A Graph-based Framework for Transmission of Correlated Sources over Multiple Access Channels *

S. Sandeep Pradhan[a], Suhan Choi[a] and Kannan Ramchandran[b],
[a] {pradhanv,suhanc}@eecs.umich.edu, EECS Dept., Univ. of Michigan, Ann Arbor, MI.
[b] {kannanr}@eecs.berkeley.edu, EECS Dept., Univ. of California, Berkeley, CA.

January 19, 2006

Abstract

In this paper we consider a graph-based framework for transmission of correlated sources over multiple access channels. We show that a graph can be used as a discrete interface between the source coding and the channel coding for this multiterminal information transmission problem. We adopt a separation-based modular approach to this problem, involving a source coding module and a channel coding module. In the former module, the correlated sources are encoded distributively into correlated messages which can be associated with a graph (called *message-graph*), and these correlated messages are then encoded by using correlated codewords and are reliably transmitted over the multiple access channel in the latter module. This leads to performance gains in terms of enlarging the class of correlated sources that can be reliably transmitted over a multiple access channel. We provide the rate of growth of the exponent (as a function of the number of channel uses) of the size of the message-graphs whose edges can be reliably transmitted over a multiple access channel. A similar characterization of message-graphs that can reliably represent a pair of correlated sources is also provided.

1 Introduction

Consider a set of transmitters wishing to reliably and simultaneously communicate with a single receiver using a multiple access channel [1, 2, 3, 4, 5]. The transmitters do not communicate among themselves. Each transmitter in the set has some independent information, and together they wish to communicate their information to a joint receiver. This channel was first studied by Ahlswede in [1] and by Liao in [2], where they obtained the capacity region.

At around the same time, another multiterminal communication problem involving separate (distributed) encoding of correlated information sources was formulated, and the corresponding optimal encoding rate region was obtained by Slepian and Wolf in [6] (also see [7]). In this problem, the goal is to represent two (or more) distributed correlated sources using a pair of indexes to be transmitted to a joint receiver in order to losslessly reproduce these sources. Conventionally, this system is referred to as Slepian-Wolf source coding. The seminal result obtained in [6] says that a no-communication constraint on the encoders leads to no loss of performance, i.e., these sources can be represented using pairs of indexes whose total rate can approach the joint entropy of the sources [8] asymptotically. One can deduce from this that since the goal of the encoders is to produce a nonredundant representation of the sources, the indexes (messages) transmitted by the encoders will be asymptotically independent in an optimal system.

*This work was supported by NSF grant CCF-0448115. This work was presented in part at the IEEE International symposium on information theory (ISIT), Chicago, IL, June/July 2004.

Now that we have a characterization of the performance limits of a pair of multiterminal source coding and channel coding problems, we can naturally extend Shannon's point-to-point communication paradigm involving transmission of a source over a channel to multiple terminals in the following way. Consider a pair of correlated sources which are observed by a pair of distributed encoder terminals. The encoder terminals wish to simultaneously transmit the corresponding sources over a multiple access channel to a joint decoder. In such a case, there are generally two ways of sending these sources over the given channel. One is separate source coding and channel coding, where one first applies the Slepian-Wolf source coding to the correlated sources in order to minimize the redundancy in the messages representing the sources, thus producing nearly independent messages, and then applies the multiple access channel coding on these nearly independent messages. The other is joint source-channel coding which may reduce both delay and complexity, where the sources are directly mapped into the channel inputs. But designing such a joint-source-channel coding scheme is generally a more difficult optimization problem. In the latter case, what one can generally say is whether a given set of sources can be reliably transmitted over a given multiple access channel or not.

Although the separation approach is conveniently modular, it was recognized early on [9, 10], that this approach will not be optimal in contrast to the point-to-point case. In other words, this separation approach is in general strictly suboptimal. Understandably the latter approach became the center of attention, and was first studied by Slepian and Wolf in [9], where they considered a special class of such problems where there are two transmitter terminals sharing three independent information sources. The first transmitter has access to the first and the second source, and the second transmitter has access to the second and the third source. In other words, the sources of information accessed by these terminals have a so-called "common part" [11, 12, 13]. In [9], a characterization of the set of such information sources that can be transmitted reliably over a given multiple access channel was given with direct and converse parts. In 1980, a joint source-channel coding theorem to this problem was given in [10] for a more general case, where a single-letter characterization of the set of sources that can be reliably transmitted over a given multiple access channel was obtained. Now, n (say) samples of the sources are distributively mapped into n samples of the channel inputs, and the joint decoder simultaneously recovers n samples of the correlated sources by observing n samples of the channel output. This includes the results of [1], [2], [6] and [9] as special cases. The authors in [10] also provided an interesting example (to be illustrated here in the next section) that shows that separate source and channel coding [14] is not optimal for multiple access channels with correlated sources. Later, Dueck [15] showed, by an example that the approach of [10] gives only sufficient conditions for the transmissibility of correlated sources over a given multiple access channel, but not necessary conditions. Further work related to joint source-channel coding in this multiterminal setting can be found in [16, 17, 18, 19]. In summary, the performance of joint-source-channel coding is strictly superior to that of separate source coding and channel coding in this setting.

To better understand why the separation approach is not optimal in this multiterminal setting, let us revisit the point-to-point case and see fundamentally why it works here. The essence of Shannon's separation approach in the point-to-point case is an efficient architecture for transmission problems through a discrete interface (a finite set) for representing information sources. Many sources can be mapped into indexes in this finite set, and these indexes can be communicated over many channels. The fundamental concept which facilitates this is the notion of typicality. It should be noted that although there are uncountably many finite alphabet sources whose entropy [8] is less than or equal to some positive number, say H, when grouped into sufficiently large blocks (say blocklength n), they exhibit a certain determinism (from the law of large numbers). So for a given source X, most of the time, only those sequences, called typical, that come from a set of size nearly $2^{nH(X)}$ are observed, and the probability of observing any sequence from this set is nearly the same, where $H(X)$ denotes the entropy of X [8]. For all these sequences, the empirical histogram is close to the probability distribution of X. So loosely speaking, all the details of the probability distribution of a source can be dispensed with, and one needs to worry only about the *cardinality* of this sequence set.

Using this observation, one can see that the straightforward extension of this approach to the multiterminal case uses a product of finite sets as a discrete interface between Slepian-Wolf source coding and multiple access channel coding with independent messages. As mentioned above, although this representation of correlated sources is efficient in terms of the total rate of the index pairs coming from the product of finite sets being equal to the joint entropy of the correlated sources, this interface falls short of the expectation in terms of achieving optimal performance for the task of transmission of these sources over multiple access channels. This is mainly because, in the process of producing a non-redundant representation of the sources by the encoders in the source coding module, all the correlation in the sources is destroyed. If the encoders could produce an efficient representation of the sources which still retains some of the correlation of the sources, then it could be potentially used by the channel coding module that follows to combat channel interference and noise in a more effective way, as the final goal of the system is to transmit the sources over the channel. Hence the question that we would like to ask is whether it is possible to obtain a structured and efficient representation of the sources that preserves a judiciously prescribed amount of source correlation in it so that the end-to-end performance is not compromised, and thus retain in the overall system the modularity, i.e., source coding and channel coding modules which are the hallmark of the separation theorem in the point-to-point case? In short, do there exist discrete objects other than products of finite sets which can be used as efficient representations of correlated sources?

A key insight into this problem may lie in the following observation. Of course, the notion of typicality can be extended to two sources, say X and Y, which says that only those sequence-pairs (called jointly typical) that come from a set of size nearly $2^{nH(X,Y)}$ will be observed most of the time, where $H(X,Y)$ denotes the joint entropy [8]. Although, there are roughly $2^{n(H(X)+H(Y))}$ sequence pairs which are individually typical, not all of them are jointly typical because the joint entropy is in general smaller than the sum of the individual entropies. Further, using these ideas, it can be shown that for every typical sequence of X (respectively Y) , there exist roughly $2^{nH(Y|X)}$ (respectively $2^{nH(X|Y)}$) typical Y (respectively X) sequences that are jointly typical, where $H(Y|X)$ is the conditional entropy [8] of Y given X. This leads us naturally to consider a bipartite undirected graph on the sets of individually typical sequences induced from the property of joint typicality. That is, the vertexes of this graph denote the individually typical sequences, and the jointly typical sequences are connected through an edge. We refer to this graph as the typicality-graph of two correlated sources. Loosely speaking, for large blocks, the sources exhibit a certain determinism, where all the sequence pairs that really matter can be associated with a bipartite graph. These graphs capture all the correlation structure of the sources. In summary, there are roughly $2^{nH(X,Y)}$ edges in this graph, and the probability of observing any edge in this graph is roughly the same, and hence this is an equally efficient representation of the sources. Note that the bipartite graphs which have the special structure–the number of edges connected to every vertex of a set is the same–are referred to as semi-regular graphs [20, 21]. Now if we associate an index with each individually typical sequence, then joint typicality induces a bipartite graph on this pair of index sets. If we consider a separate message for each vertex in the graph, this implies that only certain pairs of messages can occur (denoted by edges in the graph) most of the time. We refer to such correlated message sets as message-graphs. Thus, we have obtained a nearly semi-regular graph, induced from the typicality graph, as an efficient representation of a pair of correlated sources.

Inductively, the Slepian-Wolf source coding result can also be interpreted in this framework: since the messages produced by an optimum Slepian-Wolf source coder are nearly independent and the total rate of the message sets is nearly equal to $H(X,Y)$, the message sets thus produced can be thought of as a nearly fully connected semi-regular message-graph. This leads to the following question. Are there other nearly semi-regular graphs that are efficient representations of these sources?

The typicality graph can be thought of as being situated at one end of the spectrum. At the other end of the spectrum is the nearly fully connected graph associated with the product of index sets used in the Slepian-Wolf coding of these correlated sources. A slew of graphs which lie in between, and that are efficient representations of these sources can be obtained if we leave some redundancy in the Slepian-Wolf source coding. As the residual redundancy goes from the minimum to the maximum level, we will approach the representation involving the typicality graph from the representation involving a fully connected graph obtained in the optimum Slepian-Wolf source coding.

This leads to the possibility of a semi-regular graph being used as a discrete interface for the multiterminal information transmission problem. We have seen how source encoders can represent the pair of sources using nearly semi-regular graphs. Three examples are illustrated in Figure 1, each with two users having three messages. As the number of edges in the graph reduces, the correlation increases. The first message-graph depicts completely independent messages and the third message-graph depicts completely correlated messages. All the edges in a given graph are assumed to be equally likely.

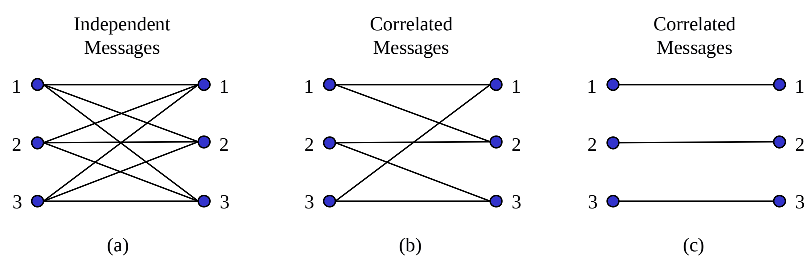

Figure 1: "Independent" and "correlated" messages: As the number of edges in the graph decreases, the correlation between the two message sets increases.

Now for the channel coding component, the encoders have to work with these correlated message sets and have to reliably transmit over the channel the edges of the message-graph produced by the source encoders. The channel encoders will now be operating on correlated messages (with the correlation structure of the graph) rather than independent ones. Thus, for a multiple access channel, at one end of the spectrum we have independent message sets (standard multiple access channel) as in Figure 1(a), and at the other end we have perfectly correlated messages, i.e., both users having the same information to send as in Figure 1(c), and in the middle there is a slew of nearly semi-regular graphs, whose edges can be reliably communicated over the channel.

Hence the act of encoding these sources into channel inputs can be divided into two operations, where the two sources are first mapped into an appropriate bipartite message-graph (in the source coding component), and the edges of this graph are reliably communicated to the receiver (in the channel coding component). The correlation of the information sources is retained by these message-graphs, and this can be directly translated to the codewords transmitted over the channel. This correlation in the channel inputs can now be exploited to combat interference and degradation introduced by the channel. Without "smearing" these two components into a joint source-channel coding block, in the proposed approach, we enhance them to work with correlated messages or graphs, thus retaining the Shannon-style modular approach to this multiuser communication system.

To see a concise summary of the results presented in this paper toward the above mentioned goal, let is consider the following definition. A nearly semi-regular bipartite graph is said to have parameters $(\theta_1, \theta_2, \theta_1', \theta_2')$, if the number of vertexes of the ith set is close to θ_i for $i = 1, 2$, and the degree of every vertex in first set is close to θ_2', and vice versa. In this paper, we first address the channel coding part, and then discuss the source coding part in the later section. In the channel coding part, we would like to know the rates of growth of the exponents, as functions of the number of uses of the channel, of the sizes and the associated degrees of the vertexes of all nearly semi-regular graphs whose edges can be reliably transmitted over a multiple access channel. In other words, our goal is to find the set of quadruples (R_1, R_2, R_1', R_2'), called the achievable rate region such that edges coming from *every* nearly semi-regular graph with parameters $(2^{nR_1}, 2^{nR_2}, 2^{nR_1}, 2^{nR_2'})$ can be reliably transmitted over a multiple access channel by using the channel n times. Similarly, in the source coding part, the goal is to find the set of quadruples (R_1, R_2, R_1', R_2') such that *every* nearly semi-regular graph with parameters $(2^{nR_1}, 2^{nR_2}, 2^{nR_1}, 2^{nR_2'})$ can be used to represent efficiently n realizations of the correlated sources. If we are successful in this task, then a pair of correlated sources can be reliably transmitted over a multiple access channel if there is a non-empty intersection of their achievable rate regions.

