1\}$. However, we show that such distribution cannot be generated by taking the marginal of $P_{\underline{S}} P_{X_1|S_1} P_{X_2|S_2} P_{X_3|S_3}$. This is because, to get X_1 and X_2 to be uniform over $\{0,1\}$, we need to set $P_{X_1|S_1}(x|s) = P_{X_2|S_2}(x|s) = \frac{1}{2}$ for all $x,s \in \{0,1\}$. This implies that, X_1 and X_2 are independent of each other and of S_1 and S_2, respectively. Hence, $P_{\underline{S},\underline{X}} = P_{\underline{S}} P_{X_1} P_{X_2} P_{X_3|S_3}$, which means that (X_1,X_2) are independent of X_3. This contradicts with the condition that $X_3 = X_1 \oplus X_2$. ■

C. New Sufficient Condition

We use the intuition behind the argument in Subsection III-B and propose a new coding strategy in which a combination of random linear codes (as in Example 2) and the extension of CES scheme is used. The coding scheme uses both univariate and additive common information among the sources. In the next Theorem, we derive sufficient conditions for transmission of correlated sources over three-user MAC.

Definition 10: Given a source $(\underline{S}, P_{\underline{S}})$ with an additive common part (T_1,T_2,T_3), and a MAC $(\underline{\mathcal{X}}, \mathcal{Y}, P_{Y|X_1X_2X_3})$, let $\mathscr{P}$ be the set of conditional distributions $P_{\underline{U},\underline{V},\underline{X}|\underline{S}}$ defined on $\underline{\mathcal{U}} \times \mathbb{F}_q^3 \times \underline{\mathcal{X}}$ which can be factored as

$$P_{U_{123}} \left[\prod_{b\in\{12,13,23\}} P_{U_b|W_bU_{123}} \right] P_{V_1V_2V_3} \times \left[\prod_{\substack{i,j,k\in\{1,2,3\} \\ j<k, i\neq j, i\neq k}} P_{X_i|S_iU_{123}U_{ij}U_{ik}V_i} \right], \quad (4)$$

where $\mathbb{F}_q$ is the finite field associated with the additive common part, the random variables $(W_{123},W_{12},W_{13},W_{23})$ are the univariate common parts of the sources,

$$P_{V_1V_2V_3} = \frac{1}{q^2} \mathbb{1}\{V_3 \oplus_q V_1 \oplus_q V_2 = 0\},$$

and with slight abuse of notation $\underline{U} \triangleq (U_{123},U_{12},U_{13},U_{23})$. $\underline{\mathcal{U}}$ and $\underline{\mathcal{V}}$ are finite alphabets associated with the auxiliary random variables $\underline{U}$ and $\underline{V}$, respectively.

Theorem 1: A source $(\underline{S}, P_{\underline{S}})$ with an additive common part (T_1,T_2,T_3) is reliably transmissible over a MAC $(\underline{\mathcal{X}}, \mathcal{Y}, P_{Y|X_1X_2X_3})$, if there exists a conditional distribution $P_{\underline{U},\underline{V},\underline{X}|\underline{S}} \in \mathscr{P}$ such that for all $a,b \in \mathbb{F}_q$, any distinct $i,j,k \in \{1,2,3\}$, and for any $\mathcal{B} \subseteq \{12,13,23\}$ the set of inequalities in (5), shown at the bottom of the page, hold.

Remark 1: Via a choice of $X_i, i=1,2,3$, that is independent of V_i, one obtains an extension of the CES scheme for the three user case (Proposition 1), i.e., the set of conditions given in Theorem 1 is weaker than that in Proposition 1.

Outline of the proof: We use a new approach which is based on affine maps to encode additive common parts. Suppose the random variables $(\underline{S},\underline{X},U_{123},U_{12},U_{13},U_{23},\underline{V})$ are distributed according to a joint distribution that factors as in (4).

Codebook Generation: At each transmitter five different codebooks are defined, one codebook for the additive common part T_i, three codebooks for univariate common parts (W_{123},W_{ij},W_{ik}), where i,j,k are distinct elements of $\{1,2,3\}$, and one codebook for generating the total output X_i^n. Fix $\epsilon > 0$.

1) The codebooks for encoding of univariate common parts are as in the proof of Proposition 1.
2) The codebook for encoding of (T_1,T_2,T_3) is defined using affine maps. Generate two vectors $\mathbf{B}_1, \mathbf{B}_2$ of length n, and an $n \times n$ matrix $\mathbf{G}$ with elements selected randomly, uniformly and independently from $\mathbb{F}_q$. Set $\mathbf{B}_3 = -(\mathbf{B}_1 \oplus_q \mathbf{B}_2)$. For each sequence $\mathbf{t}_i \in \mathbb{F}_q^n$, define $\mathbf{V}_i(\mathbf{t}_i) = \mathbf{t}_i\mathbf{G} \oplus \mathbf{B}_i$, where $i=1,2,3$, and all the additions and multiplications are modulo-q.
3) Given distinct $i,j,k \in \{1,2,3\}$, any realization $\mathbf{s}_i$ of the source, the common parts $(\mathbf{w}_{123},\mathbf{w}_{ij},\mathbf{w}_{ik},\mathbf{t}_i)$, and the corresponding sequences

$$\big(\mathbf{U}_{123}(\mathbf{w}_{123}), \mathbf{U}_{ij}(\mathbf{w}_{ij},\mathbf{U}_{123}), \mathbf{U}_{ik}(\mathbf{w}_{ik},\mathbf{U}_{123}), \mathbf{V}_i(\mathbf{t}_i)\big)$$

$$H(S_i|S_j,S_k) \le I(X_i;Y|S_j,S_kU_{123},U_{12},U_{13},U_{23},V_1,V_2,V_3,X_j,X_k) \quad (5a)$$
$$H(S_i,S_j|S_k,W_{\mathcal{B}}) \le I(X_i,X_j;Y|S_k,W_{\mathcal{B}},U_{123},U_{ik},U_{jk}U_{\mathcal{B}},V_k,X_k) \quad (5b)$$
$$H(S_i,S_j|S_k,W_{\mathcal{B}},\underline{T}) \le I(X_i,X_j;Y|S_k,W_{\mathcal{B}},U_{123},U_{ik},U_{jk}U_{\mathcal{B}},\underline{T},\underline{V},X_k) \quad (5c)$$
$$H(S_1,S_2,S_3|W_{123},W_{\mathcal{B}},\underline{T}) \le I(X_1,X_2,X_3;Y|W_{123},W_{\mathcal{B}},U_{123},U_{\mathcal{B}},\underline{T},\underline{V}) \quad (5d)$$
$$H(S_1,S_2,S_3|\underline{T}) \le I(X_1,X_2,X_3;Y|\underline{T},\underline{V}) \quad (5e)$$
$$H(S_1,S_2,S_3|aT_1 \oplus_q bT_2) \le I(X_1,X_2,X_3;Y|aT_1 \oplus_q bT_2, a\ V_1 \oplus_q bV_2) \quad (5f)$$
$$H(S_1,S_2,S_3|W_{123},W_{\mathcal{B}},aT_1 \oplus_q bT_2) \le I(X_1,X_2,X_3;Y|W_{123},W_{\mathcal{B}},U_{123},U_{\mathcal{B}},aT_1 \oplus_q bT_2, aV_1 \oplus_q bV_2) \quad (5g)$$

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

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

$$\big(\mathbf{U}_{123}(\mathbf{w}_{123}), \mathbf{U}_{ij}(\mathbf{w}_{ij}, \mathbf{U}_{123}), \mathbf{U}_{ik}(\mathbf{w}_{ik}, \mathbf{U}_{123}), \mathbf{V}_i(\mathbf{t}_i)\big)$$

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

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

$$(\underline{\tilde{\mathbf{s}}}, \tilde{\mathbf{U}}_{123}, \tilde{\mathbf{U}}_{12}, \tilde{\mathbf{U}}_{13}, \tilde{\mathbf{U}}_{23}, \underline{\tilde{\mathbf{v}}}, \underline{\tilde{\mathbf{X}}}, \mathbf{y}) \in A_\epsilon^{(n)}(\underline{S}, U_{123}, U_{12}, U_{13}, U_{23}, \underline{V}, \underline{X}, Y), \quad (6)$$

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

A decoding error will be occurred, if no unique $(\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2, \tilde{\mathbf{s}}_3)$ is found. It is shown in Appendix A that the probability of error approaches zero as $n \to \infty$, if the inequalities in (5) are satisfied. ■

Remark 2: The coding strategy explained in the proof of Theorem 1 subsumes the extension of CES scheme and identical random linear coding strategy.

D. Example With Structural Mismatch

In Example 2, the structure in the sources matches with that of the channel. In other words, the source correlation is captured via the relation given by $S_3 = S_1 \oplus S_2$, and when $X_3 = X_1 \oplus X_2$, the channel behaved obligingly. In this section, we consider an example where there is a mismatch between the structures of the source and the channel. In other words, the source correlation is still governed by $S_3 = S_1 \oplus S_2$, whereas, the channel fuses X_3 and $X_1 \oplus X_2$ in a nonlinear fashion. In what follows, we provide an application of our coding scheme in scenarios where there is a structural mismatch between the sources and the channel.

Example 3: Consider the sources denoted by (S_1, S_2, S_3), where S_1 and S_3 are independent Bernoulli random variables with parameter $\sigma, \gamma \in [0, \frac{1}{2}]$, respectively. Suppose that the second source satisfies $S_2 = S_1 \oplus_2 S_3$ with probability one. For shorthand we associate such sources with the parameters (σ, γ). The sources are to be transmitted trough a MAC with binary inputs as shown in Figure 4. In this channel the noise random variable N is assumed to be independent of other random variables. The PMF of N is given in Table I, where the parameter $\delta \in (0, \frac{1}{4}]$. As a result, $H(N) = 1 + \frac{1}{2}h_b(2\delta)$.

For this setup, we show that there exist parameters (σ, γ) whose corresponding sources in Example 3 cannot be transmitted reliably using the CES scheme. However, according to Theorem 1, such sources can be reliably transmitted. This emphasizes the fact that efficient encoding of conferencing common information contributes to improvements upon coding schemes solely based on univariate common information. In what follows, we explain the steps to show the existence of such parameters.

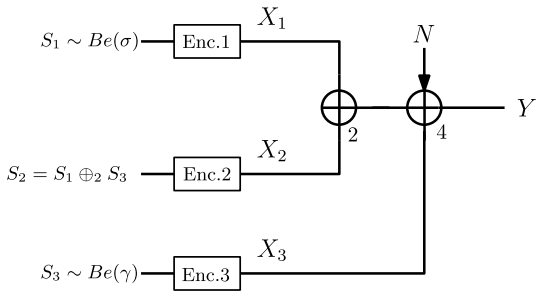

Fig. 4. The diagram the setup introduced in Example 3. Note the input alphabets of this MAC are restricted to $\{0, 1\}$.

TABLE I
DISTRIBUTION OF N

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

Remark 3: For the special case in which $\sigma = 0$, the equalities $S_1 = 0$ and $S_2 = S_3$ hold with probability one. From Proposition 1, such (S_1, S_2, S_3) can be transmitted using CES scheme, if $h_b(\gamma) \leq 2 - H(N)$ holds.

Let $\gamma^* \in [h_b^{-1}(0.5), \frac{1}{2})$ be such that $\gamma^* = h_b^{-1}(2 - H(N))$. Such a γ^* exists as $2 - H(N) = 1 - \frac{1}{2}h_b(2\delta)$ and, thus, is a number between $\frac{1}{2}$ to 1. By Remark 3, the sources (S_1, S_2, S_3) with parameter $(\sigma = 0, \gamma = \gamma^*)$ can be transmitted reliably using CES scheme. However, we argue that for small enough $\epsilon > 0$, the sources with parameter $(\sigma = \epsilon, \gamma = \gamma^* - \epsilon)$ cannot be transmitted using this scheme. Whereas, from Theorem 1, this source can be transmitted reliably. This is formally stated as follows.

Theorem 2: There exist $\sigma \in (0, \frac{1}{2}]$ and $\gamma \in (0, \gamma^*]$ such that the triplet sources (S_1, S_2, S_3) with these parameters satisfies the sufficient condition of Theorem 1, thus, transmissible over the channel in Example 3, but does not satisfy the sufficient condition in Proposition 1.

Proof: The proof is in Appendix B. ■

IV. COMMUNICATIONS OVER MAC WITH FEEDBACK: PRELIMINARIES

The problem of three user MAC with noiseless feedback is depicted in Figure 5. This communication channel consists of one receiver and multiple transmitters. After each channel use, the output of the channel is received at each transmitter noiselessly. Gaarder and Wolf [15] showed that the capacity region of the MAC can be expanded through the use of the feedback. This was shown in a binary erasure MAC. Cover and Leung [16] studied the two-user MAC with feedback, and developed a coding strategy using unstructured random codes.

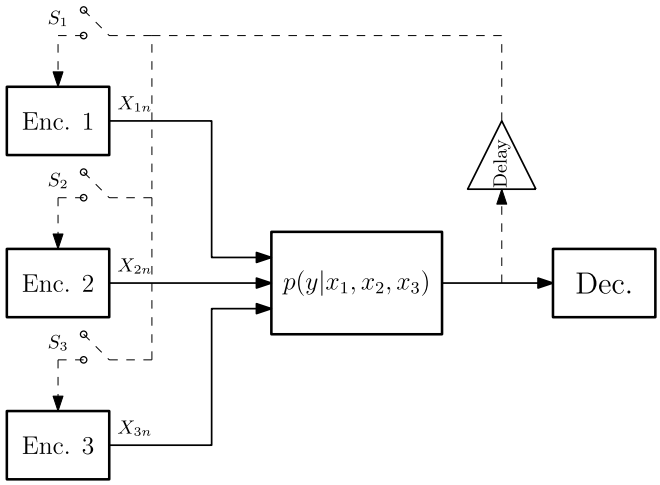

Fig. 5. The three-user MAC with noiseless feedback. If the switch S_i is closed, the feedback is available at the ith encoder, where $i = 1, 2, 3$.

A. Model and Problem Formulation

In what follows, we formulate the problem of communications over MAC-FB. We restrict ourselves to three-user MAC with noiseless feedback in which all or a subset of the transmitters have access to the feedback perfectly. Consider a three-user MAC identified by a transition probability matrix $P_{Y|X_1,X_2,X_3}$ as in Definition 1. Let $\mathbf{y}^n$ be a realization of the output of the channel after n uses, where $\mathbf{x}_i^n$ is the ith input sequence of the channel, $i \in [1, 3]$. Then, the conditional probability distribution of the channel output y_n given the current and past input and output vectors is given by

$$P_{Y_n|\mathbf{Y}^{n-1},\mathbf{X}_1^n,\mathbf{X}_2^n,\mathbf{X}_3^n}(y_n|\mathbf{y}^{n-1},\mathbf{x}_1^n,\mathbf{x}_2^n,\mathbf{x}_3^n) = P_{Y|X_1,X_2,X_3}(y_n|x_{1n},x_{2n},x_{3n}). \quad (7)$$

It is assumed that noiseless feedback is made available, with one unit of delay, to a subset $\mathcal{T} \subseteq [1, 3]$ of the transmitters. In Figure 5, the switches $S_i, i = 1, 2, 3$ determine which transmitter receives the feedback. A formal definition of a MAC-FB setup is given in the following.

Definition 11: A 3-user MAC-FB setup is characterized by a 3-user MAC $(\underline{\mathcal{X}}, \mathcal{Y}, P_{Y|X_1X_2X_3})$ and a subset $\mathcal{T} \subseteq [1, 3]$ determining the transmitters which have access to the feedback. It is assumed that at least one transmitter has access to the feedback, i.e., $|\mathcal{T}| \geq 1$. Such a MAC-FB is denoted by $(\underline{\mathcal{X}}, \mathcal{Y}, P_{Y|\underline{X}}, \mathcal{T})$.

Definition 12: For a 3-user MAC-FB $(\underline{\mathcal{X}}, \mathcal{Y}, P_{Y|\underline{X}}, \mathcal{T})$, an $(N, \Theta_1, \Theta_2, \Theta_3)$ coding scheme consists of 3 sequences of encoding functions defined as,

$$e_{i,n} : [1, \Theta_i] \times \mathcal{Y}^{n-1} \to \mathcal{X}_i,$$

for $i \in \mathcal{T}$, and

$$e_{j,n} : [1, \Theta_j] \to \mathcal{X}_j,$$

for $j \in \mathcal{T}^c$, with $n \in [1, N]$ and a decoding function denoted by

$$d : \mathcal{Y}^N \to [1, \Theta_1] \times [1, \Theta_2] \times [1, \Theta_3].$$

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

It is assumed that, transmitter i receives a message index M_i which is drawn randomly and uniformly from $[1, \Theta_i]$, where $i \in [1, 3]$. Furthermore, the message indexes (M_1, M_2, M_3) are assumed to be mutually independent. Moreover, the timeline of the random variables are in the following order

$$(M_1, M_2, M_3), (X_{11}, X_{21}, X_{31}), Y_1, (X_{12}, X_{22}, X_{32}), Y_2, \cdots$$

We assume that the channel does not have access to the messages, i.e., the Markov chain[2],

$$(M_1, M_2, M_3) - (\mathbf{Y}^{n-1}, \mathbf{X}_1^n, \mathbf{X}_2^n, \mathbf{X}_3^n) - Y_n,$$

holds for all $n \in [1, N]$. For this setup, the average probability of error is defined as

$$P_{err}(\underline{e}) \triangleq \mathbb{P}\{d(Y^N) \neq (M_1, M_2, M_3)\}, \quad (8)$$

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

Definition 13: For a 3-user MAC-FB, a rate-tuple (R_1, R_2, R_3) is said to be achievable, if for any $\epsilon > 0$ there exists, for all sufficiently large N, an $(N, \Theta_1, \Theta_2, \Theta_3)$ coding scheme such that

$$P_{err}(\underline{e}) < \epsilon, \quad \frac{1}{N} \log_2 \Theta_i \geq R_i - \epsilon, \quad i = 1, 2, 3.$$

B. CL Achievable Region: Unstructured Coding Approach

The main idea behind the CL scheme is to use superposition block-Markov encoding. The scheme operates in two stages. In stage one, the transmitters send the messages with rates outside the no-feedback capacity region, but small enough that each user can decode the other user's message using feedback. In the second stage, the encoders fully cooperate to send the messages to disambiguate the information at the receiver. Using this approach, the following rate-region is achievable for communications over a MAC with noiseless feedback available at at least on of the transmitters [16].