To see why we have emphasized the word 'every' in the above paragraph, consider the set of all graphs with a fixed quadruple of parameters $(\theta_1, \theta_2, \theta_1', \theta_2')$. Of course, there is more than one graph in this set, and the structure of the graphs in this collection could be disparate. Although we will revisit this issue formally in Section 3, at this point, it suffices to mention that these graphs can be partitioned into equivalence classes, where the graphs in an equivalence class have the same structure. This essentially means that a single codebook can be designed that works well for all graphs that belong to an equivalence class. Thus, we need one codebook for each equivalence class of graphs with a given set of parameters in both the source coding and the channel coding components.

In this paper we provide partial answers to the questions raised above using single letter information quantities. In particular, we relax the definition of the achievable rate region by replacing the word 'every' to 'at least one' in the above definition. Our main results are stated in Theorem 1, 2 and 3. As an example, we consider an achievable rate region for the Gaussian multiple access channel with jointly Gaussian channel input. We also compare our new coding scheme with the separate source and channel coding scheme which involves conventional multiple access channel coding preceded by Slepian-Wolf source coding ([6], [7]) of correlated messages. As expected, the result says that we can send the same amount of information over the multiple access channel with less power by adopting correlated codewords. Further, it is shown that the coding scheme of Cover, El Gamal, and Salehi [10] can be interpreted as a match between the typicality-graph of the pair of sources and a subgraph of the typicality-graph of some channel input distribution.

The work of Slepian and Wolf [9] is along this direction, where, as mentioned above, they considered two correlated messages with a common part [12] as inputs to the two encoders to be transmitted over a multiple access channel. However, it was shown by Gács and Körner [11] and Witsenhausen [12] that the common part of two dependent random variables is zero in most cases. Rather, in this work, we consider a more general class of correlated messages where they need not have a common part. Ahlswede and Han considered in [22] a related approach to the source-channel matching problem in multiuser communication. In [22], the authors considered the problem of representing correlated sources using bipartite graphs, and transmitting the edges of these graphs over multiple access channels without putting any structure on the graphs in terms of the distribution of the degrees of the vertexes. In contrast, in the present work, inspired from the asymptotic equipartition property, we deal with nearly semi-regular graphs, where the degrees of the vertexes are asymptotically the same. By restricting our attention to this set of "symmetric" objects, we are able to provide more concrete statements on the size of such graphs and the degrees of every vertex in those graphs such as those provided in Theorem 1, 2 and 3.

Before closing this discussion, we note that we are nowhere near achieving the ambitious goal that we began to march with. But the set of results given in this paper is possibly the first step that one needs to take to move toward this goal. For a skeptic who may not subscribe to this vision of connecting these two information-theoretic results toward a separation principle for transmission of correlated sources over a multiple access channel, we still believe that these two results would be of independent interest even when viewed separately. In other words, as an analogy, the conventional Shannon's channel coding theorem can be interpreted as finding the maximum number of codewords (colors, if each codeword has a different color) that are distinguishable at the noisy channel output. In conventional multiple access channels, the goal is to distinguish among pairs of colors at the noisy channel output, where the first color can come from one set and the second color can come from another set, and all possible combination of pairs in the two sets are allowed. A natural question to ask is: if only a fraction of all possible combination of pairs of colors is permitted, what is the maximum size of the sets of these colors for which reliable distinguishability can be guaranteed at the receiver. A similar question can be asked for the source coding problem.

The outline of the remaining part of this paper is as follows. In Section 2, we provide a brief review of the capacity of multiple access channels with independent messages and the performance limits of the Slepian-Wolf source coding.

Then, in Section 3, we consider certain properties of bipartite graphs that are relevant to the discussion of later sections. Thereafter, we will discuss the channel coding part of the problem in Section 4, resulting in an achievable rate region for the multiple access channel with correlated messages, which is one of the main results of this paper. Then, the complementary source coding part, the representation of correlated sources into message-graphs, will be described in Section 5. After that, some examples and interpretations are provided in Section 6. Section 7 provides some concluding remarks.

2 Preliminaries

In this section, we briefly overview the results available in the literature on the multiple access channel coding and Slepian-Wolf source coding. We also recall an interesting example given in [10], showing that the separate source and channel coding is not optimal, since it is closely related to our discussion.

2.1 Multiple Access Channel Capacity with Independent Messages

We summarize the well-known results [8] of the multiple access channel capacity in this section. We are given a multiple access channel characterized by a conditional distribution $p(y|x_1, x_2)$ for a two-transmitter problem, with finite input alphabets $\mathcal{X}_1$, $\mathcal{X}_2$ respectively and a finite output alphabet $\mathcal{Y}$. The channel is assumed to be memoryless and stationary. In other words, a multiple access channel is an ordered tuple $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}, p(y|x_1, x_2))$.

Definition 1 *A transmission system with parameters* $(n, \Delta_1, \Delta_2, \tau)$ *for a multiple access channel* $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}, p(y|x_1, x_2))$ *would involve*

- *a set of mappings* $\{f_1, f_2, g\}$ *where:*

$$f_i : \{1, 2, \ldots, \Delta_i\} \to \mathcal{X}_i^n \tag{1}$$

$$g : \mathcal{Y}^n \to \{1, 2, \ldots, \Delta_1\} \times \{1, 2, \ldots, \Delta_2\} \tag{2}$$

- *a performance measure, given by the average probability of error:*

$$\tau = \sum_{i=1}^{\Delta_1} \sum_{j=1}^{\Delta_2} \frac{1}{\Delta_1 \Delta_2} Pr\left[g(Y^n) \neq (i, j) | X_1^n = f_1(i), X_2^n = f_2(j)\right]. \tag{3}$$

Definition 2 *A rate pair* (R_1, R_2) *is said to be* achievable *for the given multiple access channel if* $\forall \epsilon > 0$, *and for all sufficiently large* n, *there exists a transmission system as defined above with parameters* $(n, \Delta_1, \Delta_2, \tau)$ *with* $\frac{1}{n} \log \Delta_i > R_i - \epsilon$ *for* $i = 1, 2$ *and a corresponding decoder with the average probability of error* $\tau < \epsilon$.

The capacity region of the multiple access channel, denoted by $\mathcal{R}_{MA}$, is the set of all achievable rate pairs (R_1, R_2). This is given [1, 2] by the following information-theoretic characterization: $\mathcal{R}_{MA}$ is equal to the convex closure of the set of all (R_1, R_2), such that there exists a product distribution on the input $p_1(x_1)p_2(x_2)$, and

$$R_1 \leq I(X_1; Y | X_2), \tag{4}$$

$$R_2 \leq I(X_2; Y|X_1), \tag{5}$$

$$R_1 + R_2 \leq I(X_1, X_2; Y), \tag{6}$$

where $I(\cdot;\cdot)$ denote the mutual information [8].

2.2 Noiseless Encoding of Correlated Sources

We are given a pair of correlated sources (for a two-source problem), with a joint distribution $p(s,t)$ with finite alphabets $\mathcal{S}$ and $\mathcal{T}$. The sources are assumed to be memoryless and stationary. In other words, a pair of correlated sources is an ordered tuple $(\mathcal{S}, \mathcal{T}, p(s,t))$.

Definition 3 *A transmission system with parameters $(n, \Delta_1, \Delta_2, \tau)$ for representing a pair of correlated sources ($\mathcal{S}$, $\mathcal{T}$, $p(s,t)$) would involve*

- *a set of mappings $\{f_1, f_2, g\}$ where*

$$f_1 : \mathcal{S}^n \to \{1, 2, \ldots, \Delta_1\}, \quad f_2 : \mathcal{T}^n \to \{1, 2, \ldots, \Delta_2\} \tag{7}$$

$$g : \{1, 2, \ldots, \Delta_1\} \times \{1, 2, \ldots, \Delta_2\} \to \mathcal{S}^n \times \mathcal{T}^n \tag{8}$$

- *a performance measure given by the probability of error*

$$\tau = Pr\left[(S^n, T^n) \neq g(f_1(S^n), f_2(T^n))\right]. \tag{9}$$

Definition 4 *A rate pair (R_1, R_2) is said to be* achievable *for the given correlated sources if $\forall \epsilon > 0$ and for all sufficiently large n, there exists a transmission system as defined above with parameters $(n, \Delta_1, \Delta_2, \tau)$ with $\frac{1}{n} \log \Delta_i < R_i + \epsilon$ for $i = 1, 2$ and the probability of error $\tau < \epsilon$.*

The achievable rate region $\mathcal{R}_{SW}$, is the set of achievable rate pairs (R_1, R_2). This is given by [6] the following information theoretic characterization: $\mathcal{R}_{SW}$ is equal to the set of all (R_1, R_2) such that

$$R_1 \geq H(S|T), \tag{10}$$

$$R_2 \geq H(T|S), \tag{11}$$

$$R_1 + R_2 \geq H(S,T). \tag{12}$$

2.3 Joint source-channel coding

Consider the joint source-channel coding scheme studied in [10]. We are given a pair of correlated sources (without a common part) and a multiple access channel.

Definition 5 *A transmission system with parameters (n, τ) for transmission of a pair of correlated sources $(\mathcal{S}, \mathcal{T}, p(s,t))$ over a multiple access channel $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}, p(y|x_1, x_2))$ would involve*

- *a set of mappings* $\{f_1, f_2, g\}$ *where*

$$f_1 : \mathcal{S}^n \rightarrow \mathcal{X}_1^n, \quad f_2 : \mathcal{T}^n \rightarrow \mathcal{X}_2^n, \tag{13}$$

$$g : \mathcal{Y}^n \rightarrow \mathcal{S}^n \times \mathcal{T}^n \tag{14}$$

- *a performance measure given by the probability of error*

$$\tau = Pr[(S^n, T^n) \neq g(Y^n)] \tag{15}$$

Definition 6 *A pair of correlated sources is said to be transmissible over a multiple access channel if* $\forall \epsilon > 0$*, and for all sufficiently large* n*, there exists a transmission system as defined above with parameters* (n, τ) *such that* $\tau < \epsilon$*.*

Following [10], a pair of correlated sources is transmissible over a multiple access channel if,

$$H(S|T) < I(X_1; Y|X_2, T), \tag{16}$$

$$H(T|S) < I(X_2; Y|X_1, S), \tag{17}$$

$$H(S, T) < I(X_1, X_2; Y), \tag{18}$$

for some $p_1(x_1|s)$ and $p_2(x_2|t)$, where the joint distribution is obtained as $p(s, t, x_1, x_2, y) = p(s, t)\ p_1(x_1|s)\ p_2(x_2|t)\ p(y|x_1, x_2)$.

2.4 An Example of Correlated Sources over the Multiple Access Channel

Let us consider an interesting example given in [10], which shows the advantage of encoders that directly map sources into channel inputs (joint-source-channel coding). Consider the transmission of a set of correlated sources (S, T), with the joint distribution $p(s, t)$ given by $p(s = 0, t = 0) = p(s = 0, t = 1) = p(s = 1, t = 1) = 1/3$, over a multiple access channel defined by $\mathcal{X}_1 = \mathcal{X}_2 = \{0, 1\}$, $\mathcal{Y} = \{0, 1, 2\}$, $Y = X_1 + X_2$. Here $H(S, T) = \log 3 =$1.58 bits. On the other hand,

$$\max_{p(x_1)p(x_2)} I(X_1, X_2\ ; Y) \ = \ 1.5 \text{ bits.}$$

Thus $H(S, T) > I(X_1, X_2\ ; Y)$ for all $p_1(x_1)p_2(x_2)$. Consequently, it appears that there is no way, even with the use of Slepian-Wolf source coding of S and T, to use the multiple access channel to send S and T reliably. However, it is easy to see that with the choice $X_1 \equiv S$ and $X_2 \equiv T$, error-free transmission of the sources over the channel is possible. This example shows that separate source and channel coding described above is *not* optimal — the partial information that each of the random variables S and T contains about the other is destroyed in this separation. In the proposed approach (to be discussed next), we allow our codes to depend statistically on the source outputs. This induces some dependence between the codewords, which will help combat the adversities of the channel more effectively.

3 Graphs as discrete interface

In this section, we present the problem statement, and, to better understand the significance and the limitations of the results presented in the next section, consider some structural properties of graphs. The problem we are addressing is the simultaneous transmission of two discrete memoryless stationary correlated sources S and T over a discrete memoryless stationary multiple access channel as shown in Figure 2. The encoders are given by mappings $f_1 : \mathcal{S}^n \to \mathcal{X}_1^n$ and $f_2 : \mathcal{T}^n \to \mathcal{X}_2^n$. The decoder is given by a mapping $g : \mathcal{Y}^n \to \mathcal{S}^n \times \mathcal{T}^n$. The performance measure associated with this transmission system is the probability of decoding error:

$$Pr[(S^n, T^n) \neq g(Y^n)]. \tag{19}$$

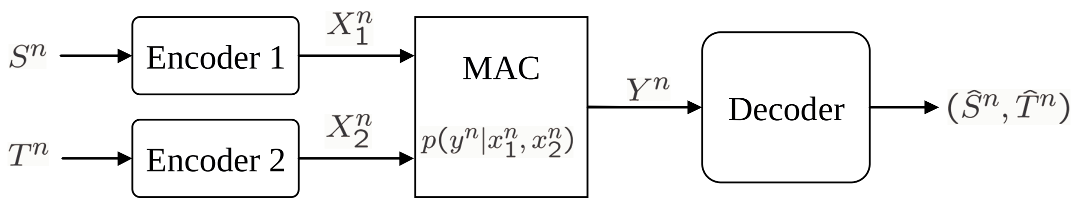

Figure 2: Transmission of correlated sources over a multiple access channel.