Fact 2: Given a two-user MAC-FB $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}, P_{Y|X_1,X_2}, \mathcal{T} \subseteq \{1, 2\})$, a rate pair (R_1, R_2) is achievable, if there exist distributions $P_U, P_{X_1|U}$, and $P_{X_2|U}$ such that

$$\begin{aligned} R_1 &\leq I(X_1; Y|X_2, U), \\ R_2 &\leq I(X_2; Y|X_1, U), \\ R_1 + R_2 &\leq I(X_1, X_2; Y), \end{aligned}$$

where U takes values from a finite set $\mathcal{U}$, and the joint distribution of all the random variables factors as

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

It was shown in [20] that, in a two-user MAC-FB, the CL rate region is achievable even if only one of the transmitters has access to the feedback ($|\mathcal{T}| = 1$).

[2]This is the standard formulation of channel coding problem with feedback following [45].

As explained in CL scheme, the decoded sub-messages $(M_{1,b}, M_{2,b})$ are used as a common information for the next block of transmission. One can extend this scheme for a multi-user MAC-FB setup (say a three-user MAC-FB) using unstructured codes. In this setup, the transmitters send the messages with rates outside the no-feedback capacity region. Hence, the receiver is not able to decode the messages. However, the transmission rates are taken to be sufficiently low so that each user can decode the sub-messages of the other users. The decoded sub-messages at the end of each block b are used as uni-variate common parts for the next block of transmission. One can derive a single-letter characterization of an achievable rate region based on such a scheme in a straightforward fashion. For conciseness we do not state this rate region in this paper.

V. Three-User MAC-FB: Structured Codes

In this section, we propose a new coding scheme for three-user MAC-FB, and derive a computable single-letter achievable rate region (an inner bound to the capacity region) using structured codes – in particular, *quasi-linear* codes that were introduced in [46]. Note that prior to the start of the communication, the messages are mutually independent; whereas after multiple uses of the channel, they become statistically correlated conditioned on the feedback. Based on this observation, we make a connection to the problem of MAC with correlated sources to design coding strategies that exploit the statistical correlation among the messages. We use the notion of conferencing common information to propose a new coding strategy for 3-user MAC-FB. The main results of this section are given in Theorem 3 and 4.

A. New Achievable Rate Region

In what follows, we give the intuition behind the use of conferencing common information in MAC-FB. Consider a three-user MAC-FB setup as depicted in Figure 6. Similar to the two-user version of the problem, the communications take place in B blocks each of length n. Moreover, the message at Transmitter i is divided into B sub-messages denoted by $(M_{i,1}, M_{i,2}, ..., M_{i,B})$, where $i = 1, 2, 3$. Suppose, the transmission rates are such that neither the decoder nor the transmitters can decode the messages. However, at each block b, the rates are sufficiently low so that each transmitter is able to decode the modulo-q sum of the other two sub-messages[3]. For instance, Transmitter 1 can decode $M_{2,b} \oplus_q M_{3,b}$ with high probability. Let $T_{i,b}$ denote the decoded sum at Transmitter i, where $i = 1, 2, 3$. Then, for binary messages,

$$T_{1,b} \oplus T_{2,b} \oplus T_{3,b} = 0$$

with high probability. As a result, (T_1, T_2, T_3) can be interpreted as additive conferencing common parts (see Definition 9). Building upon this intuition, in what follows, we propose a coding strategy for communications over 3-user MAC-FB, and we derive a new computable achievable rate region.

[3] It is understood that the messages belong to a finite field $\mathbb{F}_q$.

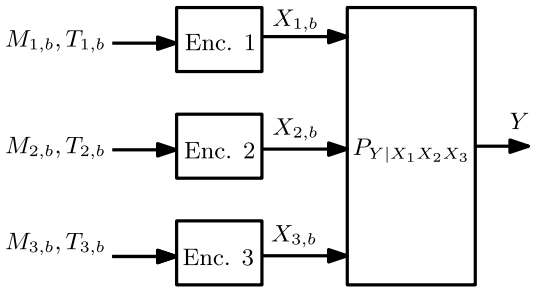

Fig. 6. Applications of conferencing common information for communications over MAC-FB. The new sub-messages at block b are denoted by $M_{i,b}$. At the end of block $b-1$, each transmitter decodes the modulo-two sum of the other two transmitters. The decoded sums are denoted by $T_{i,b}, i = 1, 2, 3$. Note that $T_{1,b} \oplus T_{2,b} \oplus T_{3,b} = 0$ with probability close to one.

We start by the following definition to characterize an achievable rate region.

Definition 14: For a prime q and a given set $\mathcal{U}$ and a three-user MAC-FB $(\underline{\mathcal{X}}, \mathcal{Y}, P_{Y|\underline{X}}, \mathcal{T})$, define $\mathscr{P}$ as the collection of all distributions on $\mathcal{U} \times \mathbb{F}_q^6 \times \underline{\mathcal{X}}, \mathcal{Y}$ factoring as

$$P_U P_{V_1V_2V_3} \prod_{i=1}^{3} P_{T_i} P_{X_i|UT_iV_i} P_{Y|X_1X_2X_3}, \tag{9}$$

where (T_1, T_2, T_3) are mutually independent with uniform distribution over a finite field $\mathbb{F}_q$, (V_1, V_2, V_3) are pairwise independent each with uniform distribution over $\mathbb{F}_q$, and

$$P_{V_1V_2V_3}(v_1, v_2, v_3) = \frac{1}{q^2} \mathbb{1}\{v_1 \oplus v_2 \oplus v_3 = 0\},$$

and for any $i \in \mathcal{T}^c$, we have $P_{X_i|UT_iV_i} = P_{X_i}$ for some distribution on $\mathcal{X}_i$.

We follow a block Markov coding approach with one step memory. To ensure stationarity we impose the following conditions. Fix a distribution $\mathsf{P} \in \mathscr{P}$ that factors as in (9). Denote $S_i = (X_i, T_i, V_i)$ for $i = 1, 2, 3$. Consider two sets of random variables (U, S_1, S_2, S_3, Y) and $(\tilde{U}, \tilde{S}_1, \tilde{S}_2, \tilde{S}_3, \tilde{Y})$. Here the first set corresponds to the current block and the second set corresponds to the previous block in the block Markov coding strategy. The distribution of each set of the random variables is P, i.e.,

$$P_{US_1\ S_2\ S_3Y} = P_{\tilde{U}\tilde{S}_1\tilde{S}_2\tilde{S}_3\tilde{Y}} = \mathsf{P}.$$

In addition, conditioned on $(\tilde{U}, \tilde{S}_1, \tilde{S}_2, \tilde{S}_3, \tilde{Y})$, we have that

$$P_{US_1S_2S_3Y|\tilde{U}\tilde{S}_1\tilde{S}_2\tilde{S}_3\tilde{Y}} = P_U P_{V_1V_2V_3|\tilde{T}_1\tilde{T}_2\tilde{T}_3} \times \prod_{i=1}^{3} P_{T_i} P_{X_i|UT_iV_i} P_{Y|X_1X_2X_3}, \tag{10}$$

and $\underline{V} = \underline{\tilde{T}}\mathbf{A}$ with probability one, where $\mathbf{A}$ is a 3×3 matrix with elements in $\mathbb{F}_q$ and the multiplications are modulo q. Further, $\mathbf{A}$ is chosen such that $P_{V_1V_2V_3} = P_{\tilde{V}_1,\tilde{V}_2,\tilde{V}_3}$. These random variables are described in Fig. 7.

Definition 15: Given a MAC-FB $(\underline{\mathcal{X}}, \mathcal{Y}, P_{Y|\underline{X}}, \mathcal{T})$, let $\mathcal{R}_{\text{MAC-FB}}$ be the set of triplets (R_1, R_2, R_3) for which there exist $\alpha \in (0, 1)$, random variables (U, S_1, S_2, S_3, Y) and $(\tilde{U}, \tilde{S}_1, \tilde{S}_2, \tilde{S}_3, \tilde{Y})$ distributed according to (10) for some

	CI	Conf. CI	New Message		Output
Enc. 1	$\tilde{U}$	$\tilde{V}_1$	$\tilde{W}_1$	$\tilde{T}_1$	$\tilde{X}_1$
Enc. 2	$\tilde{U}$	$\tilde{V}_2$	$\tilde{W}_2$	$\tilde{T}_2$	$\tilde{X}_2$
Enc. 3	$\tilde{U}$	$\tilde{V}_3$	$\tilde{W}_3$	$\tilde{T}_3$	$\tilde{X}_3$

Previous Block

	CI	Conf. CI	New Message		Output
Enc. 1	U	V_1	W_1	T_1	X_1
Enc. 2	U	V_2	W_2	T_2	X_2
Enc. 3	U	V_3	W_3	T_3	X_3

Current Block

Fig. 7. The random variables involved in the block Markov coding strategy. Here, the new message at encoder i is represented by a pair (W_i, T_i), where W_i's are to be decoded at the receiver after two blocks and T_i's are used to construct the next block's conferencing common information. Conferencing common information random variables (V_1, V_2, V_3) satisfy the relation $\underline{V} = \tilde{\underline{T}}\mathbf{A}$, where $\mathbf{A}$ is a 3×3 matrix and $\tilde{T}_i$'s are from the previous block.

$P \in \mathscr{P}$ and matrix $\mathbf{A} \in \mathbb{F}_q^{3\times 3}$ and mutually independent random variables (W_1, W_2, W_3) which are also independent of other random variables such that the following inequalities hold for any subset $\mathcal{B} \subseteq \{1, 2, 3\}$ and any distinct elements $i, j, k \in \{1, 2, 3\}$:

$$\begin{aligned}
\alpha H(W_i) &= R_i, \\
\alpha H(W_{\mathbf{A}_i}|W_i) &\le I(T_{\mathbf{A}_i}; Y|UT_iV_iX_i), \\
\alpha H(W_j, W_k|W_{\mathbf{A}_i}, W_i) &\le I(\tilde{T}_j\tilde{X}_j\tilde{T}_k\tilde{X}_k; Y\tilde{Y}|\tilde{U}\tilde{S}_iUS_i\tilde{V}_j\tilde{V}_k), \\
\alpha H(W_{\mathcal{B}}) &\le I(X_{\mathcal{B}}; Y|US_{\mathcal{B}^c}\tilde{V}_1, \tilde{V}_2, \tilde{V}_3) \\
&\quad + I(U; Y),
\end{aligned}$$

where $W_{\mathbf{A}_i}$ and $T_{\mathbf{A}_i}, i = 1, 2, 3$, are the ith element of the vector $\underline{W}\mathbf{A}$ and $\underline{T}\mathbf{A}$, respectively.

Theorem 3: For a MAC-FB $(\underline{\mathcal{X}}, \mathcal{Y}, P_{Y|\underline{X}}, \mathcal{T})$, the rate-region $\mathcal{R}_{\text{MAC-FB}}$ is achievable.

Proof: The proof is given in Appendix C. ∎

B. Necessity of Structured Codes for MAC-FB

In this section, we show that coding strategies based on structured codes are necessary for certain instances of MAC with feedback. We first provide an example of a MAC with feedback. Then, we apply Theorem 3 and show that the inner bound achieves optimality.

Example 4: Consider the three-user MAC-FB problem depicted in Figure 8. In this setup, there is a MAC with three pairs of binary inputs, where the ith input is denoted by the pair (X_{i1}, X_{i2}) for $i = 1, 2, 3$. The output of the channel is denoted by a binary vector (Y_1, Y_{21}, Y_{22}). Assume that noiseless feedback is available only at the third transmitter.

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

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

where $\tilde{N}_\delta$ is a Bernoulli random variable with bias δ, and is independent of the inputs. The second channel is a MAC with (X_{12}, X_{22}, X_{32}) as the inputs, and (Y_{21}, Y_{22}) as the output. The conditional probability distribution of this channel satisfies

$$\begin{aligned}
&(Y_{21}, Y_{22}) \\
&= \begin{cases} (X_{12} \oplus N_\delta, X_{22} \oplus N'_\delta), & \text{if } X_{32} = X_{12} \oplus X_{22}, \\ (X_{12} \oplus N_{1/2}, X_{22} \oplus N'_{1/2}), & \text{if } X_{32} \neq X_{12} \oplus X_{22}, \end{cases}
\end{aligned} \tag{11}$$

where $N_\delta, N'_\delta, N_{1/2}$ and $N'_{1/2}$ are independent Bernoulli random variables with parameter $\delta, \delta, \frac{1}{2}$, and $\frac{1}{2}$, respectively. The relation between the output and the input of the channel is depicted in Figure 9. The channel operates in two states. If the condition $X_{32} = X_{12} \oplus X_{22}$ holds, the channel would be in the first state (the left channel in Figure 9); otherwise it would be in the second state (the right channel in Figure 9). In this channel, N_δ and N'_δ are Bernoulli random variables with identical bias δ. Whereas, $N_{1/2}$ and $N'_{1/2}$ are Bernoulli random variables with bias $\frac{1}{2}$. We assume that $\tilde{N}_\delta, N_\delta, N'_\delta, N_{1/2}$, and $N'_{1/2}$ are mutually independent, and are independent of all the inputs.

We use linear codes to propose a new coding strategy for the setup given in Example 4. The scheme uses a large number L of blocks, each of length n. Each encoder has two outputs,

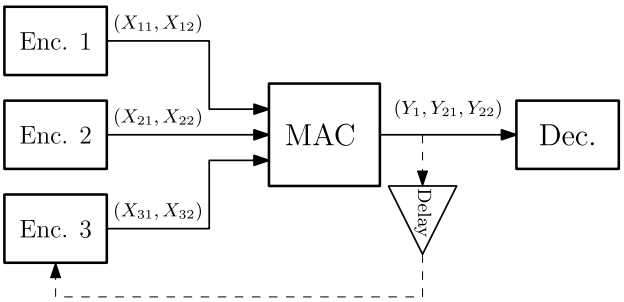

Fig. 8. The MAC with feedback setup for Example 4.

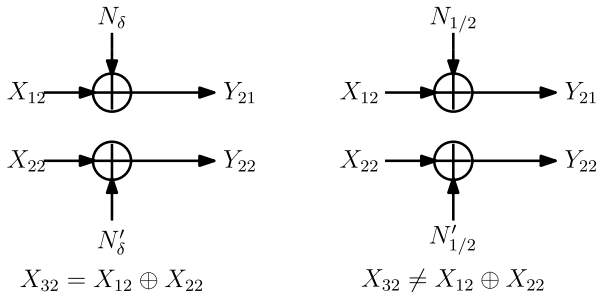

Fig. 9. The second channel for Example 4. If the condition $X_{32} = X_{12} \oplus X_{22}$ holds, the channel would be the one on the left; otherwise it would be the right channel.

one for each channel. We use identical linear codes with length n and rate $\frac{k}{n}$ for each transmitter. The coding scheme at each block is performed in two stages. In the first stage, each transmitter encodes the fresh message at the beginning of the block l, where $1 \leq l \leq L$. The encoding process is performed using identical linear codes. At the end of block l, feedback is received by the third user. In stage 2, the third user uses the feedback from the first channel (that is Y_1) to decode the binary sum of the messages of the other encoders. Then, it encodes the summation, and sends it through its second output. If the decoding process is successful at the third user, then the relation $X_{32} = X_{12} \oplus X_{22}$ holds with probability one. This is because identical linear codes are used to encode the messages. As a result of this equality, the channel in Figure 9 is in the first state with probability one. In the following theorem, we show that the rate

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

is achievable using this strategy. Further, we prove in the followng theorem that any coding scheme achieving these rates must have codebooks that are almost closed under the binary addition. Since unstructured random codes do not have this property, any coding scheme solely based in them is suboptimal.

Theorem 4: For the channel given in Example 4, the rate triple $(1-h(\delta), 1-h(\delta), 1-h(\delta))$ is achievable if and only if 1) user 3 decodes $X_1 \oplus X_2$ with average probability of error approaching zero, and 2) the codebooks in user 1 and 2 must satisfy

$$\lim_{N\to\infty} \frac{1}{N} \Big| \ \log ||\mathcal{C}_{12} \oplus \mathcal{C}_{22}|| - \log ||\mathcal{C}_{12}|| \ \Big| = 0, \quad \text{for} \ \ i = 1, 2.$$

Proof: The proof is given in Appendix D. ∎

VI. Conclusion

A new form of common information, called "conferencing common information", is defined among triplets of random variables. Based on this notion, two coding strategies are proposed for three-user version of two problems: transmission of correlated sources over MAC, and MAC with feedback. Further, achievable rate regions of such strategies are characterized in terms of single-letter information quantities. It is shown analytically that the proposed strategies outperform conventional unstructured random coding approaches in terms of achievable rates.

Appendix A
Proof of Theorem 1

Proof: There are two error events, E_0 and E_1. E_0 occurs if no triple $\underline{\tilde{\mathbf{s}}}$ was found. E_1 occurs if there exists $\underline{\tilde{\mathbf{s}}} \neq \underline{\mathbf{s}}$ such that equation (6) is satisfied. We consider a special case in which all the uni-variate common parts are trivial and that $T_i = S_i, i = 1, 2, 3$. This implies that $S_1 \oplus_q S_2 \oplus_q S_3 = 0$ with probability one. The proof for the general case follows by adopting this proof and the standard arguments as in [7].

Suppose $\mathbf{v_i}(\cdot)$ and $\mathbf{x_i}(\cdot)$ are the realizations of random functions generated as in the outline of the proof of Theorem 1. Using standard arguments one can show that $E_0 \to 0$ as $n \to \infty$. We find the condition under which $P(E_1 \cap E_0^c) \to 0$. For a given $\underline{\mathbf{s}} \in A_{\epsilon_1}(\underline{S})$, using the definition of E_1 and the union bound we obtain,

$$P(E_1 \bigcap E_0^c|\underline{\mathbf{s}}) \leq \sum_{\substack{(\underline{\mathbf{v}},\underline{\mathbf{x}},\underline{\mathbf{y}})\in \\ A_{\epsilon_2}(\underline{V},\underline{X},Y|\underline{\mathbf{s}})}} \mathbb{1}\Big\{\mathbf{v}_i = \mathbf{v}_i(\mathbf{s}_i), \mathbf{x}_i = \mathbf{x}_i(\mathbf{s}_i, \mathbf{v}_i), \ i = 1, 2, 3\Big\} P^n_{Y|\underline{X}}(\mathbf{y}|\underline{\mathbf{x}})$$
$$\sum_{\substack{(\underline{\tilde{\mathbf{s}}},\underline{\tilde{\mathbf{v}}},\underline{\tilde{\mathbf{x}}})\in A_{\epsilon_3}(\underline{S},\underline{V},\underline{X}|\mathbf{y}) \\ \underline{\tilde{\mathbf{s}}}\neq\underline{\mathbf{s}}}} \mathbb{1}\Big\{\tilde{\mathbf{v}}_j = \mathbf{v}_j(\tilde{\mathbf{s}}_j), \tilde{\mathbf{x}}_j = \mathbf{x}_j(\tilde{\mathbf{s}}_j, \tilde{\mathbf{v}}_j), j = 1, 2, 3\Big\}$$

Taking expectation over random vector functions $\mathbf{X}_i(,)$ and $\mathbf{V}_i()$ gives,

$$p_e(\underline{\mathbf{s}}) = \mathbb{E}\{P(E_1|\underline{\mathbf{s}})\}$$
$$\leq \sum_{\substack{(\underline{\mathbf{v}},\underline{\mathbf{x}},\underline{\mathbf{y}})\in \\ A_{\epsilon_2}(\underline{V},\underline{X},Y|\underline{\mathbf{s}})}} P^n_{Y|\underline{X}}(\mathbf{y}|\underline{\mathbf{x}}) \sum_{\substack{(\underline{\tilde{\mathbf{s}}},\underline{\tilde{\mathbf{v}}},\underline{\tilde{\mathbf{x}}})\in A_{\epsilon_3}(\underline{S},\underline{V},\underline{X}|\mathbf{y}) \\ \underline{\tilde{\mathbf{s}}}\neq\underline{\mathbf{s}}}} P\Big\{\mathbf{v}_l = \mathbf{V}_l(\mathbf{s}_l), \mathbf{x}_l = \mathbf{X}_l(\mathbf{s}_l, \mathbf{v}_l),$$
$$\tilde{\mathbf{v}}_l = \mathbf{V}_l(\tilde{\mathbf{s}}_l), \tilde{\mathbf{x}}_l = \mathbf{X}_l(\tilde{\mathbf{s}}_l, \tilde{\mathbf{v}}_l) \text{ for } l = 1, 2, 3\Big\}. \tag{12}$$

Let

$$\epsilon = \max_{i\in[1,3]} \epsilon_i, \tag{13}$$

where ϵ_i is as in the above summations. Note that $V_i(\cdot)$ and $\mathbf{X}_i(\cdot,\cdot)$ are generated independently. So the most inner term in (12) is simplified to

$$P\Big\{\mathbf{v}_j = \mathbf{V}_j(\mathbf{s}_j), \tilde{\mathbf{v}}_j = \mathbf{V}_j(\tilde{\mathbf{s}}_j) \ j = 1, 2\Big\} \times$$
$$P\Big\{\mathbf{x}_l = \mathbf{X}_l(\mathbf{s}_l, \mathbf{v}_l), \tilde{\mathbf{x}}_l = \mathbf{X}_l(\tilde{\mathbf{s}}_l, \tilde{\mathbf{v}}_l) \ l = 1, 2, 3\Big\}. \tag{14}$$

Note that $j = 3$ is redundant because, $\mathbf{v}_3 \oplus_q \mathbf{v}_1 \oplus_q \mathbf{v}_2 = \mathbf{0}$ and $\tilde{\mathbf{v}}_3 \oplus_q \tilde{\mathbf{v}}_1 \oplus_q \tilde{\mathbf{v}}_2 = \mathbf{0}$. By definition, $\mathbf{V}_j(\mathbf{s}_j) = \mathbf{s}_j\mathbf{G} + \mathbf{B}_j, j = 1, 2$, where $\mathbf{B}_1, \mathbf{B}_2$ are uniform and independent of $\mathbf{G}$. Then

$$P\Big\{\mathbf{v}_j = \mathbf{V}_j(\mathbf{s}_j), \tilde{\mathbf{v}}_j = \mathbf{V}_j(\tilde{\mathbf{s}}_j), \ j = 1, 2\Big\} =$$
$$\frac{1}{q^{2n}} P\Big\{(\tilde{\mathbf{s}}_j - \mathbf{s}_j)\mathbf{G} = \tilde{\mathbf{v}}_j - \mathbf{v}_j, \ j = 1, 2\Big\}. \tag{15}$$

The following lemma determines the above term.

Lemma 1: Suppose $\mathbf{G}$ is a $n \times m$ matrix with elements generated randomly and uniformly from $\mathbb{F}_q$. If $\mathbf{s}_1$ or $\mathbf{s}_2$ is nonzero, the following holds:

$$P\{\mathbf{s}_j\mathbf{G} = \mathbf{v}_j, \ j = 1, 2\} = \begin{cases} \mathbb{1}\{\mathbf{v}_j = \mathbf{0}, \ l = 1, 2\}, & \text{if } \mathbf{s}_1 = \mathbf{0}, \mathbf{s}_2 = \mathbf{0}. \\ q^{-n}\mathbb{1}\{\mathbf{v}_j = \mathbf{0}\}, & \text{if } \mathbf{s}_j = \mathbf{0}, \mathbf{s}_{j^c} \neq \mathbf{0}. \\ q^{-n}\mathbb{1}\{\mathbf{v}_1 = a\mathbf{v}_2\}, & \text{if } \mathbf{s}_1 \neq \mathbf{0}, \mathbf{s}_2 \neq \mathbf{0}, \\ & \quad \mathbf{s}_1 = a\mathbf{s}_2, \ a \in \mathbb{F}_q. \\ q^{-2n}, & \text{if otherwise.} \end{cases}$$

Proof: We can write

$$\mathbf{s}_j\mathbf{G} = \sum_{i=1}^{n} \mathbf{s}_{ji}\mathbf{G}_i, \quad j = 1, 2,$$

where $\mathbf{s}_{ji}$ is the ith component of $\mathbf{s}_j$ and $\mathbf{G}_i$ is the ith row of $\mathbf{G}$. Not that $\mathbf{G}_i$ are independent random variables with

uniform distribution over $\mathbb{F}_q^n$. Hence, if $\mathbf{s}_j \neq \mathbf{0}$, then $\mathbf{s}_j\mathbf{G}$ is uniform over $\mathbb{F}_q^n$. Then, given the condition $\mathbf{s}_j = \mathbf{0}, \mathbf{s}_{j^c} \neq \mathbf{0}$, we obtain that

$$P\{\mathbf{s}_j\mathbf{G} = \mathbf{v}_j,\ j = 1, 2\} = q^{-n}\mathbb{1}\{\mathbf{v}_j = \mathbf{0}\}.$$

If $\mathbf{s}_1 = a\mathbf{s}_2$ with $a \in \mathbb{F}_q$, then $\mathbf{s}_1\mathbf{G} = a\mathbf{s}_2\mathbf{G}$, with probability one and, thus,

$$P\{\mathbf{s}_j\mathbf{G} = \mathbf{v}_j,\ j = 1, 2\} = q^{-n}\mathbb{1}\{\mathbf{v}_1 = a\mathbf{v}_2\}.$$

If $\mathbf{s}_1 \neq a\mathbf{s}_2$ for any $a \in \mathbb{F}_q$, then $(\mathbf{s}_1, \mathbf{s}_2)$ are linearly independent. This implies that there exist indices (l, k) such that the 2×2 matrix $\mathbf{A}$ with elements $a_{11} = s_{1l}, a_{12} = s_{1,k}, a_{21} = s_{2l}$ and $a_{22} = s_{2k}$ is full rank. As a result, $s_{1l}\mathbf{G}_l \oplus s_{1k}\mathbf{G}_k$ and $s_{2l}\mathbf{G}_l \oplus s_{2k}\mathbf{G}_k$ are independent random vectors with uniform distribution over $\mathbb{F}_q^k$. In this case, one can show that $\mathbf{s}_1\mathbf{G}$ is independent of $\mathbf{s}_2\mathbf{G}$. The proof follows by arguing that if a random variables X is independent of Y and is uniform over $\mathbb{F}_q$, then $X \oplus_q Y$ is also uniform over $\mathbb{F}_q$ and is independent of Y. ∎

Finally, we are ready to characterize the conditions under which $p_e \to 0$. Let $\mathcal{L}(\underline{\mathbf{s}})$ denote the set of all the variables $(\underline{\mathbf{v}}, \underline{\mathbf{x}}, \mathbf{y}, \underline{\tilde{\mathbf{s}}}, \underline{\tilde{\mathbf{v}}}, \underline{\tilde{\mathbf{x}}})$ included in the summations in (12); more precisely,

$$\mathcal{L}(\underline{\mathbf{s}}) \triangleq \Big\{(\underline{\mathbf{v}}, \underline{\mathbf{x}}, \mathbf{y}, \underline{\tilde{\mathbf{s}}}, \underline{\tilde{\mathbf{v}}}, \underline{\tilde{\mathbf{x}}}) : (\underline{\mathbf{v}}, \underline{\mathbf{x}}, \mathbf{y}) \in A_{\epsilon_2}(\underline{V}, \underline{X}, Y|\underline{\mathbf{s}}), (\underline{\tilde{\mathbf{s}}}, \underline{\tilde{\mathbf{v}}}, \underline{\tilde{\mathbf{x}}}) \in A_{\epsilon_3}(\underline{S}, \underline{V}, \underline{X}|\mathbf{y}),\ \tilde{\mathbf{s}} \neq \underline{\mathbf{s}}\Big\}. \quad (16)$$

Based on the conditions in Lemma 1, we partition this set into five subsets $\mathcal{L}_i(\underline{\mathbf{s}}), i = 1, 2, ..., 5$. Hence, if $p_{e_i}(\underline{\mathbf{s}}), i \in [1, 5]$ represents the contribution of each subset, then

$$p_e(\underline{\mathbf{s}}) = \sum_{i=1}^{5} p_{e_i}(\underline{\mathbf{s}}).$$

In what follows, we characterize these subsets and provide an upper bound to each term $p_{e_i}(\underline{\mathbf{s}}), i \in [1, 5]$.

Case 1, $\tilde{\mathbf{s}}_1 \neq \mathbf{s}_1, \tilde{\mathbf{s}}_2 = \mathbf{s}_2$:

In this case, using Lemma 1, the right-hand side of (15) equals to $q^{-3n}\mathbb{1}\{\tilde{\mathbf{v}}_2 = \mathbf{v}_2\}$. As $\mathbf{s}_2 = \tilde{\mathbf{s}}_2$ and $\mathbf{v}_2 = \tilde{\mathbf{v}}_2$, then $X_2(\tilde{\mathbf{s}}_2, \tilde{\mathbf{v}}_2) = X_2(\mathbf{s}_2, \mathbf{v}_2)$. Therefore, we define

$$\mathcal{L}_1(\underline{\mathbf{s}}) \triangleq \Big\{(\underline{\mathbf{v}}, \underline{\mathbf{x}}, \mathbf{y}, \underline{\tilde{\mathbf{s}}}, \underline{\tilde{\mathbf{v}}}, \underline{\tilde{\mathbf{x}}}) \in \mathcal{L}(\underline{\mathbf{s}}) :\ \tilde{\mathbf{s}}_1 \neq \mathbf{s}_1, \tilde{\mathbf{s}}_2 = \mathbf{s}_2, \tilde{\mathbf{v}}_2 = \mathbf{v}_2, \mathbf{x}_2 = \tilde{\mathbf{x}}_2\Big\},$$

where $\mathcal{L}(\underline{\mathbf{s}})$ is defined as in (16). Thus, the contribution of this case equals to

$$p_{e_1}(\underline{\mathbf{s}}) \triangleq \sum_{\mathcal{L}_1(\underline{\mathbf{s}})} P^n_{Y|\underline{X}}(\mathbf{y}|\underline{\mathbf{x}})\ q^{-3n} \times P\Big\{\mathbf{x}_l = \mathbf{X}_l(\mathbf{s}_l, \mathbf{v}_l), \tilde{\mathbf{x}}_l = \mathbf{X}_l(\tilde{\mathbf{s}}_l, \tilde{\mathbf{v}}_l),\ l = 1, 2, 3\Big\}.$$

Note that $\mathbf{X}_l(\mathbf{s}_l, \mathbf{v}_l)$ is independent of $\mathbf{X}_k(\tilde{\mathbf{s}}_k, \tilde{\mathbf{v}}_k)$, if $l \neq k$ or $\mathbf{s}_l \neq \tilde{\mathbf{s}}_l$ or $\mathbf{v}_l \neq \tilde{\mathbf{v}}_l$. Moreover, since $\mathbf{X}_l(\mathbf{s}_l, \mathbf{v}_l)$ is generated IID according to $P_{X_l|S_l,V_l}$, then for jointly typical sequences $(\mathbf{x}_l, \mathbf{s}_l, \mathbf{v}_l)$,

$$\frac{-1}{n}\log_2\ P\{\mathbf{x}_l = \mathbf{X}_l(\mathbf{s}_l, \mathbf{v}_l)\} \geq H(X_l|S_lV_l)) - \delta_1(\epsilon),$$

where ϵ is defined as in (13) and $\delta_1(\epsilon) \geq 0$ is a continuous function satisfying $\lim_{\epsilon\to 0}\delta_1(\epsilon) = 0$. Therefore,

$$P\Big\{\mathbf{x}_l = X_l(\mathbf{s}_l, \mathbf{v}_l), \tilde{\mathbf{x}}_l = X_l(\tilde{\mathbf{s}}_l, \tilde{\mathbf{v}}_l)\ l = 1, 2, 3\Big\} \leq 2^{-n[2H(X_1|S_1V_1)+H(X_2|S_2V_2)+2H(X_3|S_3V_3)-\delta_2(\epsilon)]}\mathbb{1}\{\tilde{\mathbf{x}}_2 = \mathbf{x}_2\},$$

where δ_2 is a non-negative and continuous function with $\lim_{\epsilon\to 0}\delta_2(\epsilon) = 0$. Note that for jointly typical sequences $(\mathbf{y}, \underline{\mathbf{x}})$, the conditional probability $P^n_{Y|\underline{X}}(\mathbf{y}|\underline{\mathbf{x}})$ is upper bounded by $2^{-n(H(Y|\underline{X})-\delta_3(\epsilon))}$. Hence, we have:

$$p_{e_1}(\underline{\mathbf{s}}) \leq |\mathcal{L}_1(\underline{\mathbf{s}})| \times 2^{-nH(Y|\underline{X})}\frac{1}{q^{3n}} \times 2^{-n[2H(X_1|S_1V_1)+H(X_2|S_2V_2)+2H(X_3|S_3V_3)-\delta_4(\epsilon)]},$$

where $\delta_4(\epsilon) \to 0$ as $\epsilon \to 0$ and $|\mathcal{L}_1(\underline{\mathbf{s}})|$ is the cardinality of $\mathcal{L}_1(\underline{\mathbf{s}})$. Note that for ϵ_1-typical sequences $\underline{\mathbf{s}}$, the following inequality holds:

$$\frac{1}{n}\log_2|\mathcal{L}_1(\underline{\mathbf{s}})| \leq H(\underline{V}, \underline{X}, Y|\underline{S}) + H(S_1, V_1, X_1, S_3, V_3, X_3|YS_2\ V_2\ X_2) + \delta_5(\epsilon),$$

where $\delta_5(\epsilon) \to 0$ as $\epsilon \to 0$. Note that

$$\begin{aligned} H(\underline{V}, \underline{X}, Y|\underline{S}) &= H(\underline{V}|\underline{S}) + H(\underline{X}|\underline{S}, \underline{V}) + H(Y|\underline{X}) \\ &= 2\log_2 q + \sum_{i=1}^{3} H(X_i|S_i, V_i) + H(Y|\underline{X}), \end{aligned} \quad (17)$$

where the first equality holds by chain rule and the Markov chain $(\underline{S}, \underline{V}) \leftrightarrow \underline{X} \leftrightarrow Y$. The second equality holds, because, from (4), $\underline{V}$ are independent of the other random variables and $P_{V_1V_2V_3} = \frac{1}{q^2}\mathbb{1}\{V_3 = V_1 \oplus_q V_2\}$. Therefore, $p_{e_1} \to 0$ as $n \to \infty$, if

$$H(S_1, V_1, X_1, S_3, V_3, X_3|YS_2V_2X_2) \leq \log_2 q + H(X_1|S_1V_1) + H(X_3|S_3V_3). \quad (18)$$

Next, we simplify the right-hand side terms in (18). From (4), the Markov chain $(S_{i^c}, V_{i^c}, X_{i^c}) \leftrightarrow S_i \leftrightarrow X_i$ holds for all $i \in \{1, 2, 3\}$, where $i^c \triangleq \{1, 2, 3\}/\{i\}$. Therefore, the right-hand side above equals to

$$\log_2 q + H(X_1X_3|\underline{S}, V_1V_3X_2V_2) = H(X_1X_3V_1V_3|\underline{S}X_2V_2), \quad (19)$$

where the equality holds by chain rule and the following argument:

$$\begin{aligned} H(V_1V_3|\underline{S}X_2V_2) &= H(V_1|\underline{S}X_2V_2) = H(V_1|\underline{S}V_2) \\ &= H(V_1|V_2) = H(V_1) = \log_2\ q. \end{aligned}$$

We simplify the left-hand side in (18). Using the chain rule we obtain that

$$\begin{aligned} &H(S_1, V_1, X_1, S_3, V_3, X_3|YS_2\ V_2\ X_2) \\ &\quad = H(V_1, X_1, V_3, X_3|YS_2\ V_2\ X_2) + H(S_1\ S_3|YS_2\underline{V}\ \underline{X}) \\ &\quad = H(V_1, X_1, V_3, X_3|YS_2\ V_2\ X_2) + H(S_1|S_2\underline{V}\ \underline{X}), \end{aligned}$$

where the second equality holds due to the Markov chain $\underline{S} \leftrightarrow \underline{X} \leftrightarrow Y$ and the assumption that $S_1 \oplus_q S_2 \oplus_q S_3 = 0$. Note

that

$$\begin{aligned}H(S_1|S_2\underline{V}\ \underline{X}) &= H(S_1|S_2X_2V_2)\\ &\quad - I(S_1;X_1V_1X_3V_3|S_2V_2X_2)\\ &= H(S_1|S_2) - I(S_1;X_1V_1X_3V_3|S_2V_2X_2),\end{aligned}$$

where, the last equality holds because V_2 is independent of S_1 and X_2 is a function of (S_2, V_2). Therefore, using the above arguments, the inequality in (18) is simplified to

$$\begin{aligned}&H(S_1|S_2) \le I(S_1;X_1V_1X_3V_3|S_2V_2X_2)\\ &- H(V_1,X_1,V_3,X_3|YS_2\ V_2\ X_2) + H(X_1X_3V_1V_3|\underline{S}X_2V_2)\\ &= I(X_1V_1X_3V_3;Y|S_2V_2X_2)\\ &= I(X_1X_3;Y|S_2V_2X_2).\end{aligned}$$

As a result, $p_{e_1}(\underline{\mathbf{s}})$ can be made sufficiently small for large enough n, if the inequality

$$H(S_1|S_2) \le I(X_1X_3;Y|S_2V_2X_2)$$

is satisfied.