3.1 Basic concepts

Consider the following approach to this problem as shown in Figure 3. The system has two modules: the source coding module and the channel coding module. The sources are first represented efficiently using nearly semi-regular graphs in the source coding module. The edges coming from these nearly semi-regular graphs are reliably transmitted over the multiple access channel. The assumption is that the source coding module is going to produce message pairs

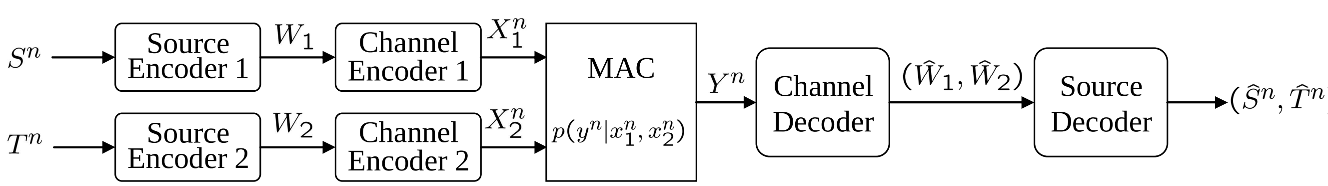

Figure 3: The sources are first mapped into edges in a nearly semi-regular graph, and the edges coming from this graph are reliably transmitted over a multiple access channel.

which have some relation between them. In other words, from the perspective of the channel coding module, the two senders have some integer message sets $\mathcal{W}_1 = \{1, 2, \ldots, |\mathcal{W}_1|\}$ and $\mathcal{W}_2 = \{1, 2, \ldots, |\mathcal{W}_2|\}$ respectively. Further, there is some correlation between the two messages, i.e., messages from each sender cannot be chosen independently. If the messages of the senders can be chosen independently, then all possible pairs (W_1, W_2) in the set $\mathcal{W}_1 \times \mathcal{W}_2$ can occur jointly. On the other hand, if they are correlated, only some pairs $(W_1, W_2) \in A$ occur, and the other pairs $(W_1, W_2) \notin A$ do not, where $A \subset \mathcal{W}_1 \times \mathcal{W}_2$. In more detail, we can think of these messages as follows.

- If the messages of the senders are independent, the message pairs (W_1, W_2) are equally likely with probability $\frac{1}{|\mathcal{W}_1 \times \mathcal{W}_2|}$.

- If the messages of senders are correlated, the message pairs $(W_1, W_2) \in A$ are equally likely with probability $\frac{1}{|A|}$, and the message pairs $(W_1, W_2) \notin A$ have probability zero.

As an example, let us consider the simple case as shown in Figure 1. In this case, two senders have $\mathcal{W}_1 = \mathcal{W}_2 = \{1, 2, 3\}$. The vertexes in the bipartite graph denote messages in the message sets, and an edge between two vertexes imply that the message pairs can occur jointly. The complete bipartite graph of Figure 1(a) corresponds to the case for which two messages from each sender can be chosen independently, so all the possible pairs can occur with equal probability $\frac{1}{9}$. Figure 1(b) and Figure 1(c) show the case for which two messages are correlated. In the case of Figure 1(b), each message pair (1, 1), (1, 2), (2, 2), (2, 3), (3, 3), and (3, 1) can occur with probability $\frac{1}{6}$, but (1, 3), (2, 1) and (3, 2) cannot occur. Similarly, only three message pairs (1, 1), (2, 2) and (3, 3) can occur with the same probability $\frac{1}{3}$ in case of Figure 1(c), which means that they are perfectly correlated. The messages of Figure 1(c) have higher correlation than those of Figure 1(b).

Before we discuss the main problem, let us first define a bipartite graph and related mathematical terms. Although our main results deal with nearly semi-regular graphs, for the purpose of illustration, we consider semi-regular graphs for this section alone.

Definition 7

- *A* bipartite graph *G is defined as an ordered tuple $G = (A_1, A_2, B)$ where A_1 and A_2 are two non-empty sets of vertexes, and B is a set of edges where every edge of B joins a vertex in A_1 to a vertex in A_2, i.e., $B \subseteq A_1 \times A_2$.*

- *If G is a bipartite graph, let $V_1(G)$ and $V_2(G)$ denote the first and the second vertex sets of G, respectively, and $E(G)$ denote the edge set of G.*

- *If $(i, j) \in E(G)$, then i and j are* adjacent*, or* neighboring *vertexes of G, and the vertexes i and j are* incident *to the edge (i, j).*

- *If each vertex in one set is adjacent to every vertex in the other set, then G is said to be a* complete *bipartite graph. In this case, $E(G) = V_1(G) \times V_2(G)$.*

- *The* degree*, or* valency*, $\deg_{G,i}(v)$ of a vertex $v \in V_i(G)$ in a graph G is the number of edges incident to v for $i = 1, 2$.*

- *A* subgraph *of a graph G is a graph whose vertex and edge sets are subsets of those of G.*

Since we consider a specific type of bipartite graphs in our discussion, let us define those bipartite graphs.

Definition 8

- *A bipartite graph G is said to have parameters $(\theta_1, \theta_2, \theta'_1, \theta'_2)$ if it satisfies:*

  – $|V_i(G)| = \theta_i$ *for i=1, 2,*

– $\forall u \in V_1(G)$, $\deg_{G,1}(u) = \theta_2'$,

– $\forall v \in V_2(G)$, $\deg_{G,2}(v) = \theta_1'$.

- *For two bipartite graphs G_1 and G_2, G_2 is said to* cover *G_1 if $E(G_1) \subseteq E(G_2)$.*

Definition 9 *With a bipartite graph G with parameters $(\theta_1, \theta_2, \theta_1', \theta_2')$, one can associate a pair of correlated messages with message sets $\mathcal{W}_1$ and $\mathcal{W}_2$, referred to as a* message-graph*, where $V_1(G) = \mathcal{W}_1$, $V_2(G) = \mathcal{W}_2$, and every edge in $E(G)$ denotes a message pair $(W_1, W_2) \in \mathcal{W}_1 \times \mathcal{W}_2$ which occurs with nonzero equal probability.*

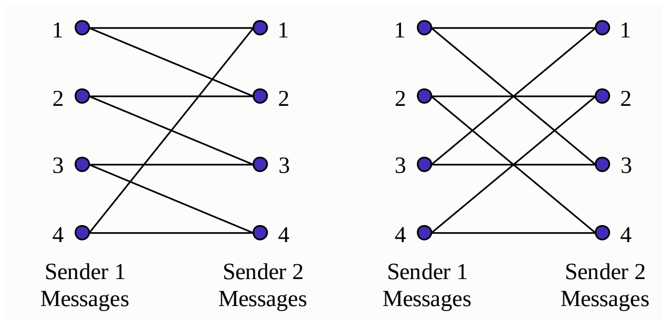

Figure 4: Examples of bipartite graphs $G(4,4,2,2)$: the message-graph characterized by the graph on the right side can be decomposed into three independent messages, with both encoders sharing a common message. This can be seen by renaming $1,2,3$ and 4 as $11,21,12$ and 22 respectively.

Figure 4 illustrates two examples of bipartite message-graphs $G(4,4,2,2)$. An example of the transmission system considered by [9], where the messages of the two users have a "common part" can be represented by the graph on the right side which can be divided into two complete bipartite graphs. In other words, the graph on the right side can be represented as a set of three independent messages, each of length 1, with the first user having the first and the second message sets, and the second user having the the first and the third message sets. This can be seen by renaming $1,2,3$ and 4 as $11,21,12$ and 22 respectively. Now each message of each user has two labels. As can be seen from the graph, for any valid message pair of the two users, the corresponding first labels are the same. Now if we consider each label as a message, then the first label of both users corresponds to the common message. Such graphs form a subset of all incomplete graphs as given in the above definition.

3.2 Equivalence classes of graphs

As shown in the previous example in Section 2.4, if we can design special codes which can translate the existing correlation between messages of two senders into the channel inputs, we might achieve higher transmission rates than those bounded by the conventional codes where all possible message pairs are assumed to jointly occur. To facilitate such an efficient discrete interface, one needs to answer the following question. For fixed $\theta_1, \theta_2, \theta_1', \theta_2'$, and n, since there are more than one bipartite graph which have the same above parameters, do we need to design a specific channel codes of block-length n for each graph having the parameters $(\theta_1, \theta_2, \theta_1', \theta_2')$ or it is possible to design a single channel code for all of these graphs?

To answer this question, let us consider the following example. Suppose we are given two message-graphs, given by A and B as shown in Figure 5. Suppose there exists a channel code for a multiple access channel with $n = 1$ which can reliably transmit the message pairs coming from A. In the first glance, it appears that this code cannot reliably transmit the message pairs coming from B as $B \neq A$. However, it turns out that one can indeed do so. This

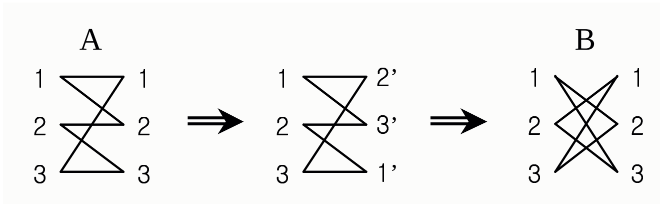

Figure 5: Example of permutation and relabeling.

is due to an interesting relation that exists between the two graphs. Note that if we permute the right vertexes of A, $(1, 2, 3)$, into $(2', 3', 1')$, relabel $(2', 3', 1')$ as $(2, 3, 1)$, and then move right vertexes together with their connected edges in natural order $(1, 2, 3)$, then we get graph B. This implies that we can use the given code to send the message pairs coming from B after simple permutation and relabeling. This procedure is illustrated in Figure 5. Clearly, we can also get graph A from graph B similarly. This motivates us to define equivalence classes of graphs having the same set of parameters.

Let us consider a set bipartite graphs having parameters (n, n, a, a) and denoted by $\mathcal{K}_{n,a}$, $n \in \mathbb{Z}^+$ where $\mathbb{Z}^+$ is the set of positive integers, and $a \in \{1, 2, \ldots, n\}$. For example, Figure 6 illustrates all the elements of $\mathcal{K}_{3,2}$. So there are totally six distinct bipartite graphs in the set $\mathcal{K}_{3,2}$. Now consider the generation of different bipartite graphs in

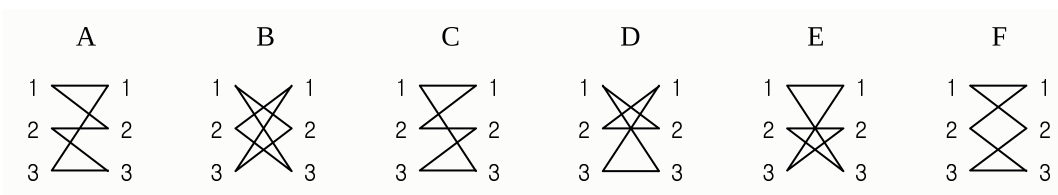

Figure 6: All the possible bipartite graphs in the set $\mathcal{K}_{3,2}$: any graph can be obtained from any other by permutation and relabeling.

$\mathcal{K}_{n,a}$ by permutation and relabeling of any one of them. Let $\mathcal{K}_{3,2}$={A, B, C, D, E, F}, where A, B, C, D, E, and F are the bipartite graphs shown in Figure 6. In this case, all the elements in $\mathcal{K}_{3,2}$ can be generated from any one element in the set by permutation and relabeling.

However, in the case of $n = 4$, $a = 2$, we cannot get all graphs in $\mathcal{K}_{4,2}$ by just permutation and relabeling of any one graph in the set. There are a total of 90 distinct bipartite graphs in $\mathcal{K}_{4,2}$, i.e., $|\mathcal{K}_{4,2}| = 90$. These 90 graphs can be divided into two mutually exclusive subsets, denoted by $\mathbf{S}_1$ and $\mathbf{S_2}$ where $|\mathbf{S}_1| = 72$ and $|\mathbf{S}_2| = 18$. Figure 7 shows one graph from each subset. It can be verified that all the graphs in the subset $\mathbf{S}_i$ (for $i = 1, 2$) can be obtained from any graph in that set by permutation and relabeling of the vertexes. However, no graph in the subset $\mathbf{S}_1$ can be obtained by permutation and relabeling of a graph in $\mathbf{S}_2$ and vice versa. This is explained

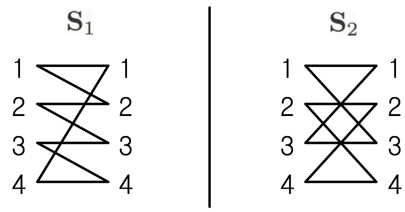

Figure 7: Graphs in the subsets of $\mathcal{K}_{4,2}$

in detail in Appendix A. In other words, the set $\mathcal{K}_{4,2}$ can be partitioned into two *equivalence classes* where the cardinalities of these classes are 72 and 18 respectively, and the equivalence relation is characterized by the feasibility of obtaining one element in the class by permutation and relabeling of the vertexes of the other. It can be shown that a message-graph characterized by every graph in the set $\mathbf{S}_2$ has a common part. This means that each graph in the set $\mathbf{S}_2$ can be divided into two complete bipartite graphs. Similarly, one can partition the set of all bipartite graphs with a given set of parameters into equivalence classes. At this point, a precise characterization of the number of equivalence classes even in $\mathcal{K}_{n,a}$ is an open question in combinatorics. However, it should be noted that some of the issues regarding the combinatorics of such graphs has been studied in the recent mathematics literature [23, 24]. Further, the complexity of the algorithms required for testing whether two graphs belong to an equivalence class are addressed in the computer science literature [25, 26, 27].

Remark 1 We summarize the conclusions of this discussion:

- All graphs having the same set of parameters can be partitioned into equivalence classes, where one element in a class can be obtained from the other in the same class by permutation and relabeling. Thus if we have a channel code which can reliably transmit message pairs coming from a graph (say G_1), then it can be easily used to reliably transmit message pairs coming from any graph that belongs to the equivalence class of G_1.
- The graphs having the same set of parameters but belonging to different equivalence classes may have different correlation structures.

4 Multiple Access Channel with Correlated Messages

In this section we give a characterization of the transmissibility of certain message-graphs over a multiple access channel.

4.1 Summary of Results

We are given a stationary discrete memoryless multiple access channel with conditional distribution $p(y|x_1, x_2)$, with input alphabets given by finite sets $\mathcal{X}_1$ and $\mathcal{X}_2$, and a finite output alphabet $\mathcal{Y}$. Although, ideally, we would want to use semi-regular graphs for source representation and communication of information, for the sake of analytical tractability, as is typical in Shannon theory, we will allow some slack with regard to the degrees of the vertexes of these graphs, and consider the asymptotic case when this slack is reduced to an arbitrarily small value. In other words, we consider bipartite graphs which are nearly semi-regular in our discussion.

Definition 10 *A bipartite graph G is said to have parameters $(\Delta_1, \Delta_2, \Delta'_1, \Delta'_2, \mu)$ if it satisfies:*

- $V_i(G) = \{1, 2, \ldots, \Delta_i\}$ *for* $i = 1, 2$,
- $\forall u \in V_1(G)$, $\Delta'_2 \mu^{-1} \leq \deg_{G,1}(u) \leq \Delta'_2 \mu$,
- $\forall v \in V_2(G)$, $\Delta'_1 \mu^{-1} \leq \deg_{G,2}(v) \leq \Delta'_1 \mu$.

Note that $\mu > 1$ is a slack parameter.

Definition 11 *An (n, τ)-transmission system for a bipartite graph G with parameters $(\Delta_1, \Delta_2, \Delta'_1, \Delta'_2, \mu)$ and a multiple access channel $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}, p(y|x_1, x_2))$ with correlated messages would involve:*

1. *encoding mappings $\{f_1, f_2\}$ and a decoding mapping g where:*

$$f_i : V_i(G) \to \mathcal{X}_i^n \quad \text{for} \quad i = 1, 2, \tag{20}$$

$$g : \mathcal{Y}^n \to E(G), \tag{21}$$

2. *a performance measure given by the following average probability of error criterion:*

$$\tau = \frac{1}{|E(G)|} \sum_{(i,j) \in E(G)} Pr\left[g(Y^n) \neq (i, j) | X_1^n = f_1(i), X_2^n = f_2(j)\right]. \tag{22}$$

Definition 12 *A tuple of rates (R_1, R_2, R'_1, R'_2) is said to be* achievable *for a given multiple access channel with correlated message sets, if for any $\epsilon > 0$, and for all sufficiently large n, there exists a bipartite graph G with parameters $(\Delta_1, \Delta_2, \Delta'_1, \Delta'_2, \mu)$ and an associated (n, τ)-transmission system as defined above satisfying: $R_i - \epsilon < \frac{1}{n} \log \Delta_i$, $R'_i - \epsilon < \frac{1}{n} \log \Delta'_i$ for $i = 1, 2$, $\frac{1}{n} \log \mu < \epsilon$ and the corresponding average probability of error $\tau < \epsilon$.*

Note that in the above definition, we have taken an optimistic point of view. As long as one can find a sequence of nearly semi-regular graphs where the number of vertexes and the degrees are increasing exponentially with given rates, such that the edges from these graphs are reliably transmitted over the given multiple access channel, we allow the corresponding rate tuples to belong to the achievable rate region. The goal is to find the achievable rate region $\mathcal{R}$ which is the set of all achievable tuple of rates (R_1, R_2, R'_1, R'_2). In the following we provide an information-theoretic characterization of an achievable rate region.

Theorem 1 *For an input probability distribution $p(x_1, x_2)$ defined on $\mathcal{X}_1 \times \mathcal{X}_2$, if a tuple (R_1, R_2, R'_1, R'_2) satisfies the following conditions,*

$$R_1 < I(X_1; Y|X_2) + I(X_1; X_2) = I(X_1; Y, X_2), \tag{23}$$

$$R_2 < I(X_2; Y|X_1) + I(X_1; X_2) = I(X_2; Y, X_1), \tag{24}$$

$$R_1 + R_2 < I(X_1, X_2; Y) + I(X_1; X_2) \tag{25}$$

$$R_i' < R_i - I(X_1; X_2), \quad \textit{for i=1, 2,} \tag{26}$$

then it belongs to the achievable rate region for the multiple access channel with correlated messages.

First, note that $I(X_1; X_2)$ is the bonus in sumrate we get by exploiting the correlation in the messages. Second, in this characterization, there is a constraint on the input distribution that one can choose for determining the rate region. For any $p(x_1, x_2)$, for a fixed sumrate of $I(X_1, X_2; Y) + I(X_1; X_2)$, the minimum value that R_1 can take is $I(X_1; Y)$, and the minimum value that R_2 can take is $I(X_2; Y)$. Hence the constraint on $p(x_1, x_2)$ is that $\min\{I(X_1; Y), I(X_2; Y)\} > I(X_1; X_2)$. Further, for the case of independent message sets, $R_i' = R_i$ for $i = 1, 2$, and the rate region reduces to that of the standard multiple access channels with $I(X_1; X_2) = 0$.

Theorem 2 *Any sequence (indexed by n) of $(n, \tau(n))$-transmission systems for a sequence of bipartite graphs G_n, respectively, with parameters $(2^{nR_1}, 2^{nR_2}, 2^{nR_1'}, 2^{nR_2'}, \mu(n))$ and a multiple access channel $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}, p(y|x_1, x_2))$ with correlated messages such that $\tau(n) \to 0$ and $\frac{1}{n} \log \mu(n) \to 0$ as $n \to \infty$ must satisfy:*

$$R_1 + R_2' \leq I(X_1, X_2; Y), \tag{27}$$

$$R_1' + R_2 \leq I(X_1, X_2; Y). \tag{28}$$

for some input distribution $p(x_1, x_2)$ on $\mathcal{X}_1 \times \mathcal{X}_2$.

Remark 2 The limitations of this theorem are illustrated in the following. Note that this theorem gives only a partial characterization of the set of all nearly semi-regular graphs whose edges can be reliably transmitted over a multiple access channel. In the formulation of the achievable rate region, we have the freedom of choosing a particular message-graph for every block-length n. The theorem characterizes the exponent of the rate of growth (as a function of the number of channel uses) of the size of certain nearly semi-regular graphs, such that edges coming from any such graph can be reliably transmitted over the multiple access channel. This obviously also means that it is possible to transmit edges coming from a graph belonging to the equivalence class of any of these graphs. However, the fact that edges coming from a graph (with certain parameters) are reliably transmitted does not mean that the edges coming from any graph with those parameters can be reliably transmitted.

4.2 Proof of Theorem 1 and 2

In this section, we present the proof of Theorem 1 and 2. We use random coding, and the notion of jointly typical sequences as given in [8].

Given the multiple access channel with distribution $p(y|x_1, x_2)$, consider a fixed joint distribution $p(x_1, x_2)$ on $\mathcal{X}_1 \times \mathcal{X}_2$. Also fix $\epsilon > 0$ and positive real numbers R_1, R_2. Without loss of generality, let us assume $R_i > I(X_1; X_2)$ for $i = 1, 2$. Let $R_i' = R_i - I(X_1; X_2)$ for $i = 1, 2$.

Codebook generation: Draw 2^{nR_1} codewords $X_1^n(i)$, $i \in \{1, 2, \ldots, 2^{nR_1}\}$, of length n, independently from the strongly ϵ-typical set $A_\epsilon^{(n)}(X_1)$. That is, $P\{X_1^n(i) = x_1^n\} = \frac{1}{|A_\epsilon^{(n)}(X_1)|}$ if $x_1^n \in A_\epsilon^{(n)}(X_1)$, and $P\{X_1^n(i) = x_1^n\} = 0$ if $x_1^n \notin A_\epsilon^{(n)}(X_1)$. Let us denote this codebook $\mathbb{C}_1$. Similarly, the second codebook $\mathbb{C}_2$ can be generated by choosing 2^{nR_2} codewords $X_2^n(i)$, $i \in \{1, 2, \ldots, 2^{nR_2}\}$ according to a uniform distribution over $A_\epsilon^{(n)}(X_2)$.

Graph generation: As shown in Figure 8, with $\mathbb{C}_1$ and $\mathbb{C}_2$, let us associate a graph $\mathbb{G}$ such that (i) $V_i(\mathbb{G}) = \mathbb{C}_i$ for $i = 1, 2$, (ii) $\forall (x_1^n, x_2^n) \in \mathbb{C}_1 \times \mathbb{C}_2$, we have $(x_1^n, x_2^n) \in E(\mathbb{G})$ if (x_1^n, x_2^n) is strongly jointly ϵ-typical, i.e., $(x_1^n, x_2^n) \in A_\epsilon^{(n)}(X_1, X_2)$. Then, there exists a set of $|E(\mathbb{G})|$ strongly jointly ϵ-typical sequence pairs in $\mathbb{C}_1 \times \mathbb{C}_2$. We label them by k, $k \in \{1, 2, \ldots, |E(\mathbb{G})|\}$. For each k, let $(X_1^n(i_k), X_2^n(j_k))$ denote the corresponding strongly jointly ϵ-typical sequence pair where $i_k \in \{1, 2, \ldots, 2^{nR_1}\}$ and $j_k \in \{1, 2, \ldots, 2^{nR_2}\}$.

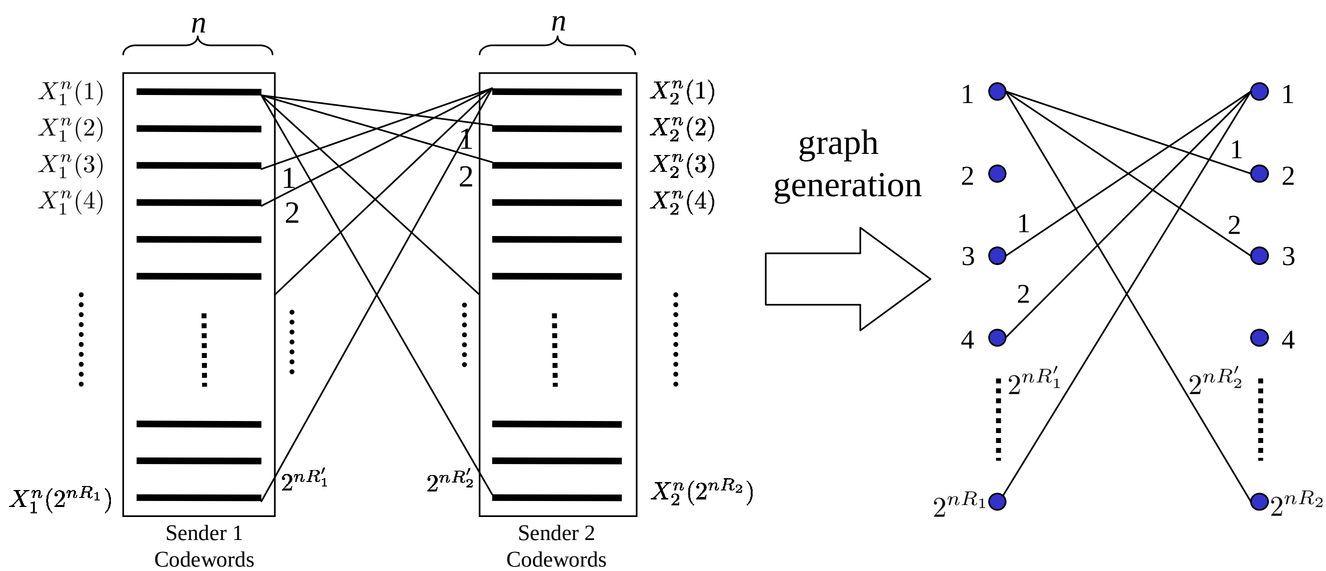

Figure 8: "Correlated" random codebook: The relation of joint typicality between two codewords, one from each codebook, induces a graph.

Encoder error events: Before we proceed to the encoding and decoding procedure, we need to make sure that the generated codebooks satisfy certain properties. If the normalized exponent of $|E(\mathbb{G})|$ of the generated graph $\mathbb{G}$ is greater than $I(X_1, X_2; Y)$, then every codeword pair $(x_1^n, x_2^n) \in E(\mathbb{G})$ cannot be reliably transmitted, resulting in errors. Moreover, if vertexes of $\mathbb{G}$ do not satisfy the degree conditions, the message pairs cannot be transmitted with arbitrary small probability of error. So, an encoding error will be declared if either one of the following events occurs. For this we need some properties [8] of strongly ϵ-typical sets.

- For a particular $x_1^n \in A_\epsilon^{(n)}(X_1)$, the probability that $(x_1^n, X_2^n) \in A_\epsilon^{(n)}(X_1, X_2)$ is bounded by

$$2^{-n(I(X_1;X_2)+\delta(\epsilon))} \leq P\{(x_1^n, X_2^n) \in A_\epsilon^{(n)}\} \leq 2^{-n(I(X_1;X_2)-\delta(\epsilon))} \tag{29}$$

  where X_2^n is obtained by using the uniform distribution on $A_\epsilon^{(n)}(X_2)$, and $\delta(\epsilon) \to 0$ as $\epsilon \to 0$,

- The probability that $(X_1^n, X_2^n) \in A_\epsilon^{(n)}(X_1, X_2)$ is bounded by

$$2^{-n(I(X_1;X_2)+\delta_1(\epsilon))} \leq P\{(X_1^n, X_2^n) \in A_\epsilon^{(n)}\} \leq 2^{-n(I(X_1;X_2)-\delta_1(\epsilon))} \tag{30}$$

  where for $i = 1, 2$, X_i^n is obtained by using the uniform distribution on $A_\epsilon^{(n)}(X_i)$, and $\delta_1(\epsilon) \to 0$ as $\epsilon \to 0$.