Case 2, $\tilde{\mathbf{s}}_1 = \mathbf{s}_1, \tilde{\mathbf{s}}_2 \neq \mathbf{s}_2$:

This case corresponds to $p_{e_2}(\underline{\mathbf{s}})$ which is defined using a similar expression as for $p_{e_1}(\underline{\mathbf{s}})$; but with the conditions in the second summation replaced with $\tilde{\underline{\mathbf{s}}} \neq \underline{\mathbf{s}}, \tilde{\mathbf{s}}_1 = \mathbf{s}_1, \tilde{\mathbf{v}}_1 = \mathbf{v}_1$. Therefore, we have

$$\begin{aligned}\mathcal{L}_2(\underline{\mathbf{s}}) \triangleq \Big\{(\underline{\mathbf{v}},\underline{\mathbf{x}},\mathbf{y},\tilde{\underline{\mathbf{s}}},\tilde{\underline{\mathbf{v}}},\tilde{\underline{\mathbf{x}}}) \in \mathcal{L}(\underline{\mathbf{s}}) : \tilde{\mathbf{s}}_1 = \mathbf{s}_1, \tilde{\mathbf{s}}_2 \neq \mathbf{s}_2,\\ \tilde{\mathbf{v}}_1 = \mathbf{v}_1, \tilde{\mathbf{x}}_1 = \mathbf{x}_1\Big\}.\end{aligned}$$

By symmetry and using a similar argument as in the first case, we can show that $p_{e_2}(\underline{\mathbf{s}}) \to 0$ as $n \to \infty$ if the following inequality holds

$$H(S_2|S_1) \le I(X_2X_3;Y|S_1V_1X_1).$$

Case 3, $\tilde{\mathbf{s}}_1 \neq \mathbf{s}_1, \tilde{\mathbf{s}}_2 \neq \mathbf{s}_2, \tilde{\mathbf{s}}_1 \oplus_q \tilde{\mathbf{s}}_2 = \mathbf{s}_1 \oplus_q \mathbf{s}_2$:

In this case, we have

$$\begin{aligned}P\Big\{\mathbf{v}_j = \mathbf{V}_j(\mathbf{s}_j), \tilde{\mathbf{v}}_j = &\mathbf{V}_j(\tilde{\mathbf{s}}_j),\ j=1,2\Big\} =\\ &q^{-3n}\mathbb{1}\Big\{\tilde{\mathbf{v}}_1 \oplus_q \tilde{\mathbf{v}}_2 = \mathbf{v}_1 \oplus_q \mathbf{v}_2\Big\}\end{aligned}$$

Further, we obtain that

$$\begin{aligned}&P\Big\{\mathbf{x}_l = \mathbf{X}_l(\mathbf{s}_l,\mathbf{v}_l), \tilde{\mathbf{x}}_l = \mathbf{X}_l(\tilde{\mathbf{s}}_l,\tilde{\mathbf{v}}_l), l=1,2,3\Big\} \le\\ &2^{-n[2H(X_1|S_1V_1)+2H(X_2|S_2V_2)+H(X_3|S_3V_3)-\delta_6(\epsilon)]}\mathbb{1}\{\tilde{\mathbf{x}}_3 = \mathbf{x}_3\}.\end{aligned}$$

By assumption $\mathbf{s}_1 \oplus_q \mathbf{s}_2 \oplus_q \mathbf{s}_3 = 0$ and $\mathbf{v}_1 \oplus_q \mathbf{v}_2 \oplus_q \mathbf{v}_3 = 0$. Therefore, the first probability is nonzero only when $\tilde{\mathbf{v}}_3 = \mathbf{v}_3$. Hence, as $\mathbf{s}_3 = \tilde{\mathbf{s}}_3$, we get $X_3(\tilde{\mathbf{s}}_3, \tilde{\mathbf{v}}_3) = X_3(\mathbf{s}_3, \mathbf{v}_3)$. As a result, we can define

$$\begin{aligned}\mathcal{L}_3(\underline{\mathbf{s}}) \triangleq \Big\{(\underline{\mathbf{v}},\underline{\mathbf{x}},\mathbf{y},\tilde{\underline{\mathbf{s}}},\tilde{\underline{\mathbf{v}}},\tilde{\underline{\mathbf{x}}}) \in \mathcal{L}(\underline{\mathbf{s}}) : \tilde{\mathbf{s}}_1 \neq \mathbf{s}_1, \tilde{\mathbf{s}}_2 \neq \mathbf{s}_2,\\ \tilde{\mathbf{s}}_1 \oplus_q \tilde{\mathbf{s}}_2 = \mathbf{s}_1 \oplus_q \mathbf{s}_2, \tilde{\mathbf{v}}_1 \oplus_q \tilde{\mathbf{v}}_2 = \mathbf{v}_1 \oplus_q \mathbf{v}_2, \tilde{\mathbf{x}}_3 = \mathbf{x}_3\Big\}.\end{aligned}$$

As a result, the contribution of this case (p_{e_3}) is bounded by

$$\begin{aligned}p_{e_3}(\underline{\mathbf{s}}) \le |\mathcal{L}_3(\underline{\mathbf{s}})|2^{-nH(Y|\underline{X})}\frac{1}{q^{3n}}\times\\ 2^{-n[2H(X_1|S_1V_1)+2H(X_2|S_2V_2)+H(X_3|S_3V_3)-\delta_7(\epsilon)]},\end{aligned}$$

Note that for ϵ_1-typical $\underline{\mathbf{s}}$, we have

$$\begin{aligned}\frac{1}{n}\log_2|\mathcal{L}_3(\underline{\mathbf{s}})| &\le H(\underline{V},\underline{X},Y|\underline{S})\\ &\quad + H(S_1,V_1,X_1,S_2,V_2,X_2|YS_3\ V_3\ X_3) + \delta_8(\epsilon).\end{aligned}$$

Therefore, from (17) and the above inequality, $p_{e_3}(\underline{\mathbf{s}}) \to 0$, if

$$\begin{aligned}H(S_1,V_1,X_1,S_2,V_2,X_2|YS_3\ V_3\ X_3) &\le \log_2\ q+\\ &\quad H(X_1|S_1V_1) + H(X_2|S_2V_2)\\ &= H(X_1,X_2,V_1,V_2|S_1S_2S_3V_3X_3),\end{aligned}$$

where the inequality above holds using a similar argument applied in (19). By symmetry and using a similar argument as in the first case, this inequality is equivalent to

$$H(S_1S_2|S_3) \le I(X_1,X_2;Y|S_3V_3X_3).$$

Case 4, $\tilde{\mathbf{s}}_1 \oplus a\tilde{\mathbf{s}}_2 = \mathbf{s}_1 \oplus a\mathbf{s}_2$, $a \in \mathbb{F}_q/\{0,1\}$:

From Lemma 1,

$$\begin{aligned}P\Big\{\mathbf{v}_j = \mathbf{V}_j(\mathbf{s}_j), \tilde{\mathbf{v}}_j = \mathbf{V}_j(\tilde{\mathbf{s}}_j),\ j=1,2\Big\} = q^{-3n}\times\\ \mathbb{1}\Big\{\tilde{\mathbf{v}}_1 \oplus_q a\tilde{\mathbf{v}}_2 = \mathbf{v}_1 \oplus_q a\mathbf{v}_2\Big\}.\end{aligned}$$

Therefore, the error probability in this case, i.e., $p_{e_4}(\underline{\mathbf{s}})$, satisfies

$$\begin{aligned}p_{e_4}(\underline{\mathbf{s}}) \triangleq \sum_{a=1}^{q-1}\sum_{\mathcal{L}_4(a,\underline{\mathbf{s}})} P^n_{Y|\underline{X}}(\mathbf{y}|\underline{\mathbf{x}})q^{-3n}\times\\ P\Big\{\mathbf{x}_l = \mathbf{X}_l(\mathbf{s}_l,\mathbf{v}_l), \tilde{\mathbf{x}}_l = \mathbf{X}_l(\tilde{\mathbf{s}}_l,\tilde{\mathbf{v}}_l),\ l=1,2,3\Big\},\end{aligned}$$

where

$$\begin{aligned}\mathcal{L}_4(a,\underline{\mathbf{s}}) \triangleq \Big\{(\underline{\mathbf{v}},\underline{\mathbf{x}},\mathbf{y},\tilde{\underline{\mathbf{s}}},\tilde{\underline{\mathbf{v}}},\tilde{\underline{\mathbf{x}}}) \in \mathcal{L}(\underline{\mathbf{s}}) : \tilde{\mathbf{s}}_1 \oplus a\tilde{\mathbf{s}}_2 = \mathbf{s}_1 \oplus a\mathbf{s}_2,\\ \tilde{\mathbf{v}}_1 \oplus a\tilde{\mathbf{v}}_2 = \mathbf{v}_1 \oplus a\mathbf{v}_2\Big\}.\end{aligned}$$

Also, observe that

$$\begin{aligned}P\Big\{\mathbf{x}_l = \mathbf{X}_l(\mathbf{s}_l,\mathbf{v}_l), \tilde{\mathbf{x}}_l = \mathbf{X}_l(\tilde{\mathbf{s}}_l,\tilde{\mathbf{v}}_l),\ l=1,2,3\Big\} \le\\ 2^{-2n[\sum_{i=1}^3 H(X_i|S_iV_i)]-\delta_9(\epsilon)}.\end{aligned}$$

where $\delta_9(\cdot)$ is a continuous function of ϵ with $\lim_{\epsilon\to 0}\delta_9(\epsilon) = 0$. Consequently, for any typical sequences $\underline{\mathbf{s}}$, the following upper bound holds:

$$\begin{aligned}p_{e_4}(\underline{\mathbf{s}}) \le \sum_{a=1}^{q-1}|\mathcal{L}_4(a,\underline{\mathbf{s}})|\times 2^{-nH(Y|\underline{X})}q^{-3n}\times\\ 2^{-2n[\sum_{i=1}^3 H(X_i|S_iV_i)]}2^{n\delta_{10}(\epsilon)}.\end{aligned}$$

Note that for any non-zero $a \in \mathbb{F}_q$ and any typical sequence $\underline{\mathbf{s}}$, the cardinality of $\mathcal{L}_4$ satisfies the inequality

$$\begin{aligned}\frac{1}{n}\log_2|\mathcal{L}_4(a,\underline{\mathbf{s}})| &\le H(\underline{V},\underline{X},Y|\underline{S})\\ &\quad + H(\underline{S},\underline{V},\underline{X}|Y,S_1 \oplus_q aS_2, V_1 \oplus_q aV_2) + \delta_{11}(\epsilon).\end{aligned}$$

Note that

$$H(\underline{V},\underline{X},Y|\underline{S}) = 2\log_2 q + \sum_{i=1}^3 H(X_i|S_i,V_i) + H(Y|\underline{X}).$$

Therefore, from the above inequalities, $p_{e_4}(\underline{\mathbf{s}}) \to 0$ as $n \to \infty$, if

$$H(\underline{S}, \underline{V}, \underline{X}|Y, S_1 \oplus_q aS_2, V_1 \oplus_q aV_2) < \log q + \sum_{i=1}^{3} H(X_i|S_i, V_i). \tag{20}$$

From the joint probability distribution given in (4), conditioned on $(\underline{S}, \underline{V})$ the random variables (X_1, X_2, X_3) are mutually independent. Hence,

$$\sum_{i=1}^{3} H(X_i|S_i, V_i) = H(\underline{X}|\underline{S}, \underline{V})$$

and the right-hand side of the above inequality simplifies to $\log q + H(\underline{X}|\underline{S}, \underline{V})$. Next, we simplify the left-hand side of the above inequality. For that we have

$$\begin{aligned}
&H(\underline{S}, \underline{V}, \underline{X}|Y, S_1 \oplus_q aS_2, V_1 \oplus_q aV_2) \\
&\quad = H(\underline{V}, \underline{X}|Y, S_1 \oplus_q aS_2, V_1 \oplus_q aV_2) \\
&\qquad + H(\underline{S}|S_1 \oplus_q aS_2, \underline{X}, \underline{V}) \\
&\quad = H(\underline{S}|S_1 \oplus_q aS_2, V_1 \oplus_q aV_2) \\
&\qquad - I(\underline{X}, \underline{V}; Y|S_1 \oplus_q aS_2, V_1 \oplus_q aV_2) \\
&\qquad + H(\underline{X}, \underline{V}|\underline{S}, S_1 \oplus_q aS_2, V_1 \oplus_q aV_2) \\
&\quad = H(\underline{S}|S_1 \oplus_q aS_2) \\
&\qquad - I(\underline{X}, \underline{V}; Y|S_1 \oplus_q aS_2, V_1 \oplus_q aV_2) \\
&\qquad + H(\underline{X}, \underline{V}|\underline{S}, V_1 \oplus_q aV_2),
\end{aligned}$$

where the first equality holds by chain rule and the Markov chain $\underline{S} \leftrightarrow \underline{X} \leftrightarrow Y$. The second equality holds by the definition of the mutual information. The last equality holds as (V_1, V_2, V_3) are independent of (S_1, S_2, S_3). As a result of the above argument, the inequality in (20) is equivalent to the following inequality:

$$\begin{aligned}
H(\underline{S}|S_1 \oplus_q aS_2) &< I(\underline{X}, \underline{V}; Y|S_1 \oplus_q aS_2, V_1 \oplus_q aV_2) \\
&\quad - H(\underline{X}, \underline{V}|\underline{S}, V_1 \oplus_q aV_2) + \log q + H(\underline{X}|\underline{S}, \underline{V}) \\
&= I(\underline{X}, \underline{V}; Y|S_1 \oplus_q aS_2, V_1 \oplus_q aV_2) - H(\underline{V}|V_1 \oplus_q aV_2) \\
&\quad - H(\underline{X}|\underline{S}, \underline{V}, V_1 \oplus_q aV_2) + \log q + H(\underline{X}|\underline{S}, \underline{V}) \\
&= I(\underline{X}, \underline{V}; Y|S_1 \oplus_q aS_2, V_1 \oplus_q aV_2) - H(\underline{V}|V_1 \oplus_q aV_2) \\
&\quad + \log q,
\end{aligned}$$

where the first equality holds by the chain rule and the fact that $\underline{V}$ is independent of $\underline{S}$. In what follows, we show that the last two terms above cancel each other. Since V_1 and V_2 are independent random variables with uniform distribution over $\mathbb{F}_q$, then so is V_1 and $V_1 \oplus_q aV_2$ for any $a \in \mathbb{F}_1/\{0\}$. Therefore, as $V_3 \oplus_q V_1 \oplus_q V_2 = 0$, we have that

$$\begin{aligned}
H(\underline{V}|V_1 \oplus_q aV_2) &= H(V_1, V_2|V_1 \oplus_q aV_2) \\
&= H(V_1, V_1 \oplus_q aV_2|V_1 \oplus_q aV_2) \\
&= H(V_1|V_1 \oplus_q aV_2) = \log q.
\end{aligned}$$

As a result, we showed that $p_{e_4}(\underline{\mathbf{s}}) \to 0$ as $n \to \infty$, if

$$H(\underline{S}|S_1 \oplus_q aS_2) \le I(\underline{X}, \underline{V}; Y|S_1 \oplus_q aS_2, V_1 \oplus_q aV_2).$$

Case 5, $\tilde{\mathbf{s}}_i \neq \mathbf{s}_i, i = 1, 2, 3$ and $\tilde{\mathbf{s}}_1 \oplus a\tilde{\mathbf{s}}_2 \neq \mathbf{s}_1 \oplus a\mathbf{s}_2, \forall \in \mathbb{F}_q$: Observe that,

$$\mathcal{L}_5(\underline{\mathbf{s}}) \triangleq \Big\{(\underline{\mathbf{v}}, \underline{\mathbf{x}}, \underline{\mathbf{y}}, \underline{\tilde{\mathbf{s}}}, \underline{\tilde{\mathbf{v}}}, \underline{\tilde{\mathbf{x}}}) \in \mathcal{L}(\underline{\mathbf{s}}) : \tilde{\mathbf{s}}_1 \neq \mathbf{s}_1, \tilde{\mathbf{s}}_2 \neq \mathbf{s}_2, \tilde{\mathbf{s}}_1 \oplus a\tilde{\mathbf{s}}_2 \neq \mathbf{s}_1 \oplus a\mathbf{s}_2, \forall a \in \mathbb{F}_q\Big\}.$$

In addition, we obtain that

$$P\Big\{\mathbf{v}_j = \mathbf{V}_j(\mathbf{s}_j), \tilde{\mathbf{v}}_j = \mathbf{V}_j(\tilde{\mathbf{s}}_j),\ j = 1, 2\Big\} = q^{-4n},$$

and that

$$P\Big\{\mathbf{x}_l = \mathbf{X}_l(\mathbf{s}_l, \mathbf{v}_l), \tilde{\mathbf{x}}_l = \mathbf{X}_l(\tilde{\mathbf{s}}_l, \tilde{\mathbf{v}}_l),\ l = 1, 2, 3\Big\} \le 2^{-2n[\sum_{l=1}^{3} H(X_l|S_lV_l) - \delta_9(\epsilon)]}.$$

Therefore, the contribution of this case is simplified to

$$p_{e_5}(\underline{\mathbf{s}}) \approx q^{-2n} 2^{nH(\underline{S}, \underline{V}, \underline{X}|Y)} 2^{-n\sum_{l=1}^{3} H(X_l|S_lV_l)}.$$

As a result, one can show that $P_{e_5} \to 0$, if

$$H(S_1S_2S_3) \le I(X_1X_2X_3; Y).$$

Finally, note that $P_e(\underline{\mathbf{s}}) = \sum_{i=1}^{5} P_{ei}(\underline{\mathbf{s}})$. Moreover, $P_{ei}(\underline{\mathbf{s}})$ depends on $\underline{\mathbf{s}}$ only through its PMF. Therefore, for any typical $\underline{\mathbf{s}}$, P_e approaches zero as $n \to \infty$, if the following bounds are satisfied:

$$\begin{aligned}
H(S_1|S_2) &\le I(X_1\ X_3; Y|S_2\ V_2\ X_2) \\
H(S_2|S_1) &\le I(X_2\ X_3; Y|S_1\ V_1\ X_1) \\
H(S_1S_2|S_1 \oplus_q S_2) &\le I(X_1\ X_2; Y|S_1 \oplus_q S_2, V_3\ X_3) \\
H(S_1S_2|S_1 \oplus_q aS_2) &\le I(X_1, X_2, X_3; Y|S_1 \oplus_q aS_2, V_1 \oplus_q aV_2) \\
H(S_1, S_2) &\le I(X_1\ X_2\ X_3; Y).
\end{aligned}$$

■

APPENDIX B
PROOF OF THEOREM 2

Lemma 2: For the MAC in Example 3, it holds that

$$I(X_1, X_2, X_3; Y) \le 2 - H(N),$$

with equality if and only if $X_3 = X_1 \oplus_2 X_2$, with probability one, and X_3 is uniform over $\{0, 1\}$.