The error events are given by

- E_0: $|E(\mathbb{G})| > 2^{n(I(X_1,X_2;Y)-\delta(\epsilon)-\epsilon)}$,
- E_1: $\exists X_1^n \in \mathbb{C}_1$ such that $\left|\frac{1}{n}\log \deg_{\mathbb{G},1}(X_1^n) - R_2'\right| > \delta(\epsilon) + \epsilon$,
- E_2: $\exists X_2^n \in \mathbb{C}_2$ such that $\left|\frac{1}{n}\log \deg_{\mathbb{G},2}(X_2^n) - R_1'\right| > \delta(\epsilon) + \epsilon$.

In the following we show that the probability of these error events can be made small under certain conditions.

Lemma 1 *For any $\epsilon > 0$, and sufficiently large n,*

$$P(E_0) < \frac{\epsilon}{7} \tag{31}$$

provided $R_1 + R_2 < I(X_1, X_2; Y) + I(X_1; X_2) - \delta(\epsilon) - \delta_1(\epsilon) - \epsilon$.

Proof: Refer to Appendix B.

Lemma 2 *Let us define two events $E_{0,1}$ and $E_{0,2}$ as follows. $E_{0,1}$: $\exists X_1^n(i) \in \mathbb{C}_1$ such that $\deg_{\mathbb{G},1}(X_1^n(i)) < 2^{n(R_2-I(X_1;X_2)-\delta(\epsilon)-\epsilon)}$ and $E_{0,2}$: $\exists X_2^n(i) \in \mathbb{C}_2$ such that $\deg_{\mathbb{G},2}(X_2^n(i)) < 2^{n(R_1-I(X_1;X_2)-\delta(\epsilon)-\epsilon)}$. Then for any $\epsilon > 0$, and sufficiently large n:*

$$P\{E_{0,i}\} < \frac{\epsilon}{14}, \quad \textit{for } i = 1, 2 \tag{32}$$

Proof: Refer to Appendix C.

Lemma 3 *Let us define two events $E_{0,1}^*$ and $E_{0,2}^*$ as follows. $E_{0,1}^*$: $\exists X_1^n(i) \in \mathbb{C}_1$ such that $\deg_{\mathbb{G},1}(X_1^n(i)) > 2^{n(R_2-I(X_1;X_2)+\delta(\epsilon)+\epsilon)}$ and $E_{0,2}^*$: $\exists X_2^n(i) \in \mathbb{C}_2$ such that $\deg_{\mathbb{G},2}(X_2^n(i)) > 2^{n(R_1-I(X_1;X_2)+\delta(\epsilon)+\epsilon)}$. Then for any $\epsilon > 0$, and sufficiently large n:*

$$P\{E_{0,i}^*\} < \frac{\epsilon}{14}, \quad \textit{for } i = 1, 2 \tag{33}$$

Proof: Refer to Appendix D.

Note that $E_1 = E_{0,1} \cup E_{0,1}^*$ and $E_2 = E_{0,2} \cup E_{0,2}^*$. So, following the above three lemmas, with high probability we can obtain a graph $\mathbb{G}$ where each vertex in $V_1(\mathbb{G})$ has degree nearly equal to $2^{n(R_2-I(X_1;X_2))}$ and each vertex in $V_2(\mathbb{G})$ has degree nearly equal to $2^{n(R_1-I(X_1;X_2))}$, and the total number of edges is nearly equal to $2^{nI(X_1,X_2;Y)}$ if the product of the vertexes of the corresponding two sets is nearly equal to $2^{n(I(X_1,X_2;Y)+I(X_1;X_2))}$.

Choosing message-graphs If any one of the above three events occurs, then choose any graph with parameters

$(2^{nR_1}, 2^{nR_2}, 2^{nR'_1}, 2^{nR'_2}, 2^{n(\delta(\epsilon)+\epsilon)})$ as the message-graph, and no guarantee will be given regarding the probability of decoding error. If none of these error events occurs, choose G as the message-graph, where (a) G has parameters $(2^{nR_1}, 2^{nR_2}, 2^{nR'_1}, 2^{nR'_2}, 2^{n(\delta(\epsilon)+\epsilon)})$, and (b) an integer pair $(i,j) \in E(G)$ if and only if $(X_1^n(i), X_2^n(j)) \in E(G)$.

Encoding: Sender 1 transmits the codeword $X_1^n(i)$ to send message index i; similarly, Sender 2 sends $X_2^n(j)$ to send message index j.

Decoding: At the receiver, the index pair (i,j) is chosen as the transmitted message pair only if there exists a unique pair (i,j) such that $(X_1^n(i), X_2^n(j), Y^n)$ is strongly jointly ϵ-typical in the sense of $p(x_1,x_2)p(y|x_1,x_2)$. Otherwise, an error is declared.

Probability of Error Analysis: The probability of error $P(E)$ can be given by

$$P(E) = P(E_0 \cup E_1 \cup E_2)P(E|E_0 \cup E_1 \cup E_2) + P(E \cap E_0^c \cap E_1^c \cap E_2^c) \tag{34}$$

$$\leq P(E_0 \cup E_1 \cup E_2) + P(E \cap E_0^c \cap E_1^c \cap E_2^c) \tag{35}$$

The second probability in the above equation can be bounded as given in the following lemma.

Lemma 4 *For any $\epsilon > 0$, and sufficiently large n,*

$$P(E \cap E_0^c \cap E_1^c \cap E_2^c) < \frac{4\epsilon}{7} \tag{36}$$

provided

$$R_1 < I(X_1; Y, X_2) - 7\epsilon_1, \tag{37}$$

$$R_2 < I(X_2; Y, X_1) - 7\epsilon_1, \tag{38}$$

$$R_1 + R_2 < I(X_1, X_2; Y) + I(X_1; X_2) - 7\epsilon_1, \tag{39}$$

where $\epsilon_1(\epsilon) \to 0$ as $\epsilon \to 0$, and $\epsilon_1(\epsilon)$ is a continuous function associated with certain strongly typical set.

Proof: Refer to Appendix E.

Therefore,

$$P(E) \leq P(E_0) + P(E_1) + P(E_2) + P(E \cap E_0^c \cap E_1^c \cap E_2^c) < \epsilon. \tag{40}$$

Since in every realization of random codebooks, we have chosen a message-graph with parameters $(2^{nR_1}, 2^{nR_2}, 2^{nR'_1}, 2^{nR'_2}, 2^{n(\delta(\epsilon)+\epsilon)})$, and averaged over the ensemble of random codebooks, the average probability of error is smaller than ϵ, there must exist a message-graph with parameters $(2^{nR_1}, 2^{nR_2}, 2^{nR'_1}, 2^{nR'_2}, 2^{n(\delta(\epsilon)+\epsilon)})$ and a codebook pair such that the average probability of error is smaller than ϵ. This is true only under the condition given by the statement of the theorem. Hence, the proof of Theorem 1 has been completed. ■

We have so far proved the achievability part of the capacity region. The proof of the converse part (Theorem 2) is very similar to that of conventional multiple access channel (with independent messages) [8]. So, we have used almost exactly the same technique except for the fact that message pairs are equally likely only if they belong to the edge set of the message-graph. Refer to Appendix F for details.

5 Representation of Correlated Sources into Message-graphs

In the previous section, we discussed the transmission of correlated messages over multiple access channels. In this section we consider the dual representation of correlated sources using nearly semi-regular bipartite graphs.

5.1 Summary of Results

We are given two correlated sources S and T with a joint probability distribution $p(s,t)$ with alphabets given by finite sets $\mathcal{S}$ and $\mathcal{T}$.

Definition 13 *An (n,τ)-transmission system for a bipartite graph G with parameters $(\Delta_1, \Delta_2, \Delta_1', \Delta_2', \mu)$ and a pair of correlated sources (S,T) would involve:*

1. *Encoder mappings f_1 and f_2:*

$$f_1 : \mathcal{S}^n \rightarrow V_1(G), \tag{41}$$

$$f_2 : \mathcal{T}^n \rightarrow V_2(G), \tag{42}$$

2. *A decoder mapping:*

$$g : E(G) \rightarrow \mathcal{S}^n \times \mathcal{T}^n, \tag{43}$$

3. *A performance measure given by the probability of error:*

$$\tau = Pr\big[\{g(f_1(S^n), f_2(T^n)) \neq (S^n, T^n)\} \cap \{(f_1(S^n), f_2(T^n)) \in E(G)\}\big] + Pr\big[(f_1(S^n), f_2(T^n)) \notin E(G)\big]. \tag{44}$$

Note that the rationale for choosing the above performance measure is the following. Since a channel coder for a multiple access channel with correlated messages provides guarantees on the probability of error only if the message pair belongs to a graph of certain parameters and no guarantees will be given otherwise, the source coder has to take this event into account while calculating the probability that the reconstruction source vectors are not equal to the vectors observed by the encoders.

Definition 14 *A tuple of rates (R_1, R_2, R_1', R_2') is said to be* achievable *for a distributed source coding problem with correlated sources (S,T), if for any $\epsilon > 0$, and for all sufficiently large n, there exists a bipartite graph G with parameters $(\Delta_1, \Delta_2, \Delta_1', \Delta_2', \mu)$ and an associated (n,τ)-transmission system as defined above satisfying: $\frac{1}{n}\log \Delta_i < R_i + \epsilon$, $\frac{1}{n}\log \Delta_i' < R_i' + \epsilon$ for $i = 1, 2$, $\frac{1}{n}\log \mu < \epsilon$ and the corresponding average probability of error $\tau < \epsilon$.*

The goal is to find the achievable rate region $\mathcal{R}_{DS}$ which is the set of all achievable tuple of rates (R_1, R_2, R_1', R_2'). An achievable rate region $\mathcal{R}_{DS}$ is given by the following theorem, which is the main result.

Theorem 3 *The achievable rate region for a distributed source coding problem with correlated sources (S, T) is given by the set of all (R_1, R_2, R_1', R_2') such that*

$$R_1 \geq H(S|T), \tag{45}$$

$$R_2 \geq H(T|S), \tag{46}$$

$$R_1 + R_2 \geq H(S, T), \tag{47}$$

$$R_1 + R_2' = R_1' + R_2 \geq H(S, T). \tag{48}$$

Remark 3 In Theorem 3, as in Theorems 1 and 2, there are limitations. The theorem gives only a partial characterization of the set of all nearly semi-regular bipartite graphs that can be used to represent the given pair of correlated sources. As in channel coding, in the above formulation of the achievable rate region, we have the freedom to select the message-graph for every block-length n.

5.2 Proof of Theorem 3

In this section, we present the proof of the main result. We use the random binning technique used by Berger [28], and the notion of strongly jointly typical sequences. First, we will prove the direct coding theorem. Let us consider a fixed joint distribution $p(s,t)$ on $\mathcal{S} \times \mathcal{T}$. Also fix $\epsilon > 0$ and real numbers R_1, R_2. Without loss of generality, let us assume that $R_i < H(S,T)$ for $i = 1, 2$. Let $R_1' = H(S,T) - R_2$ and $R_2' = H(S,T) - R_1$.

Bin generation: Let us define $\alpha = 2^{n(H(S)-R_1+\gamma)}$ and $\beta = 2^{n(H(T)-R_2+\gamma)}$ where γ will be specified shortly[1]. Draw α sequences S^n of length n independently and uniformly with replacement from the strongly ϵ-typical set $A_\epsilon^{(n)}(S)$. Then, put the all selected α sequences into a bin named B_1. Repeat the same procedure 2^{nR_1} times independently, resulting in 2^{nR_1} bins denoted by B_i for $i \in \{1, 2, \ldots, 2^{nR_1}\}$. Similarly, generate 2^{nR_2} bins denoted by C_j for $j \in \{1, 2, \ldots, 2^{nR_2}\}$, where each bin contains β sequences T^n, from the strongly ϵ-typical set $A_\epsilon^{(n)}(T)$.

Graph generation: As shown in Figure 9, we can associate a bipartite graph G with the bin indexes i and j of the generated bins B_i and C_j where (i) $V_i(G) = \{1, 2, \ldots, 2^{nR_i}\}$ for $i = 1, 2$, (ii) $\forall (i,j) \in V_1(G) \times V_2(G)$, $(i,j) \in E(G)$ if there exists a strongly jointly ϵ-typical sequence pair $(S^n, T^n) \in B_i \times C_j$.

Encoding error events: As done in the previous section regarding channel coding for the multiple access channel, before we proceed further, let us make sure the generated codebooks satisfy certain properties. If the vertexes of G do not satisfy certain degree requirements, we may not be able to reliably represent the sources using this graph. So, an encoding error will be declared if either one of the following events occurs. For this, let us define a continuous function $\epsilon_1(\epsilon)$ [8] as follows: $\epsilon_1(\epsilon) \to 0$ as $\epsilon \to 0$ and

[1] Note that this is a standard technique which uses a slack parameter to construct random bins from typical sets [28].