Proof: Note that

$$I(X_1, X_2, X_3; Y) = H(Y) - H(N).$$

We proceed by finding all the necessary and sufficient conditions on P_{X_1, X_2, X_3} for which Y is uniform over $\mathbb{Z}_4$. From Figure 4,

$$Y = (X_1 \oplus_2 X_2) \oplus_4 X_3 \oplus_4 N.$$

Let $X_2' = X_1 \oplus_2 X_2$ and $P(X_2' \oplus_4 X_3 = i) = q(i)$, where $i = 1, 2, 3, 4$. Since X_2' and X_3 are binary, the $q(3) = 0$. Given the distribution of N is Table I, the distribution of Y is as follows:

$$P(Y = 0) = q(0)(\frac{1}{2} - \delta) + q(2)\delta, \tag{21a}$$

$$P(Y = 1) = q(0)\frac{1}{2} + q(1)(\frac{1}{2} - \delta), \tag{21b}$$

$$P(Y = 2) = q(0)\delta + q(1)\frac{1}{2} + q(2)(\frac{1}{2} - \delta), \tag{21c}$$

$$P(Y = 3) = q(2)\frac{1}{2} + q(1)\delta. \tag{21d}$$

It's not difficult to check that the only solution for the equations in (21) is

$$q(0) = q(2) = \frac{1}{2}, \quad q(1) = 0.$$

Note that by definition

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

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

$$1)\ X_3 = X_1 \oplus_2 X_2, \qquad 2)\ X_3 \text{ is uniform over } \{0, 1\}.$$

■

Lemma 3: Let $\mathscr{P}_1$ be the set of all distributions $P^*_{X_1,X_2,X_3}$ that satisfies the conditions in Lemma 2. Let $\mathscr{P}_2$ be the set of all distributions P_{X_1,X_2,X_3} which is the marginal of

$$P_{S_1,S_2,S_3} P_{X_1,X_2,X_3|S_1,S_2,S_3}$$

for some source triplet (S_1, S_2, S_3) in Example 3 with parameters $\sigma \in (0, \frac{1}{2}], \gamma \in (0, \gamma^*]$ and conditional distribution of the form

$$P_{X_1,X_2,X_3|S_1,S_2,S_3} = \prod_{i=1}^{3} P_{X_i|S_i}.$$

Then the total variation distance between $\mathscr{P}_1$ and $\mathscr{P}_2$ satisfies

$$TV(\mathscr{P}_1, \mathscr{P}_2) \geq \frac{1}{6} - \frac{\gamma^*}{3}.$$

Moreover, there exists $\alpha(\gamma^*) > 0$ such that $\forall P_{X_1,X_2,X_3} \in \mathscr{P}_2$

$$I(X_1, X_2, X_3; Y) \leq 2 - H(N) - \alpha(\gamma^*).$$

Proof: Let $\overline{\gamma^*} \triangleq 1 - \gamma^*$ and assume for some $\epsilon \geq 0$ there exist sources with parameters $\sigma_\epsilon \in (0, \frac{1}{2}]$ and $\gamma_\epsilon \in (0, \gamma^*]$ and conditional distributions $P^\epsilon_{X_i|S_i}, i = 1, 2, 3$ and a distribution $P^*_{X_1,X_2,X_3}$ satisfying the conditions in Lemma 2 such that total variation distance between the resulted PMF $P^\epsilon_{X_1,X_2,X_3}$ and $P^*_{X_1,X_2,X_3}$ is equal to ϵ. Then, for $P^\epsilon_{X_1,X_2,X_3}$, the following inequalities hold:

$$P^\epsilon(X_3 \neq X_1 \oplus X_2) \leq \epsilon, \quad \text{and} \quad \left| P^\epsilon(X_3 = 1) - \frac{1}{2} \right| \leq \epsilon. \quad (22)$$

The second inequality implies

$$\gamma_\epsilon\ P^\epsilon_{X_3|S_3}(1|1) + \overline{\gamma_\epsilon}\ P^\epsilon_{X_3|S_3}(1|0) \in [\frac{1}{2} - \epsilon, \frac{1}{2} + \epsilon], \quad (23a)$$

$$\gamma_\epsilon\ P^\epsilon_{X_3|S_3}(0|1) + \overline{\gamma_\epsilon}\ P^\epsilon_{X_3|S_3}(0|0) \in [\frac{1}{2} - \epsilon, \frac{1}{2} + \epsilon]. \quad (23b)$$

Since the first terms in (23a) and (23b) are non-negative and $\gamma_\epsilon \leq \gamma^*$, then

$$\frac{1}{2} + \epsilon \geq \overline{\gamma_\epsilon}\ P^\epsilon_{X_3|S_3}(1|0) \geq \overline{\gamma^*}\ P^\epsilon_{X_3|S_3}(1|0),$$
$$\frac{1}{2} + \epsilon \geq \overline{\gamma_\epsilon}\ P^\epsilon_{X_3|S_3}(0|0) \geq \overline{\gamma^*}\ P^\epsilon_{X_3|S_3}(0|0).$$

Since $P^\epsilon_{X_3|S_3}(0|0) + P^\epsilon_{X_3|S_3}(1|0) = 1$, then the above inequalities imply the following

$$\frac{1}{2} + \epsilon \geq \overline{\gamma^*}\ P^\epsilon_{X_3|S_3}(1|0) \geq \overline{\gamma^*} - \frac{1}{2} - \epsilon \quad (24a)$$

$$\frac{1}{2} + \epsilon \geq \overline{\gamma^*}\ P^\epsilon_{X_3|S_3}(0|0) \geq \overline{\gamma^*} - \frac{1}{2} - \epsilon \quad (24b)$$

From the law of total probability, the first condition in (22) is equivalent to

$$\sum_{\underline{s}} \sum_{x_1,x_2} P^\epsilon_{\underline{S}}(\underline{s}) P^\epsilon_{X_1|S_1}(x_1|s_1) P^\epsilon_{X_2|S_2}(x_2|s_2) P^\epsilon_{X_3|S_3}(\overline{x_1 \oplus x_2}|s_3) \leq \epsilon,$$

where $P^\epsilon_{\underline{S}}$ is the joint PMF of the sources with parameters $\sigma_\epsilon, \gamma_\epsilon$, and

$$\overline{x_1 \oplus_2 x_2} \triangleq 1 \oplus_2 x_1 \oplus_2 x_2.$$

By considering the case in which $s_1 = s_2 = s_3 = 0$, the above inequality implies that

$$\begin{aligned}
\epsilon &\geq \sum_{x_1,x_2} \overline{\gamma_\epsilon}\ \overline{\sigma_\epsilon}\ P^\epsilon_{X_1|S_1}(x_1|0) P^\epsilon_{X_2|S_2}(x_2|0) P^\epsilon_{X_3|0}(\overline{x_1 \oplus_2 x_2}|0) \\
&\geq \sum_{x_1,x_2} \overline{\gamma^*}\ \frac{1}{2}\ P^\epsilon_{X_1|S_1}(x_1|0) P^\epsilon_{X_2|S_2}(x_2|0) P^\epsilon_{X_3|0}(\overline{x_1 \oplus_2 x_2}|0) \\
&\geq \sum_{x_1,x_2} \frac{1}{2}\ (\overline{\gamma^*} - \frac{1}{2} - \epsilon) P^\epsilon_{X_1|S_1}(x_1|0) P^\epsilon_{X_2|S_2}(x_2|0) \\
&= \frac{1}{2}\ (\overline{\gamma^*} - \frac{1}{2} - \epsilon),
\end{aligned}$$

where the third inequality holds from the bounds in (24). As a result, these inequalities imply that $\epsilon \geq \frac{1}{3}(\overline{\gamma^*} - \frac{1}{2})$. From Lemma 2 and the continuity of the mutual information in total variation distance [43], the second statement of the lemma follows. ■

Lemma 4: For the setup in Example 3, there exists $\epsilon > 0$ such that any source triple (S_1, S_2, S_3) with parameters $(\sigma > 0, \gamma \geq \gamma^* - \epsilon)$ does not satisfy the sufficient conditions stated in Proposition 1.

Proof: We prove the lemma by a contradiction. Suppose $\forall \epsilon > 0$ there exist $\sigma > 0$ and $\gamma \geq \gamma^* - \epsilon$ such that the sufficient conditions in Proposition 1 are satisfied. Consider the fourth inequality in Proposition 1. Since $\sigma > 0$ there is no common part. Let $U' = U_{123} U_{12} U_{13} U_{23}$. Then, the following holds

$$h(\gamma) + h(\sigma) \leq \max_{p(u')p(\underline{x}|u'\underline{s})} I(X_1 X_2 X_3; Y|U'), \quad (25)$$

where

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

Since U' is independent of the sources, and appears in the conditioning in the mutual information term, the inequality in (25) is equivalent to

$$h(\gamma) + h(\sigma) \leq \max_{p(\underline{x}|\underline{s})} I(X_1 X_2 X_3; Y), \quad (26)$$

where $p(\underline{s}, \underline{x}) = p(\underline{s}) p(x_1|s_1) p(x_2|s_2) p(x_3|s_3)$. From Lemma 3, the right-hand side in (26) is less than $2 - H(N) - \alpha$, for some $\alpha > 0$ (which depends only on γ^* which is a function of δ). As $h(\gamma^*) = 2 - H(N)$, by the bound above,

$$h(\gamma) + h(\sigma) \leq h(\gamma^*) - \alpha.$$

Thus, as $h(\sigma) > 0$, we get that $h(\gamma) < h(\gamma^*) - \alpha$. By the continuity and monotonicity of the binary entropy function,

$$\gamma < h^{-1}(h(\gamma^*) - \alpha) = \gamma^* - \lambda(\alpha),$$

where $\lambda(\alpha) > 0$. Hence, as $\gamma \geq \gamma^* - \epsilon$, then ϵ must be greater than $\lambda(\alpha)$ which is a contradiction. ∎

Lemma 5: There exists a non-negative function $\sigma_0(\gamma)$ such that 1) $\sigma_0(\gamma) > 0$ for all $\gamma \in [0, \gamma^*)$, and 2) any source with parameters (γ, σ) satisfying $0 \leq \gamma \leq \gamma^*$, and $0 \leq \sigma \leq \sigma_0(\gamma)$ is transmissible.

Proof: For the setup in Example 3, the bounds given in Theorem 1 are simplified to

$$h(\gamma) \leq I(X_2X_3; Y|X_1S_1V_1) \tag{27a}$$
$$h(\sigma) \leq I(X_1X_2; Y|X_3S_3V_3) \tag{27b}$$
$$h(\gamma) + h(\sigma) - h(\sigma * \gamma) \leq I(X_1X_3; Y|X_2S_2V_2) \tag{27c}$$
$$h(\gamma) + h(\sigma) \leq I(X_1X_2X_3; Y). \tag{27d}$$

Let $E_1 \sim Ber(\alpha)$, and set $X_1 = V_1 \oplus E_1$ and $X_2 = V_2, X_3 = V_3$, where (V_1, V_2, V_3) are as in Theorem 1; that is they are pairwise independent Bernoulli random variables with joint PMF $P_{V_1,V_2,V_3} = \frac{1}{4}\mathbb{1}\{V_3 = V_1 \oplus_2 V_2\}$. Next, using these random variables, we further simplify the conditions in (27).

We start by the first condition given in (27a). The right-hand side is simplified to

$$\begin{aligned} I(X_2X_3; Y|X_1S_1V_1) &= H((X_1 \oplus_2 X_2) \oplus_4 X_3 \oplus_4 N|X_1V_1) \\ &\quad - H(N) \\ &= H((E_1 \oplus_2 V_1 \oplus_2 V_2) \oplus_4 (V_1 \oplus_2 V_2) \oplus_4 N|E_1, V_1) \\ &\quad - H(N) \\ &= P(E_1 = 0)\ \big[H((V_1 \oplus_2 V_2) \oplus_4 (V_1 \oplus_2 V_2) \oplus_4 N|V_1) \\ &\quad - H(N)\big] \\ &= (1 - \alpha)(2 - H(N)), \end{aligned} \tag{28}$$

where the first equality holds as

$$Y = (X_1 \oplus_2 X_2) \oplus_4 X_3 \oplus_4 N$$

and $X_i, i = 1, 2, 3$ are independent of the sources. The fourth equality holds as $H(X \oplus_4 X \oplus_4 N) = 2$ and

$$H((1 \oplus_2 X) \oplus_4 X \oplus_4 N) = H(N)$$

when X is uniform over $\{0, 1\}$. Therefore, from (28), the first condition gives $h_b(\gamma) \leq 2 - H(N)$. This condition is always satisfied for any $\gamma \leq \gamma^*$. This is due to the monotonicity of the binary entropy function.

Next, we evaluate the second condition given by (27b). Using a similar argument, the right-hand side of (27b) is simplified to

$$\begin{aligned} I(X_1X_2; Y|X_3S_3V_3) &= H((X_1 \oplus_2 X_2) \oplus_4 N|X_3V_3) \\ &\quad - H(N) \\ &= H(E_1 \oplus_4 N) - H(N). \end{aligned} \tag{29}$$

Hence, the second condition gives $h_b(\sigma) \leq \eta_1(\alpha)$, where $\eta_1(\alpha) \triangleq H(E_1 \oplus_4 N) - H(N)$. We show that $\eta_1(\alpha)$ is strictly positive for all $\alpha \in (0, \frac{1}{2}]$. For that we have $H(N) = 1 + \frac{1}{2}h_b(2\delta)$ and

$$\begin{aligned} H(E_1 \oplus_4 N) &= 1 + \frac{1}{2}[h_b(2\alpha\delta) + h_b(2(1 - \alpha)\delta + \alpha)] \\ &\geq 1 + \frac{1}{2}[h_b(2\alpha\delta) + (1 - \alpha)h_b(2\delta)], \end{aligned}$$

where the first inequality holds due to the convexity of binary entropy function and the fact that $h_b(1) = 0$. Hence,

$$\eta_1(\alpha) \geq \frac{1}{2}[h_b(2\alpha\delta) - \alpha h_b(2\delta)].$$

When $\delta \in (0, \frac{1}{4}]$, the equality $h_b(2\alpha\delta) = \alpha h_b(2\delta)$ holds if and only if $\alpha \in \{0, 1\}$. As a result of this and due to the convexity of binary entropy, the strict inequality $h_b(2\alpha\delta) > \alpha h_b(2\delta)$ holds.

For the third condition, the right-hand side of (27c) equals to

$$I(X_1\, X_3; Y|X_2\, S_2\, V_2) = H\Big((V_1 \oplus_2 E_1) \oplus_4 V_1 \oplus_4 N\Big) - H(N).$$

As for the fourth condition, the right-hand side of (27d) is simplified to

$$\begin{aligned} I(X_1\ X_2\ X_3; Y) = H((E_1 \oplus_2 V_1 \oplus_2 V_2) \oplus_4 (V_1 \oplus_2\ V_2) \oplus_4 N) \\ - H(N). \end{aligned}$$

Since V_1 and $V_1 \oplus_2 V_2$ are both uniform over $\{0, 1\}$, then the above two terms are equal. Let

$$\eta_2(\alpha) \triangleq 2 - H((V_1 \oplus_2 E_1) \oplus_4 V_1 \oplus_4 N).$$

Note that $0 \leq \eta_2(\alpha) \leq 2 - H(N)$. Moreover, from Lemma 2, $\eta_2(\alpha)$ is strictly positive for any $\alpha \in (0, \frac{1}{2}]$. With this argument, the third and fourth conditions become

$$\begin{aligned} h(\gamma) + h(\sigma) - h(\sigma * \gamma) &\leq 2 - H(N) - \eta_2(\alpha), \\ h(\gamma) + h(\sigma) &\leq 2 - H(N) - \eta_2(\alpha). \end{aligned}$$

Since the right-hand sides are equal and $h(\sigma * \gamma) \geq 0$, the third condition is trivial.

As a result of the above argument, we obtain the following sufficient conditions:

$$h(\gamma) \leq (1 - \alpha)[2 - H(N)] \tag{30a}$$
$$h(\sigma) \leq \eta_1(\alpha) \tag{30b}$$
$$h(\gamma) + h(\sigma) \leq 2 - H(N) - \eta_2(\alpha) \tag{30c}$$

For any $\gamma \leq \gamma^*$, inequality (30a) holds if

$$\alpha \leq 1 - \frac{h_b(\gamma)}{h_b(\gamma^*)}.$$

Note that $\eta_1(\alpha) > 0$ and $\eta_2(\alpha) > 0$ for all $\alpha \in (0, \frac{1}{2}]$, and $\eta_1(0) = \eta_2(0) = 0$. Further, they are continuous functions of α with $\lim_{\alpha \to 0} \eta_i(\alpha) = 0, i = 1, 2$. Therefore, for any $\gamma < \gamma^*$, there exists $\alpha_0 > 0$ such that for any $\alpha \in (0, \alpha_0)$, inequality (30a) holds and

$$h_b(\gamma^*) - h_b(\gamma) - \eta_2(\alpha) > 0.$$

TABLE II
THE DECODING AND ENCODING PROCESSES FOR USER 1 IN BLOCKS $l = 1, 2, 3$

	$l=1$	$l=2$	$l=3$
Decoding 1	$\mathbf{v}_{1,[1]} = \mathbf{0}$	$\hat{\mathbf{w}}_{\mathbf{A}_1,[1]},\ \mathbf{v}_{1,[2]} = \mathbf{t}_{\mathbf{A}_1,[1]}$	$\mathbf{w}_{\mathbf{A}_1,[2]}, \mathbf{v}_{1,[3]} = \mathbf{t}_{\mathbf{A}_1,[2]}$
Decoding 2	—	—	$(\hat{\mathbf{w}}_{2,[1]}, \hat{\mathbf{w}}_{3,[1]})$
Encoding 1	$M_{0,[1]} = 1, \mathbf{u}_{[1]}$	$M_{0,[2]} = 1, \mathbf{u}_{[2]}$	$\mathcal{L}_{[1]},\ M_{0,[3]} = \lvert\mathcal{L}_{[1]}\rvert, \mathbf{u}_{[3]}$
Encoding 2	$\mathbf{w}_{1,[1]}, \mathbf{t}_{1,[1]}(\mathbf{w}_{1,[1]})$	$\mathbf{w}_{1,[2]}, \mathbf{t}_{1,[2]}(\mathbf{w}_{1,[2]})$	$\mathbf{w}_{1,[3]}, \mathbf{t}_{1,[3]}(\mathbf{w}_{1,[3]})$
Encoding 3	$\mathbf{x}_1(\mathbf{u}_{[1]}, \mathbf{t}_{1,[1]}, \mathbf{v}_{1,[1]}, \mathbf{w}_{1,[1]})$	$\mathbf{x}_1(\mathbf{u}_{[2]}, \mathbf{t}_{1,[2]}, \mathbf{v}_{1,[2]}, \mathbf{w}_{1,[2]})$	$\mathbf{x}_1(\mathbf{u}_{[3]}, \mathbf{t}_{1,[3]}, \mathbf{v}_{1,[3]}, \mathbf{w}_{1,[3]})$

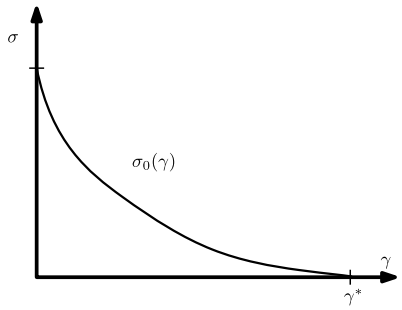

Fig. 10. The parameters σ and γ described in Lemma 5.