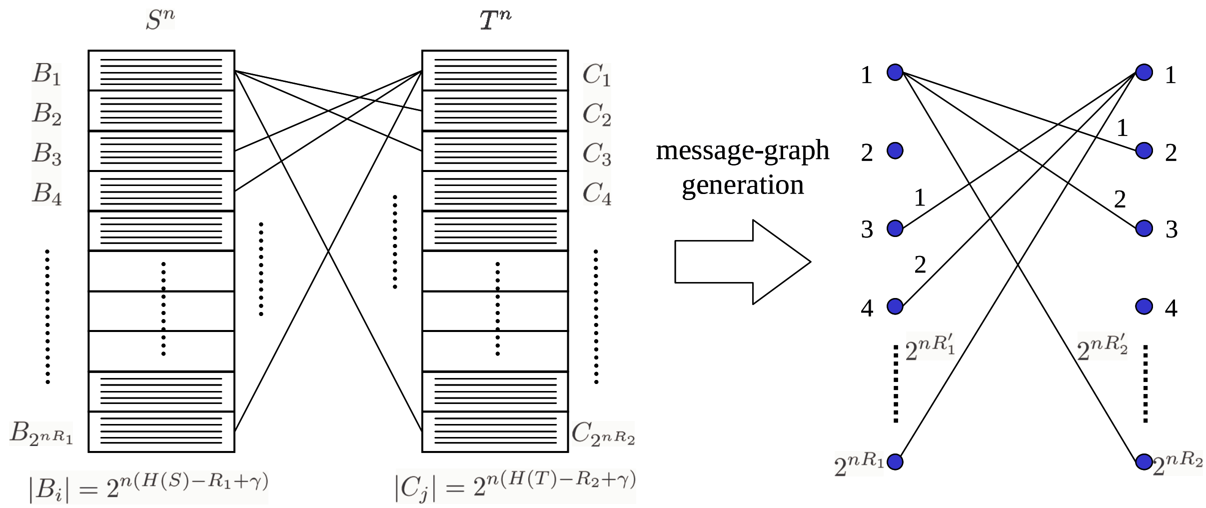

Figure 9: A bin-index graph (message-graph) generation from the pair of correlated sources (S, T).

- $2^{n(H(S)-\epsilon_1(\epsilon))} \leq |A_\epsilon^{(n)}(S)| \leq 2^{n(H(S)+\epsilon_1(\epsilon))}$
- $2^{n(H(T)-\epsilon_1(\epsilon))} \leq |A_\epsilon^{(n)}(T)| \leq 2^{n(H(T)+\epsilon_1(\epsilon))}$.

Choose $\gamma > \epsilon_1(\epsilon)$, and ϵ' such that $\epsilon' > 2\gamma + 3\epsilon_1(\epsilon)$. The error events are defined as:

- E_1: $\exists i \in V_1(G)$ such that $\left|\frac{1}{n}\log \deg_{G,1}(i) - R'_2\right| > \epsilon'$,
- E_2: $\exists j \in V_2(G)$ such that $\left|\frac{1}{n}\log \deg_{G,2}(j) - R'_1\right| > \epsilon'$,

We now show that the probability of these events can be made arbitrarily small for sufficiently large n. To bound these probabilities we first need the following lemma about a technical result.

Lemma 5 *Suppose U and V are two correlated finite-alphabet random variables with joint distribution $p(u, v)$. For any $\epsilon > 0$ and any positive real numbers R_1 and R_2 such that $R_1 + R_2 > I(U; V)$, if two collections of sequences $\mathcal{C}_U$ and $\mathcal{C}_V$ are generated with uniform distribution (with replacement) on the typical sets $A_\epsilon^{(n)}(U)$ and $A_\epsilon^{(n)}(V)$ of size 2^{nR_1} and 2^{nR_2}, respectively, then the probability $P_\epsilon(n)$ of not finding any jointly strongly ϵ-typical pair from these collections satisfies the following relation:*

$$\lim_{n\to\infty} -\frac{1}{n}\log P_\epsilon(n) = \infty. \tag{49}$$

Proof: The proof of Lemma 5 is long and technical in nature, hence omitted.

Lemma 6 *Let us define two events $E_{0,1}$ and $E_{0,2}$ as follows: $E_{0,1}$: $\exists i \in V_1(G)$ such that $\deg_{G,1}(i) < 2^{n(H(S,T)-R_1-\epsilon')}$ and $E_{0,2}$: $\exists j \in V_2(G)$ such that $\deg_{G,2}(j) < 2^{n(H(S,T)-R_2-\epsilon')}$. Then for any $\epsilon > 0$, and sufficiently large n:*

$$P\{E_{0,i}\} < \frac{\epsilon}{12}, \quad \textit{for } i = 1, 2 \tag{50}$$

Proof: Refer to Appendix G.

Lemma 7 *Let us define two events $E^*_{0,1}$ and $E^*_{0,2}$ as follows: $E^*_{0,1}$: $\exists i \in V_1(G)$ such that $\deg_{G,1}(i) > 2^{n(H(S,T)-R_1+\epsilon')}$, and $E^*_{0,2}$: $\exists j \in V_2(G)$ such that $\deg_{G,2}(i) > 2^{n(H(S,T)-R_2+\epsilon')}$. Then for any $\epsilon > 0$, and sufficiently large n:*

$$P\{E^*_{0,i}\} < \frac{\epsilon}{12}, \quad \textit{for } i = 1, 2 \tag{51}$$

Proof: Refer to Appendix H.

So, by Lemma 6 and Lemma 7, with high probability we can obtain a bipartite message-graph G where each vertex in $V_1(G)$ has degree nearly equal to $2^{n(H(S,T)-R_1)}$ and each vertex in $V_2(G)$ has degree nearly equal to $2^{n(H(S,T)-R_2)}$.

Choosing message-graphs: If any of E_1 or E_2 occurs, then choose any graph with parameters ($2^{nR_1}, 2^{nR_2}$, $2^{nR'_1}, 2^{nR'_2}$, $2^{n\epsilon'}$) as the message-graph G_M, and no guarantees will be given regarding the probability of error. If none of these events occur, then choose G as the message graph G_M.

Encoding: Define an encoding function $f_1(S^n)$ as follows. If a source sequence S^n belongs to at least one of the bins (B_i's), then $f_1(S^n)$ is the smallest index i such that $S^n \in B_i$; otherwise $f_1(S^n) = 0$. For the other source sequence T^n, $f_2(T^n)$ can be similarly defined, i.e., $f_2(T^n)$ is the smallest index j such that $T^n \in C_j$; otherwise $f_2(T^n) = 0$.

Decoding: Given the received index pair (i_0, j_0), declare the reconstruction pair $g(i_0, j_0)$ as $(\hat{s}^n, \hat{t}^n)$ if there exists a unique pair of sequences $(\hat{s}^n, \hat{t}^n)$ such that $(\hat{s}^n, \hat{t}^n) \in B_{i_0} \times C_{j_0}$ and strongly jointly ϵ-typical. Otherwise, declare an error.

Probability of error analysis: Let E denote the event

$$(\{g(f_1(S^n), f_2(T^n)) \neq (S^n, T^n)\} \cap \{(f_1(S^n), f_2(T^n)) \in E(G_M)\}) \cup \{(f_1(S^n), f_2(T^n)) \notin E(G_M)\}, \tag{52}$$

that the index pair transmitted by the encoders do not belong to $E(G_M)$, or that the reconstruction vector pair is not equal to the source vector pair with the transmitted index pair belonging to $E(G_M)$. The probability of error $P(E)$ can be given by

$$P(E) = P(E_1 \cup E_2)P(E|E_1 \cup E_2) + P(E \cap E_1^c \cap E_2^c) \tag{53}$$

$$\leq P(E_1 \cup E_2) + P(E \cap E_1^c \cap E_2^c) \tag{54}$$

The second probability in the above equation can be bounded as given in the following lemma.

Lemma 8 *For any $\epsilon > 0$, and sufficiently large n,*

$$P(E \cap E_1^c \cap E_2^c) \leq \frac{2\epsilon}{3} \tag{55}$$

provided

$$R_1 > H(S|T) + 2\gamma + 2\epsilon_1, \tag{56}$$

$$R_2 > H(T|S) + 2\gamma + 2\epsilon_1, \tag{57}$$

$$R_1 + R_2 > H(S,T) + 2\gamma + 2\epsilon_1, \tag{58}$$

Proof: Refer to Appendix I.

Therefore $P(E) < \epsilon$ for sufficiently large n and under the conditions given by the theorem. As in the previous section, in every realization of random codebooks, we have obtained a message-graph G_M with the same constraint on its parameters, and averaged over this ensemble, we have made sure that the probability of error is within the tolerance level of ϵ. Hence the proof of the direct coding theorem is completed.

The converse part of Theorem 3 can be obtained using techniques that are similar to those used in the Slepian-Wolf source coding theorem [8]. The only difference is that the messages are correlated. The proof is given in Appendix J.

5.3 Different Message-Graphs for a pair of Correlated Sources (S, T)

To shed more light on the representation of sources into nearly semi-regular graphs, let us consider the following illustration shown in Figure 10. Consider three important points A, B, and C in the shaded area in Figure 10. We

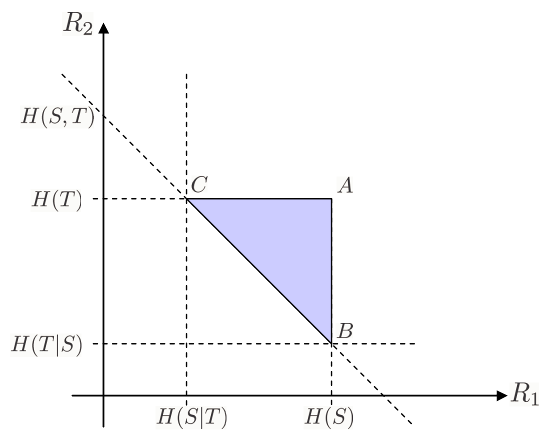

Figure 10: Achievable rate region where each point (R_1, R_2) can be associated with a different message-graph that represents the pair of correlated sources (S, T) for transmission over multiple access channels.

can make the following observations.

- Point A: In this case, the the achievable rate tuple is $(H(S), H(T), H(S|T), H(T|S))$. Roughly speaking, this corresponds to the typicality-graph of (S, T). In other words, this is an efficient representation of the source that has maximum redundancy in the conventional sense. For this point, in the direct coding theorem, the bin size that is used is roughly unity.

- Point B: In this case, the achievable rate tuple is $(H(S), H(T|S), H(S), H(T|S))$. Roughly speaking, this is an efficient representation of the source that has least redundancy in the conventional sense. For this point, in the direct coding theorem, the bin sizes that are used are roughly unity and $2^{nI(S;T)}$, respectively. In this case we get a nearly complete graph.

Hence Point A can be thought of as situated at one end of the spectrum, and points on the line BC as situated on the other end of the spectrum. For every point in the triangle ABC, we get an equally efficient representation of the sources into a nearly semi-regular graph.

6 Examples and Interpretations

6.1 End-to-End Performance

In the previous sections, we have considered a discrete interface to transmit correlated sources over multiple access channels. The main idea is that we can send these source more reliably and efficiently by exploiting the correlation structure in the given sources without merging the source coding and the channel coding blocks. We use a nearly semi-regular bipartite graph as an interface between source coding and channel coding to capture and translate the correlation from the sources to the channel inputs. Now if one considers the overall end-to-end performance, the performance bound given by the theorems of the previous sections is that $H(S,T) < I(X_1, X_2; Y)$. Using the max-flow-min-cut theorems [8], clearly one can see that this the best that one can hope for.

6.2 Gaussian Multiple Access Channel

The coding theorem given in the previous section can be extended to continuous-alphabet sources using the standard techniques [29, 30]. Consider the Gaussian multiple access channel with the channel input distribution being jointly Gaussian. There are two senders and one receiver. Each of the inputs has a power constraint, given by $E[X_i^2] \leq P_i$ for $i = 1, 2$. The received signal Y is given by

$$Y = X_1 + X_2 + Z \tag{59}$$

where Z is zero mean Gaussian random variable with variance N, denoted by $Z \sim \mathcal{N}(0, N)$, and X_1 and X_2 are zero mean jointly Gaussian random variables with covariance matrix K given by

$$K = \begin{bmatrix} P_1 & \rho\sqrt{P_1 P_2} \\ \rho\sqrt{P_1 P_2} & P_2 \end{bmatrix} \tag{60}$$

where ρ is the correlation coefficient. By evaluating the information quantities, we can obtain the following achievable rate region

$$R_1 \leq I(X_1; Y, X_2) = \frac{1}{2} \log \left(\frac{1}{1-\rho^2} + \frac{P_1}{N} \right) \tag{61}$$

$$R_2 \leq I(X_2; Y, X_1) = \frac{1}{2} \log \left(\frac{1}{1-\rho^2} + \frac{P_2}{N} \right) \tag{62}$$

$$R_1 + R_2 \leq I(X_1, X_2; Y) + I(X_1; X_2) = \frac{1}{2} \log \left[\frac{1}{1-\rho^2} \left(1 + \frac{P_1 + P_2 + 2\rho\sqrt{P_1 P_2}}{N} \right) \right] \tag{63}$$

$$R_i' \leq R_i - I(X_1; X_2) = R_i - \frac{1}{2} \log \frac{1}{(1-\rho^2)}, \tag{64}$$

where $0 \le \rho \le \rho_{\max}$, and $\rho_{\max}$ is the maximum value of ρ within the interval $[0, 1]$ such that

$$\frac{\left(P_1 + P_2 + 2\rho\sqrt{P_1 P_2} + N\right)(1-\rho^2)}{(1-\rho^2)\left(\max\{P_1, P_2\}\right) + N} \ge 1. \tag{65}$$

The variation of the information quantities and the pair (R_1, R_2) as functions of the correlation coefficient ρ are

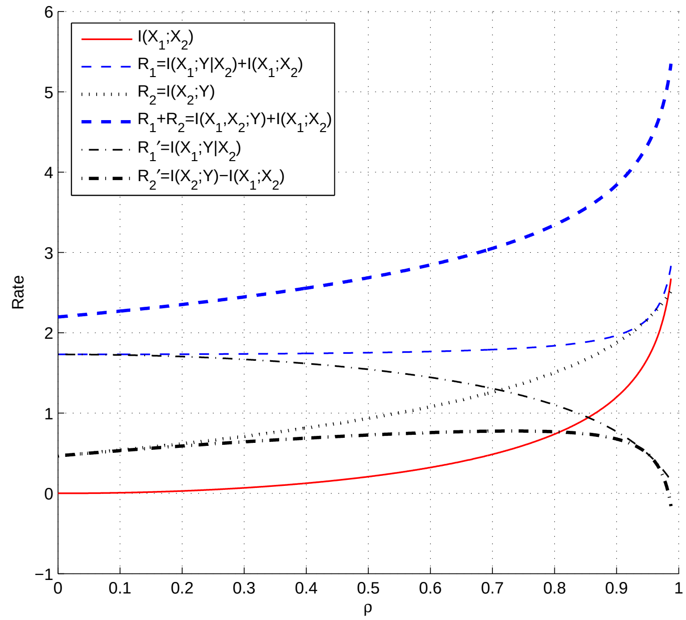

Figure 11: Variation of information quantities and the corresponding rates associated with the Gaussian multiple access channel as functions of the correlation coefficient ρ, for $P_1 = P_2 = 10$ and $N = 1$. $\rho_{\max} = 0.9877$.

plotted in Figure 11 and 12 for a particular choice of $P_1 = P_2 = 10$ and $N = 1$. Note that as ρ varies from zero to one, mutual information $I(X_1; X_2)$ increases from zero to ∞.

If X_1 and X_2 are independent, i.e., $\rho = 0$, this gives the well known capacity region of Gaussian multiple access channel with independent messages, which is the set of rate (R_1, R_2) pairs satisfying

$$R_1 < \frac{1}{2}\log\left(1 + \frac{P_1}{N}\right), \quad R_2 < \frac{1}{2}\log\left(1 + \frac{P_2}{N}\right), \quad R_1 + R_2 < \frac{1}{2}\log\left(1 + \frac{P_1 + P_2}{N}\right). \tag{66}$$

When ρ is less than about 0.9, as ρ become larger, R_1 increases very slowly, but R_2 increases rapidly. So the corner point B in Figure 12 moves almost upward in the (R_1, R_2) plane. When the roles of sender 1 and sender 2 are exchanged, the corner point C in Figure 12 moves almost rightwards. When $\rho \approx 0.9$, R_1 and R_2 have almost the same value, so the capacity region has a square shape. After that point both R_1 and R_2 increase very fast for $\rho > 0.9$, but R_1 become larger than R_2 again.

To get a picture of the variation of the entire rate region as a function of the correlation coefficient ρ, let us consider a special case where $P_1 = P_2 = P$, $R_1 = R_2 = R$ and $R_1' = R_2' = R'$ for the same Gaussian multiple access channel with jointly Gaussian channel input. In this case, an achievable rate region is the set of rate pairs (R, R')

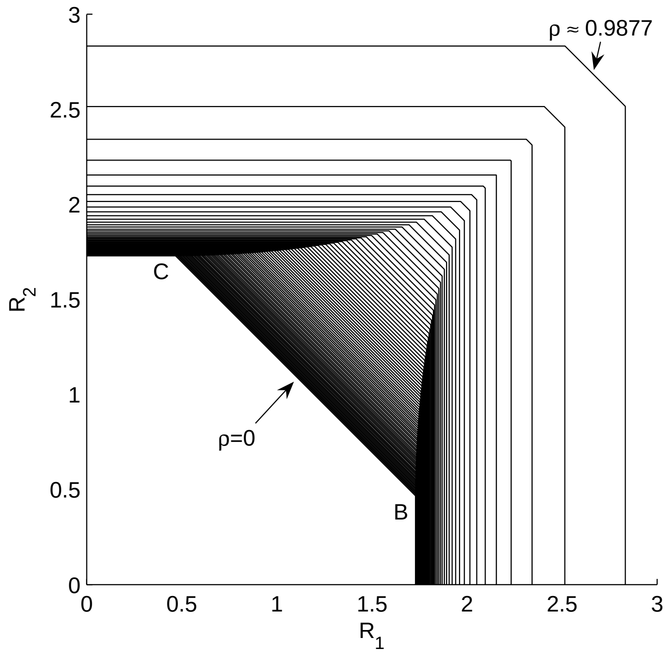

Figure 12: The variation of achievable tuples (R_1, R_2) (with non-negative (R_1', R_2')) as functions of correlation coefficient ρ for Gaussian multiple access channel with jointly Gaussian channel input, for $P_1 = P_2 = 10$ and $N = 1$. $\rho_{\max} = 0.9877$.

satisfying

$$R \leq \frac{1}{4} \log \left[\frac{1}{1-\rho^2} \left(1 + \frac{2P + 2\rho P}{N} \right) \right], \tag{67}$$

$$R' \leq R - \frac{1}{2} \log \left(\frac{1}{1-\rho^2} \right). \tag{68}$$

The boundary of the achievable rate region is thus given by

$$(R, R') = \left(\frac{1}{4} \log \left[\frac{1}{1-\rho^2} \left(1 + \frac{2P + 2\rho P}{N} \right) \right], \frac{1}{4} \log \left[(1-\rho^2) \left(1 + \frac{2P + 2\rho P}{N} \right) \right] \right). \tag{69}$$

Note that for this case, $\rho_{\max}$ is a monotone increasing function of P/N, approaching 1 as $P/N \to \infty$.

If X_1 and X_2 are independent, i.e., $\rho = 0$, this gives the well known capacity region of the Gaussian multiple access channel with independent messages, which is the set of rate pairs (R, R') satisfying

$$R \leq \frac{1}{4} \log \left(1 + \frac{2P}{N} \right), \qquad R' = R. \tag{70}$$

The achievable rate region is illustrated in Figure 13 as a function of the correlation coefficient. Using the boundary values for R and R' as given by (69), we get the bound for $R + R'$, which is given by

$$R + R' < \frac{1}{2} \log \left(1 + \frac{2P(1+\rho)}{N} \right). \tag{71}$$

We can use separate source and channel coding in order to send the same correlated messages over the Gaussian multiple access channel. In this case, we apply Slepian-Wolf source coding on the given correlated messages to

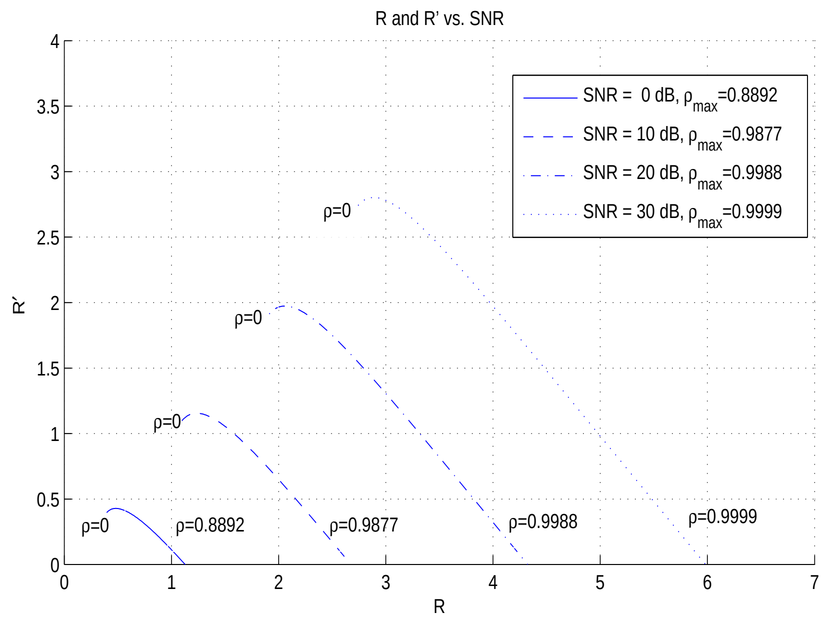

Figure 13: Variation of R and R' versus correlation coefficient ρ, for different signal to noise ratios.

transform them into new independent messages. This is followed by conventional multiple access channel coding, working on these new and transformed independent messages. For the same Gaussian multiple access channel, we consider the case where P' is the power constraint on the inputs, and the rates of the two encoders are the same. According to the Slepian-Wolf theorem (see Section 2.2), we can encode the given correlated message-graph into two messages of length nR and nR'. Using time-sharing, we can now assume that each encoder of the channel coding module has access to an independent message of length $(nR + nR')/2$. If we now use the given multiple access channel n times, and use conventional multiple access channel coding, the transmission power required to sustain reliable communication is given by

$$P' = \frac{N}{2}\left[2^{2(R+R')} - 1\right]. \tag{72}$$

Now if we substitute for R and R', the values on the boundary of achievable rate region given by Theorem 1, then we get

$$P' = P(1+\rho). \tag{73}$$

Now we can compare the two different schemes for sending correlated messages over the Gaussian multiple access channel. One is a coding scheme with correlated codewords which exploits the existing message correlation, the other is separate source and channel coding, working with independent codewords after applying Slepian-Wolf source coding. In order to have the same achievable rates in both schemes, we have the condition that $P' = (1+\rho)P > P$ if we choose a positive correlation coefficient. This means that if the given messages are not correlated, i.e., $\rho = 0$, then the required power in both schemes are exactly the same, but if the messages are correlated with $\rho > 0$, as expected, we can send the same amount of information with less power by encoding with correlated codewords.

6.3 Example of Binary-Input Multiple Access Channel of Section 2.4

Let us revisit the example considered in Section 2.4 that shows that error-free transmission of the given correlated sources over the given binary-input multiple access channel is possible with the special code $X_1 \equiv S$ and $X_2 \equiv T$. This can be considered as a match between the typicality-graph of the source and that of the channel input, i.e., a match between message-graph and the graph associated with the channel code. So by applying the theorem, we can calculate the achievable rate region for this case.

$$I(X_1; X_2) = \log 3 - \frac{4}{3} \simeq 0.2516. \tag{74}$$

$$I(X_1, X_2; Y) = \log 3 \simeq 1.5850 \tag{75}$$

Since the system is symmetric with respect to the two encoders, the achievable rate region (for the symmetric case) is given by

$$R \leq \frac{1}{2}\left[I(X_1, X_2; Y) + I(X_1; X_2)\right] \simeq 0.9183, \tag{76}$$

$$R' \leq R - 0.2516. \tag{77}$$

Clearly $R = 0.9183$ and $R' = 0.6667$ is on the boundary of the given achievable rate region. It can be easily seen that $H(S) = H(T) = 0.9183$ and $H(S|T) = H(T|S) = 0.6667$. Thus, while Sender 1 sends at a rate $R_1 = 0.9183$, Sender 2 also can send at a rate $R_2 = 0.9183$, along as their messages are correlated. This means that all the typical sequences of S and T can be sent over the channel without any error. In other words, the channel code is the jointly strongly ϵ-typical set of (X_1, X_2) (with distribution as given above), and the message-graph is just a relabeled version of the typicality graph of the source (S, T). Hence there is a match between the two. Here $H(S,T) = I(X_1, X_2; Y)$.

6.4 An Interpretation of Cover, El Gamal and Salehi's Coding [10] with Graphs

For multiple access channels with correlated sources, Cover, El Gamal and Salehi [10] gave a coding theorem, by considering direct mapping of source symbols into channel input symbols. In this section, we interpret their coding scheme (denoted by *CES coding*) by using bipartite graphs. We show that CES coding can be interpreted as instances where typicality graph of the given correlated sources is a subset of the typicality graph of some channel input distribution. In other words matching of the message-graph with the graph associated with the channel code.

Following Section 2.3, first note that the channel input distribution must obey the Markov chain $X_1 \to S \to T \to X_2$ for successful transmission. In CES coding, for a given source sequence pair (s^n, t^n), channel inputs x_1^n and x_2^n are generated from $\prod_{i=1}^n p_1(x_{1i}|s_i)$ and $\prod_{i=1}^n p_2(x_{2i}|t_i)$, respectively. So, for each jointly typical sequence pair (s^n, t^n), a corresponding channel input pair (x_1^n, x_2^n) can be generated. In particular, this pair $(x_1^n, x_2^n) \in A_\epsilon^{(n)}(X_1, X_2)$ with high probability by the Markov Lemma [8, 28].

Roughly speaking, for every sequence $s^n \in A_\epsilon^{(n)}(S)$, one can obtain a sequence $x_1^n(s^n) \in A_\epsilon^{(n)}(X_1)$. Similarly for every sequence $v^n \in A_\epsilon^{(n)}(T)$, one can obtain $x_2^n(t^n) \in A_\epsilon^{(n)}(X_2)$. Further, for nearly every pair $(s^n, t^n) \in A_\epsilon^{(n)}(S,T)$, the corresponding pair $(x_1^n(s^n), x_2^n(t^n)) \in A_\epsilon^{(n)}(X_1, X_2)$. Hence the typicality graph of (X_1, X_2) nearly covers the typicality graph of (S,T). One can now imagine a discrete interface with a message-graph which is essentially a relabeled version of the typicality graph of (S,T). Now the sources can be distributively mapped to the edges in this graph. Nearly every edge in this graph can be reliably communicated to the joint receiver by transmitting channel input sequence pair that is jointly strongly ϵ-typical.

7 Conclusion

We have considered a multiterminal communication system with correlated information sources being transmitted reliably over a multiple access channel. We have considered bipartite undirected nearly semi-regular graphs as digital interfaces that can capture the correlation between sources. This leads to a modular architecture involving two components: a channel coding component and a source coding component. Correlated sources are first mapped into such graphs, and the edges coming from these graphs are reliably transmitted over a multiple access channel.

We have given a partial characterization of the set of all graphs that can be used to represent a given pair of correlated sources, and similarly given a partial characterization of the set of all graphs such that edges coming those graphs are reliably transmitted over a given multiple access channel. We have applied our analysis to two examples, one involving the Gaussian multiple access channel and the other involving binary-input multiple access channel to corroborate the claims made.