For any $\gamma \le \gamma^*$, define

$$\sigma_0(\gamma) \triangleq h_b^{-1}\Big(\max_{0\le\alpha\le 1-\frac{h_b(\gamma)}{h_b(\gamma^*)}} \min\Big\{\eta_1(\alpha), h_b(\gamma^*) - \eta_2(\alpha) - h_b(\gamma)\Big\}\Big).$$

Note that the inequalities in (30) are satisfied for $\gamma \le \gamma^*$ and $\sigma = \sigma_0(\gamma)$. Hence, from the monotonicity of binary entropy function, these inequalities are also satisfied for $\sigma \le \sigma_0(\gamma)$. This implies that any source with such parameters are transmissible. ∎

The final step in our argument is as follows. Fix $\gamma \in (\gamma^* - \epsilon, \gamma^*)$, where ϵ is as in Lemma 4. From Lemma 5, the source with such γ and the parameter $\sigma = \sigma_0(\gamma) > 0$ is transmissible; whereas from Lemma 4 it is not transmissible using CES. Figure 10 shows the set of parameters whose sources are transmissible.

APPENDIX C
PROOF OF THEOREM 3

A. Codebook Construction

We build upon a class of codes called Quasi Linear Codes (QLCs) [46]. A QLC is defined as a subset of a linear code. By definition, any linear codebook can be viewed as the image of a linear transformation $\phi : \mathbb{F}_q^k \mapsto \mathbb{F}_q^n$, where q is a prime number. In another words, the codewords of such a linear code are $\phi(\mathbf{u}^k), \mathbf{u}^k \in \mathbb{F}_q^k$. In this representation, a QLC over a finite field $\mathbb{F}_q$ is defined as

$$\mathcal{C} \triangleq \{\phi(\mathbf{u}) : \mathbf{u} \in \mathcal{U}\}, \tag{31}$$

where $\mathcal{U}$ is a given subset of $\mathbb{F}_q^k$. If $\mathcal{U} = \mathbb{F}_q^k$, then $\mathcal{C}$ is a linear codebook.

We begin the proof by generating a QLC for each user. Let (W_1, W_2, W_3) be the random variables as in the statement of the theorem. For a fixed $\epsilon \in (0, 1)$, consider the set of all ϵ-typical sequences $\mathbf{w}_i^k$. Without loss of generality assume that the new message at the ith encoder is a sequence $\mathbf{w}_i^k$ which is selected randomly and uniformly from $A_\epsilon^{(k)}(W_i)$. In this case $M_i = |A_\epsilon^{(k)}(W_i)|, i = 1, 2, 3$.

We generate three codebooks for each user at each block $l \in [1, L]$. The codebook generations are described in the following:

Codebook 1: For each block $l \in [1, L]$ generate $M_{0,[l]}$ sequences randomly and independently according to P_U^n. The parameter $M_{0,[l]}$ is to be defined later. Denote such sequences by $\mathbf{U}_{[l]}(m)$, where $m \in [1, M_{0,[l]}]$.

Codebook 2: At each user $i = 1, 2, 3$ and for any vector $\mathbf{w}_i^k \in \mathbb{F}_q^k$, denote

$$\mathbf{t}_i(\mathbf{w}_i^k) \triangleq \mathbf{w}_i^k \mathbf{G} \oplus_q \mathbf{b}_i^n, \quad i = 1, 2, 3,$$

where $\mathbf{G}$ is a $k \times n$ matrix with elements chosen randomly and uniformly from $\mathbb{F}_q$, and $\mathbf{b}_i^n$ is a vector selected randomly and uniformly from $\mathbb{F}_q^n$.

Codebook 3: For each user $i = 1, 2, 3$ and given $\mathbf{u}^n \in \mathcal{U}^n$ and $\mathbf{t}^n, \mathbf{v}^n \in \mathbb{F}_q^n$ generate M_i sequences randomly and independently according to the conditional distribution $\prod_{j=1}^n P(\cdot|u_j, t_j, v_j)$. Denote such sequences by $\mathbf{x}_i(\mathbf{u}^n, \mathbf{t}^n, \mathbf{v}^n, m_i)$, where $m_i \in [1 : M_i], i = 1, 2, 3$.

Initialization: Note that we are using the following notation: the subscript with bracket $[\cdot]$ denotes the index of a block, subscript without a bracket denotes the index of a user, and in line bracket $(\cdot)$ denotes the index of a codeword in a corresponding codebook. When it is clear from the context, we drop the index of the codewords.

For block $l = 0$, set $M_{0,[0]} = 1$. For block $l = 1$, set $M_{0,[1]} = 1, \mathbf{v}_{i,[1]} = \mathbf{0}$ for $i = 1, 2, 3$. For block $l = 2$, set $M_{0,[2]} = 1$. Let $\mathbf{A} \in \mathbb{F}_{\mathbf{q}}^{\mathbf{3}\times\mathbf{3}}$, and by a_{ij} denote the element in ith row and jth column. By $\mathbf{A}_i, i = 1, 2, 3$, denote the ith column of $\mathbf{A}$. At each block User i intends to decode a linear combination of the messages with coefficients determined by $\mathbf{A}_i$.

B. Encoding and Decoding

Block $l = 1$: At block $l = 1$, a new message $\mathbf{w}_{i,[1]} \in A_\epsilon^{(k)}(W_i)$, $i = 1, 2, 3$, is observed by the ith user. Given the message, the ith encoder calculates $\mathbf{t}_i(\mathbf{w}_{i,[1]})$. This sequence is denoted by $\mathbf{t}_{i,[1]}$. Next, the encoder sends $\mathbf{x}_i(\mathbf{u}_{[1]}, \mathbf{t}_{i,[1]}, \mathbf{v}_{i,[1]}, \mathbf{w}_{i,[1]})$ over the channel. For shorthand, we denote such sequence by $\mathbf{x}_{i,[1]}$. The encoding and decoding processes in this block are shown in Table II.

Block $l = 2$:

At the beginning of this block, each user receives $Y_{[1]}$ as feedback from the channel. User $i, i = 1, 2, 3$, wishes to decode the linear combination

$$\mathbf{w}_{\mathbf{A}_i,[1]} \triangleq a_{1i}\mathbf{w}_{1,[1]} \oplus a_{2i}\mathbf{w}_{2,[1]} \oplus a_{3i}\mathbf{w}_{3,[1]}.$$

Since, $\mathbf{w}_{i,[1]}$ is known at User i, then it finds a sequence

$$\hat{\mathbf{w}}_{\mathbf{A}_i,[1]} \in A_\epsilon^{(k)}(W_{\mathbf{A_i}}|\mathbf{w}_{i,[1]})$$

such that

$$(\hat{\mathbf{w}}_{\mathbf{A}_i,[1]}\mathbf{G} \oplus \mathbf{b}_{\mathbf{A}_i}, Y_{[1]}) \in A_\epsilon^{(n)}(T_{\mathbf{A}_i}, Y|\mathbf{u}_{[1]}, \mathbf{t}_{1,[1]}, \mathbf{x}_{1,[1]}), \tag{32}$$

where

$$\begin{aligned} W_{\mathbf{A}_i} &\triangleq a_{1i}W_1 \oplus a_{2i}W_2 \oplus a_{3i}W_3,\\ \mathbf{b}_{\mathbf{A}_i} &\triangleq a_{1i}\mathbf{b}_1 \oplus a_{2i}\mathbf{b}_2 \oplus a_{3i}\mathbf{b}_3,\\ T_{\mathbf{A}_i} &\triangleq a_{1i}T_1 \oplus a_{2i}T_2 \oplus a_{3i}T_3. \end{aligned}$$

A decoding error $E_{i,[2]}, i = 1, 2, 3$, is declared if $\hat{\mathbf{w}}_{\mathbf{A}_i,[1]}$ is not found or is not unique. If it is unique, the encoder sets

$$\mathbf{v}_{i,[2]} = \hat{\mathbf{w}}_{\mathbf{A}_i,[1]}\mathbf{G} \oplus \mathbf{b}_{\mathbf{A}_i}.$$

Otherwise, $\mathbf{v}_{i,[2]}$ is generated at random from $\mathbb{F}_q^n$.

Next, a new message $\mathbf{w}_{i,[2]}, i = 1, 2, 3$, is observed at the ith encoder. Similar to the encoding process at the first block, the ith encoder calculates $\mathbf{t}_{i,[2]}$ and sends $\mathbf{x}_i(\mathbf{u}_{[2]}, \mathbf{t}_{i,[2]}, \mathbf{v}_{i,[2]}, \mathbf{w}_{i,[2]})$. For shorthand, such sequence is denoted by $\mathbf{x}_{i,[2]}$. The encoding and decoding processes in this block are shown in Table II.

Block $l > 2$: Each user performs two decoding and three encoding processes in this block. It is assumed that each encoder knows the common information given by $\mathbf{u}_{[l-2]}$ and $\mathbf{u}_{[l-1]}$. For $l = 3$, this is clear because $M_{0,[1]} = M_{0,[2]} = 1$. We will explain how this knowledge is acquired, and how $\mathbf{u}_{[l]}$ is generated after describing the decoding process.

The first decoding process is the same as the decoding process in block $l = 2$. At the beginning of the block $l > 2$, User i observes $Y_{[l-1]}$ as feedback from the channel and wishes to decode the linear combination

$$\mathbf{w}_{\mathbf{A}_i,[l-1]} \triangleq a_{1i}\mathbf{w}_{1,[l-1]} \oplus a_{2i}\mathbf{w}_{2,[l-1]} \oplus a_{3i}\mathbf{w}_{3,[l-1]}.$$

This decoding process is the same as the one in block $l = 2$; it is successful, if the sequences $\hat{\mathbf{w}}_{\mathbf{A}_i,[l-1]}$ is unique. Then, the codeword $\mathbf{v}_{i,[l]}$ is generated at User i, where $i = 1, 2, 3$. If the decoding process at User $i, i = 1, 2, 3$, is not successful, an error event $E_{i,[l]}$ is declared and a codeword $\mathbf{v}_{i,[l]}$ is generated at random.

Next, we explain the second decoding process. Given $(\mathbf{Y}_{[l-2]}, \mathbf{Y}_{[l-1]})$, User i decodes the messages of the other two encoders from block $l-2$. For that, User 1 finds unique $\hat{\mathbf{w}}_{2,[l-2]} \in A_\epsilon^{(k)}(W_2)$ and $\hat{\mathbf{w}}_{3,[l-2]} \in A_\epsilon^{(k)}(W_3)$ such that the conditions in (33), shown at the bottom of the page, hold, where $\mathbf{u}_{[l-1]}, \mathbf{u}_{[l-2]}, \mathbf{v}_{i,[l-2]}$ are known at the encoder from the previous blocks and

$$\begin{aligned} &\hat{\mathbf{t}}_{i,[l-2]} \triangleq \mathbf{t}_i\big(\hat{\mathbf{w}}_{i,[l-2]}\big),\\ &\hat{\mathbf{x}}_{i,[l-2]} \triangleq \mathbf{x}_i\big(\mathbf{u}_{[l-2]}, \hat{\mathbf{t}}_{i,[l-2]}, \mathbf{v}_{i,[l-2]}, \hat{\mathbf{w}}_{i,[l-2]}\big),\\ &\hat{\mathbf{x}}_{i,[l-2]} \triangleq \mathbf{x}_i\big(\mathbf{u}, \hat{\mathbf{t}}_{i,[l-2]}, \mathbf{v}_{i,[l-2]}, \hat{\mathbf{w}}_{i,[l-2]}\big),\\ &\hat{\mathbf{v}}_{2,[l-1]} \triangleq \big(a_{1,2}\mathbf{w}_{1,[l-2]} \oplus a_{2,2}\hat{\mathbf{w}}_{2,[l-2]} \oplus a_{3,2}\hat{\mathbf{w}}_{3,[l-2]}\big)\mathbf{G} \oplus \mathbf{b}_{\mathbf{A}_2},\\ &\hat{\mathbf{v}}_{3,[l-1]} \triangleq \big(a_{1,3}\mathbf{w}_{1,[l-2]} \oplus a_{2,3}\hat{\mathbf{w}}_{2,[l-2]} \oplus a_{3,3}\hat{\mathbf{w}}_{3,[l-2]}\big)\mathbf{G} \oplus \mathbf{b}_{\mathbf{A}_3}. \end{aligned}$$

If the messages are not unique, an error event will be declared. This decoding process is repeated for User 2 and 3. With these decoding processes each user obtains an estimate of the messages of the other two users. By $\tilde{E}_{i,[l]}$ denote the error event in the second phase of the decoding process at User i and block l.

Next, the transmitters and the receiver generate a common list of highly likely messages for block $l-2$. In what follows, we define this list. For any triplet of the messages $(\tilde{\mathbf{w}}_1, \tilde{\mathbf{w}}_2, \tilde{\mathbf{w}}_3)$ let

$$\tilde{\mathbf{x}}_{i,[l-2]}(\tilde{\mathbf{w}}_i) \triangleq \mathbf{x}_i\big(\mathbf{u}_{[l-2]}, \mathbf{t}_i(\tilde{\mathbf{w}}_i), \mathbf{v}_{i,[l-2]}, \tilde{\mathbf{w}}_i\big)$$

where $\mathbf{u}_{[l-2]}$ and $\mathbf{v}_{i,[l-2]}, i = 1, 2, 3$, are known from previous block. For shorthand denote

$$\begin{aligned} \underline{\tilde{\mathbf{x}}}_{[l-2]}(\underline{\tilde{\mathbf{w}}}) &\triangleq \big(\tilde{\mathbf{x}}_{i,[l-2]}(\tilde{\mathbf{w}}_i)\big)_{i=1,2,3},\\ \underline{\tilde{\mathbf{t}}}(\underline{\tilde{\mathbf{w}}}) &\triangleq \big(\mathbf{t}_i(\tilde{\mathbf{w}}_i)\big)_{i=1,2,3}. \end{aligned}$$

Next, given the channel output $Y_{[l-2]}$, define the list of highly likely messages corresponding to block $l-2$ as

$$\begin{aligned} \mathcal{L}[l-2] \triangleq \Big\{\underline{\tilde{\mathbf{w}}} \in A_\epsilon^{(n)}(W_1, W_2, W_3) : \big(Y_{[l-2]}, \mathbf{u}_{[l-2]}, \underline{\tilde{\mathbf{x}}}_{[l-2]}(\underline{\tilde{\mathbf{w}}}),\\ \underline{\tilde{\mathbf{t}}}(\underline{\tilde{\mathbf{w}}})\big) \in A_\epsilon^{(n)}(\tilde{Y}, \tilde{U}, \underline{\tilde{X}}, \underline{\tilde{T}})\Big\} \end{aligned} \tag{34}$$

where $\underline{\tilde{\mathbf{w}}} \triangleq (\tilde{\mathbf{w}}_1, \tilde{\mathbf{w}}_2, \tilde{\mathbf{w}}_3)$, $\underline{\tilde{X}} \triangleq (\tilde{X}_1, \tilde{X}_2, \tilde{X}_3)$ and $\underline{\tilde{T}} \triangleq (\tilde{T}_1, \tilde{T}_2, \tilde{T}_3)$. Note that the set $\mathcal{L}[l-2]$ represents the uncertainty of the receiver about the transmitted messages at block $l-2$. This list can be calculated at the transmitters as well as the receiver. Set $M_{0,[l]} = |\mathcal{L}[l-2]|$ as the size of codebook 1. Index all members of $\mathcal{L}[l-2]$ by $m \in [1, M_{0,[l]}]$.

Suppose the decoding processes in the transmitters are successful, which means the messages are estimated correctly. Suppose $\hat{\mathbf{w}}_{2,[l-2]}, \hat{\mathbf{w}}_{3,[l-2]}$ are the estimated messages at User 1. If $(\mathbf{w}_{1,[l-2]}, \hat{\mathbf{w}}_{2,[l-2]}, \hat{\mathbf{w}}_{3,[l-2]}) \in \mathcal{L}[l-2]$, then the first encoder finds its index (say m_1) in $\mathcal{L}[l-2]$. Similarly, User 2 and 3 find the index of their estimated messages (say m_2 and m_3). Since the decoding processes are assumed to be

$$a_{1,1}\mathbf{w}_{1,[l-2]} \oplus a_{2,1}\hat{\mathbf{w}}_{2,[l-2]} \oplus a_{3,1}\hat{\mathbf{w}}_{3,[l-2]} = \hat{\mathbf{w}}_{\mathbf{A}_1,[l-1]}, \tag{33a}$$

$$\begin{aligned} \Big(\hat{\mathbf{t}}_{2,[l-2]}, \hat{\mathbf{x}}_{2,[l-2]}, \hat{\mathbf{t}}_{3,[l-2]}, \hat{\mathbf{x}}_{3,[l-2]}, \hat{\mathbf{v}}_{2,[l-1]}, \hat{\mathbf{v}}_{3,[l-1]}, \mathbf{Y}_{[l-2]}, \mathbf{Y}_{[l-1]}\Big) \in A_\epsilon^{(n)}\Big(\tilde{T}_2\tilde{X}_2\tilde{T}_3\tilde{X}_3V_2V_3\tilde{Y}Y|\mathbf{s}_{1,[l-2]}, \mathbf{s}_{1,[l-1]}, \mathbf{v}_{2,[l-2]},\\ \mathbf{v}_{3,[l-2]}, \mathbf{u}_{[l-1]}, \mathbf{u}_{[l-2]}\Big) \end{aligned} \tag{33b}$$

TABLE III
THE DECODING AND ENCODING PROCESSES FOR USER I IN BLOCK l

	block: l
Decoding 1	$\hat{\mathbf{w}}_{\mathbf{A}_i,[l-1]} = a_{1i}\mathbf{w}_{1,[l-1]} \oplus a_{2i}\mathbf{w}_{2,[l-1]} \oplus a_{3i}\mathbf{w}_{3,[l-1]}$ $\mathbf{v}_{i,[l]} = \mathbf{t}_{\mathbf{A}_i,[l-1]}$
Decoding 2	$(\hat{\mathbf{w}}_{j,[l-2]}, \hat{\mathbf{w}}_{k,[l-2]})$
Encoding 1	$\mathcal{L}_{[l-2]}, \mathbf{u}_{[l]}$
Encoding 2	$\mathbf{w}_{i,[l]}, \mathbf{t}_{i,[l]}(\mathbf{w}_{i,[l]})$
Encoding 3	$\mathbf{x}_i(\mathbf{u}_{[l]}, \mathbf{t}_{i,[l]}, \mathbf{v}_{i,[l]}, \mathbf{w}_{i,[l]})$

successful, these indices are equal, i.e., $m_1 = m_2 = m_3 = m$. Therefore, the transmitters can calculate the corresponding codeword in codebook 1, i.e., $\mathbf{u}_{[l]}(m)$. Note that the receiver is not able to find m. This is because each transmitter knows its own message and has less uncertainty comparing to the receiver. The objective of Codebook 1 is to resolve the uncertainty at the decoder.

The next step is the encoding process for block l which is similar to the previous blocks. Given a new message $\mathbf{w}_{i,[2]}, i = 1, 2, 3$, at User i, the sequence $\mathbf{t}_{i,[l]}$ is calculated and the codeword $\mathbf{x}_i(\mathbf{u}_{[l]}, \mathbf{t}_{i,[l]}, \mathbf{v}_{i,[l]}, \mathbf{w}_{i,[l]})$ is sent to the channel. For shorthand, the transmitted codeword is denoted by $\mathbf{x}_{i,[l]}$. The encoding and decoding processes in this block are shown in Table II and III.

Decoding at block l: The decoder knows the list of highly likely messages. This list is $\mathcal{L}[l-2]$ as defined in (34). Given $Y_{[l]}$ the decoder wishes to decode $U_{[l]}$ using which it can find the transmitted messages at block $l-2$. This decoding process is performed by finding an index $m \in [1 : M_{0,[l]}]$ such that

$$(U_{[l,m]}, Y_{[l]}) \in A_\epsilon^{(n)}(U, Y).$$

If the index is not found or is not unique, then an error event $E_{d,[l]}$ is declared.

C. Error Analysis

There are three types of decoding errors:

1) Error in decoding the linear combination of the messages, i.e., $E_{i,[l]}, i = 1, 2, 3, l \geq 2$.
2) Error in the decoding of the messages of the other encoders, i.e., $\tilde{E}_{i,[l]}, i = 1, 2, 3, l \geq 3$.
3) Error at the decoder, i.e. $E_{d,[l]}, l \geq 3$.

The total error probability is the probability of the union of above error events:

$$\begin{aligned}
P_e &= \mathbb{P}\left\{\bigcup_{l\geq 2}\Big(E_{d,[l]} \bigcup \Big[\bigcup_{i=1}^3 E_{i,[l]} \bigcup \tilde{E}_{i,[l]}\Big]\Big)\right\} \\
&\leq B\ \mathbb{P}\Big\{E_{d,[3]} \bigcup \Big[\bigcup_{i=1}^3 E_{i,[3]} \bigcup \tilde{E}_{i,[3]}\Big]\Big\} \\
&\leq B\ \mathbb{P}\Big\{\bigcup_{i=1}^3 E_{i,[3]} \bigcup \tilde{E}_{i,[3]}\Big\} \\
&\quad + B\ \mathbb{P}\Big\{E_{d,[3]} \ \Big|\ \bigcap_{i=1}^3 E_{i,[3]}^c \bigcap \tilde{E}_{i,[3]}^c\Big\} \\
&\leq B \sum_{i=1}^3 \Big[\mathbb{P}\{E_{i,[3]}\} + \mathbb{P}\{\tilde{E}_{i,[3]} \ \Big|\ E_{i,[3]}^c\}\Big] \\
&\quad + B\ \mathbb{P}\Big\{E_{d,[3]} \ \Big|\ \bigcap_{i=1}^3 E_{i,[3]}^c \bigcap \tilde{E}_{i,[3]}^c\Big\}, \qquad (35)
\end{aligned}$$

where B is the number of blocks. The first inequality holds due to the union bound on l and the fact that l does not change the probability of the error events. The second and third inequality hold because $P(A \bigcup B) \leq P(A) + P(B|A^c)$ and the union bound on i. Using standard arguments for each type of the errors we get the following bounds:

The probability of the first type of the errors ($\mathbb{P}\{E_{i,[3]}\}$) can be made arbitrary small for sufficiently large n, if for any distinct $i, j, k \in \{1, 2, 3\}$ the following bound holds:

$$\frac{k}{n} H(W_{\mathbf{A}_i}|W_i) \leq I(T_{\mathbf{A}_i}; Y|UT_iV_iX_i) - \delta_1(\epsilon). \qquad (36)$$

The argument follows by standard error analysis for decoding $\mathbf{w}_{\mathbf{A}_i}$ at User i. At User i, with probability sufficiently close to 1, $\mathbf{w}_{\mathbf{A}_i}$ satisfies (32). Hence, to analyze $E_{i,[3]}$, it suffices to find the probability that a codewrod $\hat{\mathbf{w}}_{\mathbf{A}_i} \neq \mathbf{w}_{\mathbf{A}_i}$ satisfies (32). Note that $\mathbf{w}_i$ is known at User i. Hence, there are approximately $2^{kH(W_{\mathbf{A}_i}|W_i)}$ ϵ-typical sequences $\hat{\mathbf{w}}_{\mathbf{A}_i}$. From standard arguments, one can show that the probability that each of such sequences satisfies (32) is approximately equals to 2^{-nI}, where I is the mutual information on the right-hand side of (36). Therefore, the error probability $\mathbb{P}\{E_{1,[3]}\}$ approaches zero, if (36) is satisfied.

The probability of the second type of the errors given by $\mathbb{P}\{\tilde{E}_{i,[3]}|E_{i,[3]}^c\}$ approaches zero for sufficiently large n, if

$$\begin{aligned}
\frac{k}{n} H(W_j, W_k|W_i, W_{\mathbf{A}_i}) \leq I(\tilde{T}_j\tilde{X}_j\tilde{T}_k\tilde{X}_k; Y\tilde{Y}|\tilde{U}\tilde{S}_iUS_i\tilde{V}_j\tilde{V}_k) \\
- \delta_2(\epsilon). \qquad (37)
\end{aligned}$$

For this type of error it is assumed that the linear combination $\mathbf{w}_{\mathbf{A}_i}$ is decoded correctly. Hence, one needs to find the probability that (33) is satisfied for a pair $(\hat{\mathbf{w}}_j, \hat{\mathbf{w}}_k) \neq (\mathbf{w}_j, \mathbf{w}_k)$. There are approximately $2^{kH(W_j,W_k|W_i,W_{\mathbf{A}_i})}$ such jointly typical pairs satisfying (33a). The probability that any of such pairs satisfies (33b) is sufficiently small for large enough n if the following inequality holds

$$\begin{aligned}
\frac{k}{n} H(W_j, W_k|W_i, W_{\mathbf{A}_i}) \leq I(\tilde{T}_j\tilde{X}_j\tilde{T}_k\tilde{X}_kV_jV_k; Y\tilde{Y}|\tilde{U}\tilde{S}_iUS_i\tilde{V}_j\tilde{V}_k) \\
- \delta_3(\epsilon).
\end{aligned}$$

The mutual information above equals to the one in (37). This is due to the fact that $V_i = \tilde{T}_{\mathbf{A}_i}$, as stated below the equation in (10).

The third type of error ($\mathbb{P}\{E_{d,[3]}| \bigcap_{i=1}^3 E_{i,[3]}^c \bigcap \tilde{E}_{i,[3]}^c\}$) approaches zero, if $|\mathcal{L}[l]| < 2^{nI(U;Y)}$. It can be shown that for sufficiently large n,

$$\mathbb{P}\Big\{|\mathcal{L}[l]| < 2^{n \max_{\mathcal{B}\subseteq\{1,2,3\}} F_{\mathcal{B}} + o(\epsilon)}\Big\} > 1 - \epsilon,$$

where for any $\mathcal{B} \subseteq \{1, 2, 3\}$

$$F_{\mathcal{B}} \triangleq \frac{k}{n} H(W_{\mathcal{B}}) - I(X_{\mathcal{B}}; Y|US_{\mathcal{B}^c}\tilde{V}_1, \tilde{V}_2, \tilde{V}_3). \qquad (38)$$

Therefore, the probability of third type of the errors approaches zero with rate $2^{-n\delta}$ for $\delta \in (0, 1)$ and sufficiently large n, if the following bounds hold for any subset $\mathcal{B} \subseteq \{1, 2, 3\}$:

$$F_{\mathcal{B}} \leq I(U; Y) - \delta - o(\epsilon),$$

Using the definition of $F_{\mathcal{B}}$ in (38), the above bounds are equivalent to the following:

$$\frac{k}{n}H(W_{\mathcal{B}}) \leq I(X_{\mathcal{B}};Y|US_{\mathcal{B}^c}\tilde{V}_1,\tilde{V}_2,\tilde{V}_3) + I(U;Y) - \delta - o(\epsilon), \quad \forall \mathcal{B} \subseteq \{1,2,3\} \tag{39}$$

Consequently, if the bounds in (36), (37), and (39) are satisfied for a fixed $\delta > 0$, then, from the inequality in (35), we obtain

$$P_e \leq 7 \times B \times 2^{-n\delta}$$

Hence, if B grows sub-exponentially as a function of n, then $P_e \to 0$ as $n \to \infty$. Note that the effective rate of our coding scheme is

$$R_i \triangleq \frac{1}{n}\log_2 \ M_i = \frac{k}{n}H(W_i), \qquad i = 1,2,3.$$

Therefore, from the bounds in (36), (37), and (39), a rate triplet (R_1, R_2, R_3) is achievable if there exist $\alpha \in (0,1)$ and random variables $W_r, T_r, V_r, X_r, \tilde{T}_r, \tilde{V}_r, \tilde{X}_r, r = 1,2,3$, distributed as described in Theorem 2, such that

$$\begin{aligned}
\alpha H(W_i) &= R_i,\\
\alpha H(W_{\mathbf{A}_i}|W_i) &\leq I(T_{\mathbf{A}_i};Y|UT_iV_iX_i),\\
\alpha H(W_j,W_k|W_{\mathbf{A}_i},W_i) &\leq I(\tilde{T}_j\tilde{X}_j\tilde{T}_k\tilde{X}_k;Y\tilde{Y}|\tilde{U}\tilde{S}_iUS_i\tilde{V}_j\tilde{V}_k),\\
\alpha H(W_{\mathcal{B}}) &\leq I(X_{\mathcal{B}};Y|US_{\mathcal{B}^c}\tilde{V}_1,\tilde{V}_2,\tilde{V}_3) + I(U;Y)
\end{aligned}$$

APPENDIX D
PROOF OF THEOREM 4

We begin the proof by the following lemma.

Lemma 6: For the channel given in Example 4, the rate triple $(1-h(\delta), 1-h(\delta), 1-h(\delta))$ is achievable.

Proof: The proof is given in Appendix E-A. ∎

Remark 4: The triple $(1-h(\delta), 1-h(\delta), 1-h(\delta))$ is a corner point in the capacity region of the channel in Example 4. This implies the optimality of the above coding strategy in terms of achievable rates.

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

One implication of Remark 4 is that the proposed coding scheme achieves optimality. We show a stronger result in this Subsection. We prove that every coding scheme that achieves $(1-h(\delta), 1-h(\delta), 1-h(\delta))$, should carry certain algebraic structures such as closeness under the binary addition.

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

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

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

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

Lemma 7: For any fixed $c > 0$, define

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

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

Proof: The proof is given in Appendix E-B. ∎

The Lemma implies that in order to achieve $(1-h(\delta), 1-h(\delta), 1-h(\delta))$, the third user needs to decode $X_{12,l} \oplus X_{22,l}$ for "almost all" $l \in [1:N]$. This requirement is necessary to insure that the channel given in Figure 9 is in the first state.

In the next step, we use the results of Lemma 7, and drive two necessary conditions for decoding $X_{12} \oplus X_{22}$.

Lemma 8: The following holds

$$\begin{aligned}
\frac{1}{N}\big|\ \log||\mathcal{C}_{12} \oplus \mathcal{C}_{22}|| - \log||\mathcal{C}_{12}||\ \big| &\leq \lambda_1(\epsilon),\\
\frac{1}{N}\big|\ \log||\mathcal{C}_{12} \oplus \mathcal{C}_{22}|| - \log||\mathcal{C}_{22}||\ \big| &\leq \lambda_2(\epsilon),
\end{aligned}$$

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

Proof: The proof is given in Appendix E-C. ∎

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

Remark 5: The three-user extension of CL scheme is suboptimal. Because, the conditions in Lemma 8 are not satisfied.

APPENDIX E
PROOF OF LEMMA 6 TO 8

A. Proof of Lemma 6

Outline of the proof: We start by proposing a coding scheme. There are L blocks of transmissions in this scheme, with new messages available at each user at the beginning of each block. The scheme sends the messages with n uses of the channel. Let $\mathbf{W}_{i,[l]}^k$ denotes the message of the ith transmitter at the lth block, where $i = 1,2,3$, and $1 \leq l \leq L$. Let $\mathbf{W}_{i,[l]}^k$ take values randomly and uniformly from $\mathbb{F}_2^k$. In this case, the transmission rate of each user is $R_i = \frac{k}{n}, i = 1,2,3$. The first and the second outputs of the ith encoder in block l is denoted by $\mathbf{X}_{i1,[l]}^n$ and $\mathbf{X}_{i2,[l]}^n$, respectively.

Codebook Construction: Select a $k \times n$ matrix $\mathbf{G}$ randomly and uniformly from $\mathbb{F}_2^{k\times n}$. This matrix is used as the generator matrix of a linear code. Each encoder is given the matrix $\mathbf{G}$. Therefore, the encoders use an identical linear code generated by $\mathbf{G}$.

Encoder 1 and 2: For the first block set $\mathbf{X}_{i2,[1]}^n = 0$, for $i = 1,2,3$. For the block l, encoder 1 sends $\mathbf{X}_{11,[l]}^n = \mathbf{W}_{1,[l]}^k\mathbf{G}$ through its first output. For the second output, encoder 1 sends $\mathbf{X}_{11,[l-1]}^n$ from block $l-1$, that is $\mathbf{X}_{12,[l]}^n = \mathbf{X}_{11,[l-1]}^n$. Similarly, the outputs of the second encoder are $\mathbf{X}_{21,[l]}^n = \mathbf{W}_{2,[l]}^k\mathbf{G}$, and $\mathbf{X}_{22,[l]}^n = \mathbf{X}_{21,[l-1]}^n$.

Encoder 3: The third encoder sends $\mathbf{X}^n_{31,[l]} = \mathbf{W}^k_{3,[l]}\mathbf{G}$ though its first output. This encoder receives the feedback from the block $l-1$ of the channel. This encoder wishes to decode $\mathbf{W}^k_{1,[l-1]} \oplus \mathbf{W}^k_{2,[l-1]}$ using $\mathbf{Y}^n_{1,[l-1]}$. For this purpose, it subtracts $\mathbf{X}^n_{31,[l-1]}$ from $\mathbf{Y}^n_{1,[l-1]}$. Denote the resulting vector by $\mathbf{Z}^n$. Then, it finds a unique vector $\tilde{\mathbf{w}}^k \in \mathbb{F}_2^k$ such that $(\tilde{\mathbf{w}}^k\mathbf{G}, \mathbf{Z}^n)$ is ϵ-typical with respect to P_{XZ}, where X is uniform over $\mathbb{F}_2$, and $Z = X \oplus \tilde{N}_\delta$. If the decoding process is successful, the third encoder sends $\mathbf{X}^n_{32,[l]} = \tilde{\mathbf{w}}^k_{[l-1]}\mathbf{G}$. Otherwise, an event $E_{1,[l]}$ is declared.

Decoder: The decoder receives the outputs of the channel from the lth block, that is $\mathbf{Y}^n_{1,[l]}$ and $\mathbf{Y}^n_{2,[l]}$. The decoding is performed in three steps. First, the decoder uses $\mathbf{Y}^n_{2,[l]}$ to decode $\mathbf{W}^k_{1,[l-1]}$, and $\mathbf{W}^k_{2,[l-1]}$. In particular, it finds unique $\tilde{\mathbf{w}}^k_1, \tilde{\mathbf{w}}^k_2 \in \mathbb{F}_2^k$ such that $(\tilde{\mathbf{w}}^k_1\mathbf{G}, \tilde{\mathbf{w}}^k_2\mathbf{G}, \mathbf{Y}^n_{2,[l]})$ are jointly ϵ-typical with respect to $P_{X_{12}X_{22}Y_2}$. Otherwise, an error event $E_{2,[l]}$ will be declared.

Suppose the first part of the decoding process is successful. At the second step, the decoder calculates $\mathbf{X}^n_{11,[l-1]}$, and $\mathbf{X}^n_{21,[l-1]}$. This is possible, because $\mathbf{X}^n_{11,[l-1]}$, and $\mathbf{X}^n_{21,[l-1]}$ are functions of the messages. The decoder then subtracts $\mathbf{X}^n_{11,[l-1]} \oplus \mathbf{X}^n_{21,[l-1]}$ from $\mathbf{Y}^n_{1,[l-1]}$. The resulting vector is

$$\tilde{\mathbf{Y}}^n = \mathbf{X}^n_{31,[l-1]} \oplus \tilde{N}^n_\delta.$$

In this situation, the channel from X_{31} to $\tilde{Y}$ is a binary additive channel with δ as the bias of the noise. At the third step, the decoder uses $\tilde{\mathbf{Y}}^n$ to decode the message of the third user, i.e., $\mathbf{W}^k_{3,[l-1]}$. In particular, the decoder finds unique $\tilde{\mathbf{w}}^k_3 \in \mathbb{F}_2^k$ such that $(\tilde{\mathbf{w}}^k_3\mathbf{G}, \tilde{\mathbf{Y}}^n)$ are jointly ϵ-typical with respect to $P_{X_{31}\tilde{Y}}$. Otherwise, an error event $E_{3,[l]}$ is declared.

Error Analysis: We can show that this problem is equivalent to a point-to-point channel coding problem, where the channel is described by $Z = X \oplus \tilde{N}_\delta$. The average probability of error approaches zero, if $\frac{k}{n} \leq 1 - h_b(\delta)$.

Suppose there is no error in the decoding process of the third user. That is $E^c_{1,[l]}$ occurs. Therefore, with probability one,

$$\mathbf{X}^n_{32,[l]} = \mathbf{X}^n_{22,[l]} \oplus \mathbf{X}^n_{12,[l]}.$$

As a result, the channel in Fig. 9 is in the first state. This implies that the corresponding channel consists of two parallel binary additive channel with independent noises and bias δ. Similar to the argument for E_1, it can be shown that $P(E_{2,[l]}|E_{1,[l]}) \to 0$, if $\frac{k}{n} \leq 1 - h_b(\delta)$. Lastly, we can show that conditioned on $E^c_{1,[l]}$ and $E^c_{2,[l]}$, the probability of $E_{3,[l]}$ approaches zero, if $\frac{k}{n} \leq 1 - h_b(\delta)$.

As a result of the above argument, the average probability of error approaches 0, if $\frac{k}{n} \leq 1 - h_b(\delta)$. This implies that the rates $R_i = 1 - h_b(\delta), i = 1, 2, 3$ are achievable, and the proof is completed. ∎

B. Proof of Lemma 7

Proof: Let R_i be the rate of the ith encoder. We have $R_i \geq 1 - h_b(\delta) - \epsilon$. We apply the generalized Fano's inequality (Lemma 4.3 in [17]) for decoding of the messages. More precisely, as $\bar{P} \leq \epsilon$, we have

$$\frac{1}{\Theta_1\Theta_2\Theta_3} H(M_1, M_2, M_3|\mathbf{Y}^N) \leq h(\bar{P}) \leq h(\epsilon)$$

By the definition of the rate we have

$$\begin{aligned}
R_1 + R_2 + R_3 &= \frac{1}{N} H(M_1, M_2, M_3) \\
&\leq \frac{1}{N} I(M_1, M_2, M_3; \mathbf{Y}^N) + o(\epsilon) \\
&= \frac{1}{N} H(\mathbf{Y}^N) - \frac{1}{N}\sum_{l=1}^{N} H(Y_l|\mathbf{Y}^{l-1}, M_1, M_2, M_3) + o(\epsilon) \\
&\overset{(a)}{=} \frac{1}{N} H(\mathbf{Y}^N) \\
&\quad - \frac{1}{N}\sum_{l=1}^{N} H(Y_l|\mathbf{Y}^{l-1}, \mathbf{X}^l_1, \mathbf{X}^l_2, \mathbf{X}^l_3, M_1, M_2, M_3) + o(\epsilon) \\
&\overset{(b)}{=} \frac{1}{N} H(\mathbf{Y}^N) - \frac{1}{N}\sum_{l=1}^{N} H(Y_l|\mathbf{Y}^{l-1}, \mathbf{X}^l_1, \mathbf{X}^l_2, \mathbf{X}^l_3) + o(\epsilon) \\
&\overset{(c)}{=} \frac{1}{N} H(\mathbf{Y}^N) - \frac{1}{N}\sum_{l=1}^{N} H(Y_l|X_{1,l}, X_{2,l}, X_{3,l}) + o(\epsilon) \\
&\overset{(d)}{\leq} 3 - \frac{1}{N}\sum_{l=1}^{N} H(Y_l|X_{1,l}, X_{2,l}, X_{3,l}) + o(\epsilon), \qquad (40)
\end{aligned}$$

where (a) holds as $\mathbf{X}^l_j$ is a function of $(Y^{l-1}, M_j), j = 1, 2, 3$, equality (b) is due to the fact that conditioned on $(\mathbf{Y}^{l-1}, \mathbf{X}^l_1, \mathbf{X}^l_2, \mathbf{X}^l_3)$ the random variable Y_l is independent of (M_1, M_2, M_3), equality (c) is because of (7), and lastly, inequality (d) holds as Y is a vector of three binary random variables which implies that

$$\frac{1}{N} H(Y^N) \leq 3.$$

Let $P(X_{32,l} \neq X_{12,l} \oplus X_{12,l}) = q_l$, for $l \in [1 : N]$. Denote $\bar{q}_l = 1 - q_l$. We can show that,

$$H(Y_l|X_{1,l}, X_{2,l}, X_{3,l}) = (1 + 2\bar{q}_l)h_b(\delta) + 2q_l.$$

We use the above argument, and the last inequality in (40) to give the following bound

$$\begin{aligned}
R_1 + R_2 + R_3 &\leq 3 - \frac{1}{N}\sum_{l=1}^{N}[(1 + 2\bar{q}_l)h_b(\delta) + 2q_l] + o(\epsilon) \\
&= 3 - 3\ h_b(\delta) + \frac{1}{N} 2(1 - h_b(\delta)) \sum_{l=1}^{N} q_l + o(\epsilon)
\end{aligned}$$

By assumption $R_1 + R_2 + R_3 \geq 3(1 - h_b(\delta) - \epsilon)$. Therefore, using the above bound we obtain,

$$\frac{3\epsilon + o(\epsilon)}{2(1 - h_b(\delta))} \geq \frac{1}{N}\sum_{l=1}^{N} q_l \overset{(a)}{\geq} \frac{1}{N}\sum_{l \in \mathcal{I}_c^N} q_l,$$

where (a) holds, because we remove the summation over all $l \notin \mathcal{I}_c^N$. We defined $\mathcal{I}_c^N$ as in the statement of this Lemma. Note that if $l \in \mathcal{I}_c^N$, then $q_l \geq c$. Finally, we obtain

$$\frac{|\mathcal{I}_c^N|}{N} \leq \frac{3\epsilon + o(\epsilon)}{2\ c(1 - h_b(\delta))}$$

∎

C. Proof of Lemma 8

Proof: Let $\mathcal{I}_c^N$ be as in Lemma 7. The average probability of error for decoding $X_{12}^N \oplus X_{22}^N$ is bounded as

$$\begin{aligned}\bar{P}_e &= \frac{1}{N}\sum_{l=1}^{N} P(X_{32,l} \neq X_{12,l} \oplus X_{22,l}) \\ &= \frac{1}{N}\sum_{l\in\mathcal{I}_c^N} P(X_{32,l} \neq X_{12,l} \oplus X_{22,l}) \\ &\quad + \frac{1}{N}\sum_{l\notin\mathcal{L}_c^N} P(X_{32,l} \neq X_{12,l} \oplus X_{22,l}) \\ &\leq \frac{|\mathcal{I}_c^N|}{N} + c(1-\frac{|\mathcal{I}_c^N|}{N}) \\ &= (1-c)\frac{|\mathcal{I}_c^N|}{N} + c \\ &\leq (1-c)\frac{\eta(\epsilon)}{2c(1-h(\delta))} + c\end{aligned}$$

As a result as $\epsilon \to 0$, then $\bar{P}_e \to c$. Since $c > 0$ is arbitrary, $\bar{P}_e$ can be made arbitrary small. Hence, for any $\epsilon' > 0$, and there exist $\epsilon > 0$ and large enough N such that $\bar{P}_e < \epsilon'$. Note that X_{32}^N is a function of M_3, Y_1^N, Y_{12}^N and Y_{22}^N. Next we argue that to get $\bar{P}_e < \epsilon'$, it is enough for X_{32}^N to be a function of M_3, Y_1^N. More precisely, given $X_{32,l}$, the random variables $Y_{12,l}$ and $Y_{22,l}$ are independent of $X_{12,l} \oplus X_{22,l}$. To see this, we need to consider two cases. If $X_{32,l} = X_{12,l} \oplus X_{22,l}$ then the argument follows trivially. Otherwise,

$$Y_{12,l} = X_{12,l} \oplus N_{1/2},$$

where $N_{1/2} \sim Ber(1/2)$, and it is independent of $X_{12,l}$. Hence in this case, $Y_{12,l}$ is independent of $X_{12,l}$. Similarly, $Y_{22,l}$ is independent of $X_{22,l}$. By subtracting X_{31}^N from Y_1^N, we get

$$Z^N \triangleq X_{11}^N \oplus X_{21}^N \oplus N_\delta^N.$$

Next, we argue that the third encoder uses Z^N to decode $X_{12}^N \oplus X_{22}^N$. Since M_3 is independent of M_1 and M_2, it is independent of X_{1j}^N, X_{j2}^N for $j = 1, 2$. Therefore Z^N is independent of M_3. Hence, X_{32}^N is function of Z^N. Intuitively, we convert the problem of decoding $X_{11}^N \oplus X_{21}^N$ to a point to point channel coding problem. The channel in this case is a binary additive channel with noise $N_\delta \sim Ber(\delta)$. In this channel coding problem the codebook at the encoder is $\mathcal{C}_{12} \oplus \mathcal{C}_{22}$. The capacity of this channel equals $1 - h_b(\delta)$. Since the average probability of error is small, we can use the generalized Fano's inequality to bound the rate of the encoder. As a result, it can be shown that

$$\frac{1}{N}\log_2 ||\mathcal{C}_{12} \oplus \mathcal{C}_{22}|| \leq 1 - h_b(\delta) + \eta(\epsilon), \tag{41}$$

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

Lemma 9: The following bound holds

$$\frac{1}{N}\log_2 ||\mathcal{C}_{j2}|| \geq 1 - h_b(\delta) - \gamma_j(\epsilon), \tag{42}$$

where $j = 1, 2$ and $\gamma_j(\epsilon) \to 0$ as $\epsilon \to 0$.

Outline of the proof: First, we show that the decoder must decode M_3 from Y_1^N. We argued in the above that X_{32}^N is independent of M_3. Hence, the message M_3 is encoded only to X_{31}^N. Since X_{31}^N is sent though the first channel in Example 1, the decoder must decode M_3 from Y_1^N. Next, we argue that the receiver must decode M_1 and M_2 from Y_{21}^N and Y_{22}^N, respectively. Note that the rate of the third encoder is $1-h_b(\delta)$, which equals to the capacity of the first channel given $X_{11}^N \oplus X_{21}^N$. Therefore, the decoder can decode M_3, if it has $X_{11}^N \oplus X_{21}^N$. Hence, the decoder must reconstruct $X_{11}^N \oplus X_{21}^N$ from the second channel. It can be shown that this is possible, if the decoder can decode M_1 and M_2 from the second channel. As a result, from Fano's inequality, the bounds in the Claim hold. ∎

Finally, using (41) and (42) we get

$$0 \leq \frac{1}{N}\log_2 ||\mathcal{C}_{12} \oplus \mathcal{C}_{22}|| - \frac{1}{N}\log_2 ||\mathcal{C}_{j2}|| \leq \eta(\epsilon) + \gamma_j(\epsilon),$$

where $j = 1, 2$. This completes the proof. ∎

ACKNOWLEDGMENT

The authors would like to thank F. Shirani of New York University for extensive and insightful discussions related to this work. They would also like to thank anonymous reviewers and the associate editor for their careful reading of the article and their comments which helped them in preparing a better article.

REFERENCES

[1] P. Gacs and J. Krner, "Common information is far less than mutual information," *Problems Control Inf. Theory*, vol. 2, Jan. 1973.
[2] H. S. Witsenhausen, "On sequences of pairs of dependent random variables," *SIAM J. Appl. Math.*, vol. 28, no. 1, pp. 100–113, Jan. 1975.
[3] A. Wyner, "The common information of two dependent random variables," *IEEE Trans. Inf. Theory*, vol. 21, no. 2, pp. 163–179, Mar. 1975.
[4] G. R. Kumar, C. T. Li, and A. El Gamal, "Exact common information," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2014, pp. 161–165.
[5] R. M. Gray and A. D. Wyner, "Source coding for a simple network," *Bell Syst. Tech. J.*, vol. 53, no. 9, pp. 1681–1721, Nov. 1974.
[6] A. B. Wagner, B. G. Kelly, and Y. Altug, "Distributed rate-distortion with common components," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4035–4057, Jul. 2011.
[7] T. Cover, A. E. Gamal, and M. Salehi, "Multiple access channels with arbitrarily correlated sources," *IEEE Trans. Inf. Theory*, vol. IT-26, no. 6, pp. 648–657, Nov. 1980.
[8] W. Liu and B. Chen, "Interference channels with arbitrarily correlated sources," in *Proc. 47th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Sep./Oct. 2009, pp. 585–592.
[9] C. T. Li and A. El Gamal, "Distributed simulation of continuous random variables," *IEEE Trans. Inf. Theory*, vol. 63, no. 10, pp. 6329–6343, Oct. 2017.
[10] R. Ahlswede, "Multi-way communication channels," in *Proc. 2nd Int. Symp. Inf. Theory, Tsahkadsor, Armenia, USSR*, Sep. 1973, pp. 1–28.
[11] D. Slepian and J. K. Wolf, "A coding theorem for multiple access channels with correlated sources," *Bell Syst. Tech. J.*, vol. 52, no. 7, pp. 1037–1076, Sep. 1973.
[12] G. Dueck, "A note on the multiple access channel with correlated sources (corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 2, pp. 232–235, Mar. 1981.
[13] A. Lapidoth and M. Wigger, "A necessary condition for the transmissibility of correlated sources over a MAC," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 2024–2028.
[14] A. Lapidoth, S. S. Bidokhti, and M. Wigger, "Dependence balance in multiple access channels with correlated sources," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 1663–1667.
[15] N. Gaarder and J. Wolf, "The capacity region of a multiple-access discrete memoryless channel can increase with feedback (corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-21, no. 1, pp. 100–102, Jan. 1975.
[16] T. Cover and C. Leung, "An achievable rate region for the multiple-access channel with feedback," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 3, pp. 292–298, May 1981.

[17] G. Kramer, "Directed information for channels with feedback," Ph.D. dissertation, Swiss Federal Inst. Technol., Zürich, Switzerland, 1998.
[18] R. Venkataramanan and S. S. Pradhan, "A new achievable rate region for the multiple-access channel with noiseless feedback," *IEEE Trans. Inf. Theory*, vol. 57, no. 12, pp. 8038–8054, Dec. 2011.
[19] F. Willems, "The feedback capacity region of a class of discrete memoryless multiple access channels (corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 1, pp. 93–95, Jan. 1982.
[20] F. Willems and E. van der Meulen, "Partial feedback for the discrete memoryless multiple access channel (corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 2, pp. 287–290, Mar. 1983.
[21] J. Schalkwijk and T. Kailath, "A coding scheme for additive noise channels with feedback-I: No bandwidth constraint," *IEEE Trans. Inf. Theory*, vol. IT-12, no. 2, pp. 172–182, Apr. 1966.
[22] L. Ozarow, "The capacity of the white Gaussian multiple access channel with feedback," *IEEE Trans. Inf. Theory*, vol. 30, no. 4, pp. 623–629, Jul. 1984.
[23] S. I. Bross and A. Lapidoth, "An improved achievable region for the discrete memoryless two-user multiple-access channel with noiseless feedback," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 811–833, Mar. 2005.
[24] D. Krithivasan and S. S. Pradhan, "Distributed source coding using Abelian group codes: A new achievable rate-distortion region," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1495–1519, Mar. 2011.
[25] 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. Inf. Theory*, vol. IT-29, no. 3, pp. 396–412, May 1983.
[26] T. Han and K. Kobayashi, "A unified achievable rate region for a general class of multiterminal source coding systems," *IEEE Trans. Inf. Theory*, vol. IT-26, no. 3, pp. 277–288, May 1980.
[27] T. Han and K. Kobayashi, "A dichotomy of functions F(X, Y) of correlated sources (X, Y)," *IEEE Trans. Inf. Theory*, vol. IT-33, no. 1, pp. 69–76, Jan. 1987.
[28] B. Nazer and M. Gastpar, "Computation over multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3498–3516, Oct. 2007.
[29] J. Zhan, S. Y. Park, M. Gastpar, and A. Sahai, "Linear function computation in networks: Duality and constant gap results," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 4, pp. 620–638, Apr. 2013.
[30] R. Appuswamy, M. Franceschetti, N. Karamchandani, and K. Zeger, "Linear codes, target function classes, and network computing capacity," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5741–5753, Sep. 2013.
[31] A. Padakandla and S. Sandeep Pradhan, "Computing sum of sources over an arbitrary multiple access channel," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2013, pp. pp. 2144–2148.
[32] T. Philosof, R. Zamir, U. Erez, and A. J. Khisti, "Lattice strategies for the dirty multiple access channel," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 5006–5035, Aug. 2011.
[33] T. Philosof and R. Zamir, "On the loss of single-letter characterization: The dirty multiple access channel," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2442–2454, Jun. 2009.
[34] A. Padakandla and S. S. Pradhan, "An achievable rate region based on coset codes for multiple access channel with states," *IEEE Trans. Inf. Theory*, vol. 63, no. 10, pp. 6393–6415, Oct. 2017.
[35] M. Heidari, F. Shirani, and S. S. Pradhan, "A new achievable rate region for multiple-access channel with states," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 36–40.
[36] S. Sridharan, A. Jafarian, S. Vishwanath, S. A. Jafar, and S. Shamai, "A layered lattice coding scheme for a class of three user Gaussian interference channels," in *Proc. 46th Annu. Allerton Conf. Commun., Control, Comput.*, Sep. 2008, pp. 531–538.
[37] S.-N. Hong and G. Caire, "On interference networks over finite fields," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 4902–4921, Aug. 2014.
[38] G. Bresler, A. Parekh, and D. N. C. Tse, "The approximate capacity of the many-to-one and one-to-many Gaussian interference channels," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4566–4592, Sep. 2010.
[39] U. Niesen and M. A. Maddah-Ali, "Interference alignment: From degrees of freedom to constant-gap capacity approximations," *IEEE Trans. Inf. Theory*, vol. 59, no. 8, pp. 4855–4888, Aug. 2013.
[40] A. Jafarian and S. Vishwanath, "Achievable rates for K-user Gaussian interference channels," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4367–4380, Jul. 2012.
[41] O. Ordentlich, U. Erez, and B. Nazer, "The approximate sum capacity of the symmetric Gaussian K-user interference channel," in *Proc. IEEE Int. Symp. Inf. Theory Proc.*, Jul. 2012, pp. 2072–2076.
[42] A. Padakandla and S. S. Pradhan, "Achievable rate region for three user discrete broadcast channel based on coset codes," *IEEE Trans. Inf. Theory*, vol. 64, no. 4, pp. 2267–2297, Apr. 2018.
[43] I. Csiszar and J. Korner, *Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
[44] A. E. Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
[45] J. Massey, "Causality, feedback and directed information," in *Proc. Int. Symp. Inf. Theory Appl. (ISITA)*, 1990, pp. 303–305.
[46] F. Shirani, M. Heidari, and S. S. Pradhan, "Quasi linear codes: Application to point-to-point and multi-terminal source coding," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 730–734.

Mohsen Heidari received the B.Sc. and M.Sc. degrees in electrical engineering from the Sharif University of Technology in 2011 and 2013, respectively, and the M.Sc. degree in applied mathematics and the Ph.D. degree in electrical engineering from the University of Michigan in 2017 and 2019, respectively. He is currently a Post-Doctoral Research Fellow with the Center for Science of Information, NSF Science and Technology Center, Purdue University. His research interests lie in information theory, quantum information theory, and computational learning theory.

S. Sandeep Pradhan (Senior Member, IEEE) received the M.E. degree from the Indian Institute of Science in 1996 and the Ph.D. degree from the University of California at Berkeley in 2001. From 2002 to 2008, he was an Assistant Professor, and from 2008 to 2015, he was an Associate Professor with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, where he is currently a Professor. His research interests include network information theory, coding theory, and quantum information theory. He was a recipient of the 2001 Eliahu Jury Award given by the University of California at Berkeley for outstanding research in the areas of systems, signal processing, and communications and control, the CAREER Award given by the National Science Foundation (NSF), and the Outstanding Achievement Award in 2009 from the University of Michigan. He was an Associate Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY in the area of Shannon theory from 2014 to 2016.