Appendix

A Detailed Explanation of Graphs in $\mathcal{K}_{4,2}$

Let us consider a graph in the subset $\mathbf{S}_1$ as shown in Figure 7. If we permute and relabel the left vertexes of the graph in the same way as explained previously, then we get $4! = 24$ distinct graphs belonging to a set $\mathbf{A}$ such that $\mathbf{A} \subset \mathbf{S}_1$. Now let us do this operation on the right vertexes. By changing the right vertexes 1 and 2 in the graph, we can get a graph in the subset $\mathbf{B}$ such that $\mathbf{A}$ and $\mathbf{B}$ are disjoint and $\mathbf{B} \subset \mathbf{S}_1$. Similarly by permuting and relabeling the left vertexes again, we can obtain all the distinct graphs in $\mathbf{B}$. By changing the right vertexes 2 and 3 in the graph, we can get another graph in the subset $\mathbf{C}$ such that $\mathbf{A}, \mathbf{B}$ and $\mathbf{C}$ are disjoint and $\mathbf{C} \subset \mathbf{S}_1$. So, we can obtain all the graphs in $\mathbf{S}_1$ in this way.

Up to now, we could generate a total of 72 distinct graphs in the set $\mathcal{K}_{4,2}$. Note that even after we permute the right and the left vertexes of the original graph in $\mathbf{S}_1$, we can not get the graph for which the edges belong to the set given by $\{(1,3),(2,4),(1,3),(2,4)\}$. This means that there are some graphs in $\mathcal{K}_{4,2}$ which can not be obtained from a graph in the subset $\mathbf{S}_1$ by just permutation and relabeling. Now consider the graph shown in the right side of Figure 7. This is one of the graph in the subset $\mathbf{S}_2$, for which the edges belong to the set given by $\{(1,3),(2,4),(1,3),(2,4)\}$. If we permute and relabel the left vertexes, then we can obtain $\frac{4!}{2!2!} = 6$ distinct graphs.

Similarly, we can obtain all the remaining distinct 18 graphs in $\mathcal{K}_{4,2}$ by changing the right vertexes of the graph. Hence we can obtain all the remaining 18 graphs in the subset $\mathbf{S}_2$ as well.

B Proof of Lemma 1

Let $\mathcal{Z} = \{(i,j) : (X_1^n(i), X_2^n(j)) \in A_\epsilon^{(n)}(X_1, X_2)\}$. Then,

$$|\mathcal{Z}| = \sum_{i=1}^{2^{nR_1}} \sum_{j=1}^{2^{nR_2}} \psi(i,j) \tag{78}$$

$$\text{where} \quad \psi(i,j) = \begin{cases} 1, & \text{if } (X_1^n(i), X_2^n(j)) \in A_\epsilon^{(n)}(X_1, X_2), \\ 0, & \text{otherwise.} \end{cases} \tag{79}$$

From the property of strongly jointly typical sequences given in (30),

$$E|\mathcal{Z}| = \sum_{i=1}^{2^{nR_1}} \sum_{j=1}^{2^{nR_2}} P\{(X_1^n(i), X_2^n(j)) \in A_\epsilon^{(n)}\} \tag{80}$$

$$\leq 2^{nR_1} 2^{nR_2} 2^{-n(I(X_1;X_2)-\delta_1(\epsilon))}, \tag{81}$$

where $P\{(X_1^n(i), X_2^n(j)) \in A_\epsilon^{(n)}\} \leq 2^{-n(I(X_1;X_2)-\delta_1(\epsilon))}$, and $\delta_1(\epsilon) \to 0$ as $\epsilon \to 0$. So, by applying the Markov's inequality,

$$P\{|\mathcal{Z}| > 2^{n(I(X_1,X_2;Y)-\delta(\epsilon)-\epsilon)}\} \leq \frac{E|\mathcal{Z}|}{2^{n(I(X_1,X_2;Y)-\delta(\epsilon)-\epsilon)}} \tag{82}$$

$$\leq 2^{n(R_1+R_2-I(X_1,X_2;Y)-I(X_1;X_2)+\delta(\epsilon)+\delta_1(\epsilon)+\epsilon)} \tag{83}$$

Thus, for sufficiently large n,

$$P(E_0) = P\{|\mathcal{Z}| > 2^{n(I(X_1,X_2;Y)-\delta(\epsilon)-\epsilon)}\} < \frac{\epsilon}{7} \tag{84}$$

provided $R_1 + R_2 < I(X_1, X_2; Y) + I(X_1; X_2) - \delta(\epsilon) - \delta_1(\epsilon) - \epsilon$. ■

C Proof of Lemma 2

Partition the codebook $\mathbb{C}_2$ uniformly into bins $B_2(i)$, $i = 1, 2, \ldots, 2^{n(R_2-I(X_1;X_2)-\delta(\epsilon)-\epsilon)}$ with the same size $2^{n(I(X_1;X_2)+\delta(\epsilon)+\epsilon)}$. Similarly, Partition the codebook $\mathbb{C}_1$ uniformly into bins $B_1(j)$, $j = 1, 2, \ldots, 2^{n(R_1-I(X_1;X_2)-\delta(\epsilon)-\epsilon)}$ with the same size $2^{n(I(X_1;X_2)+\delta(\epsilon)+\epsilon)}$.

The event $E_{0,1}$ can be considered as

$$E_{0,1} = \bigcup_{i=1}^{2^{nR_1}} E_{0,1}(i) \tag{85}$$

where $E_{0,1}(i)$ is the event that for a particular (random) $X_1^n(i) \in \mathbb{C}_1$, $\deg_{\mathbb{G},1}(X_1^n(i)) < 2^{n(R_2-I(X_1;X_2)-\delta(\epsilon)-\epsilon)}$. The event $E_{0,1}(i)$ can be expressed as

$$E_{0,1}(i) \subset \bigcup_{j=1}^{2^{n(R_2-I(X_1;X_2)-\delta(\epsilon)-\epsilon)}} E_{0,1}(i,j) \tag{86}$$

where $E_{0,1}(i,j)$ is the event that for a particular (random) $X_1^n(i) \in \mathbb{C}_1$ and $\forall X_2^n \in B_2(j)$, $(X_1^n(i), X_2^n) \notin A_\epsilon^{(n)}(X_1, X_2)$. So, by using the union bound, the probability of this event $P(E_{0,1}(i))$ can be bounded as:

$$P(E_{0,1}(i)) \leq P\left(\bigcup_{j=1}^{2^{n(R_2-I(X_1;X_2)-\delta(\epsilon)-\epsilon)}} E_{0,1}(i,j)\right) \tag{87}$$

$$\leq \sum_{j=1}^{2^{n(R_2-I(X_1;X_2)-\delta(\epsilon)-\epsilon)}} P(E_{0,1}(i,j)) \tag{88}$$

$$\overset{(a)}{\leq} 2^{n(R_2-I(X_1;X_2)-\delta(\epsilon)-\epsilon)}\left[1-2^{-n(I(X_1,X_2)+\delta(\epsilon))}\right]^{2^{n(I(X_1;X_2)+\delta(\epsilon)+\epsilon)}} \tag{89}$$

$$\overset{(b)}{\leq} 2^{n(R_2-I(X_1;X_2)-\delta(\epsilon)-\epsilon)}2^{-2^{n\epsilon}} \tag{90}$$

where

(a) is from the property of strongly jointly ϵ-typical sequences [8],

(b) follows from Lemma 13.5.3 in [8]: for $0 \leq x, y \leq 1$, $n > 0$, $(1-xy)^n \leq 1-x+2^{-yn}$.

Therefore, for sufficiently large n, the probability of the event $E_{0,1}$ can be bounded by applying the union bound:

$$P(E_{0,1}) \leq \sum_{i=1}^{2^{nR_1}} P(E_{0,1}(i)) \tag{91}$$

$$\leq 2^{nR_1}2^{n(R_2-I(X_1;X_2)-\delta(\epsilon)-\epsilon)}2^{-2^{n\epsilon}} \tag{92}$$

$$= 2^{n(R_1+R_2-I(X_1;X_2)-\delta(\epsilon)-\epsilon)-2^{n\epsilon}} \tag{93}$$

$$\overset{(c)}{\leq} \frac{\epsilon}{14} \tag{94}$$

where (c) is from the fact that $n(R_1+R_2-I(X_1;X_2)-\delta(\epsilon)-\epsilon)$ is polynomially increasing but $2^{n\epsilon}$ is exponentially increasing as n increases.

In a similar way, we can also show that $P(E_{0,2}) < \frac{\epsilon}{14}$ for sufficiently large n. ∎

D Proof of Lemma 3

The event $E_{0,1}^*$ can be expressed as

$$E_{0,1}^* = \bigcup_{i=1}^{2^{nR_1}} E_{0,1}^*(i) \tag{95}$$

where $E_{0,1}^*(i)$ is the event that for a particular (random) $X_1^n(i) \in \mathbb{C}_1$, $\deg_{\mathbb{G},1}(X_1^n(i)) > 2^{n(R_2-I(X_1;X_2)+\delta(\epsilon)+\epsilon)}$.

For a particular (random) $X_1^n(i) \in \mathbb{C}_1$, let the random set $\mathcal{S} = \{X_2^n \in \mathbb{C}_2 : (X_1^n(i), X_2^n) \in A_\epsilon^{(n)}(X_1, X_2)\}$. Then,

$$|\mathcal{S}| = \sum_{j=1}^{2^{nR_2}} \psi(j), \text{ where } \psi(j) = \begin{cases} 1, & \text{if } (X_1^n(i), X_2^n(j)) \in A_\epsilon^{(n)}(X_1, X_2), \\ 0, & \text{otherwise.} \end{cases} \tag{96}$$

Then,

$$P\{E_{0,1}^*(i)\} = P\{|\mathcal{S}| > \underbrace{2^{n(R_2-I(X_1;X_2)+\delta(\epsilon)+\epsilon)}}_{\triangleq a}\} \tag{97}$$

$$\overset{(a)}{<} e^{-at}E\{e^{t|\mathcal{S}|}\} \tag{98}$$

for any $t > 0$ where (a) follows from the Chernoff bound [8]. Now we calculate an upper bound for $P\{E^*_{0,1}(i)\}$. Let $x_1^n[l]$ be the l-th sequence (using some ordering) in $A_\epsilon^{(n)}(X_1)$.

$$E\{e^{t|\mathcal{S}|}\} = E\bigg\{\exp\Big(t\sum_{j=1}^{2^{nR_2}}\psi(j)\Big)\bigg\} \tag{99}$$

$$= E\bigg\{\prod_{j=1}^{2^{nR_2}} e^{t\psi(j)}\bigg\} \tag{100}$$

$$= \sum_{l=1}^{|A_\epsilon^{(n)}(X_1)|} Pr\{X_1^n(i) = x_1^n[l]\}E\bigg\{\prod_{j=1}^{2^{nR_2}} e^{t\psi(j)}\bigg|X_1^n(i) = x_1^n[l]\bigg\} \tag{101}$$

$$\overset{(*)}{=} \sum_{l=1}^{|A_\epsilon^{(n)}(X_1)|} \frac{1}{|A_\epsilon^{(n)}(X_1)|}\prod_{j=1}^{2^{nR_2}} E\bigg\{e^{t\psi(j)}\bigg|X_1^n(i) = x_1^n[l]\bigg\} \tag{102}$$

where $(*)$ is from the fact that $\psi(j)$'s are independent when the outcome of $X_1^n(i)$ is fixed.

Let us denote $p_j = P\{\psi(j) = 1|X_1^n(i) = x_1^n[l]\}$. Then,

$$E\{e^{t\psi(j)}|X_1^n(i) = x_1^n[l]\} = e^t p_j + 1\cdot(1-p_j) \tag{103}$$

$$= 1 - p_j(1-e^t) \tag{104}$$

$$\leq e^{-p_j(1-e^t)} \quad \text{since } 1-x \leq e^{-x}. \tag{105}$$

So,

$$\prod_{j=1}^{2^{nR_2}} E\{e^{t\psi(j)}|X_1^n(i) = x_1^n[l]\} \leq \prod_{j=1}^{2^{nR_2}} e^{p_j(e^t-1)} \tag{106}$$

$$= \exp\bigg\{(e^t-1)\sum_{j=1}^{2^{nR_2}} p_j\bigg\} \tag{107}$$

$$= \exp\bigg\{(e^t-1)E\left\{|\mathcal{S}|\big|X_1^n(i) = x_1^n[l]\right\}\bigg\} \tag{108}$$

Then,

$$E\{e^{t|\mathcal{S}|}\} \leq \sum_{l=1}^{|A_\epsilon^{(n)}(X_1)|} \frac{1}{|A_\epsilon^{(n)}(X_1)|}\exp\bigg\{(e^t-1)E\left\{|\mathcal{S}|\big|X_1^n(i) = x_1^n[l]\right\}\bigg\} \tag{109}$$

$$\overset{(a)}{\leq} \exp\bigg\{(e^t-1)2^{n(R_2 - I(X_1;X_2)+\delta(\epsilon))}\bigg\} \tag{110}$$

where (a) follows from the following inequality:

$$E\left\{|\mathcal{S}|\big|X_1^n(i) = x_1^n[l]\right\} = \sum_{j=1}^{2^{nR_2}} P\{\psi(j) = 1|X_1^n(i) = x_1^n[l]\} \leq 2^{nR_2}2^{-n(I(X_1;X_2)-\delta(\epsilon))} \leq 2^{n(R_2-I(X_1;X_2)+\delta(\epsilon))}. \tag{111}$$

Therefore, for $t > 0$,

$$P\{E^*_{0,1}(i)\} \leq e^{-at}\exp\bigg\{(e^t-1)2^{n(R_2-I(X_1;X_2)+\delta(\epsilon))}\bigg\} \tag{112}$$

$$= \exp\Big\{ -at + (e^t - 1)\underbrace{2^{n(R_2 - I(X_1;X_2)+\delta(\epsilon))}}_{\triangleq\, b} \Big\} \tag{113}$$

To get a tighter upper bound, let us denote $f(t) = -at + b(e^t - 1)$, for $t > 0$. Then, $f'(t) = -a + be^t$ and $f''(t) = be^t > 0$. So, $f(t)$ has the minimum value when $t = \ln\left(\frac{a}{b}\right)$.

Thus, $P\{E^*_{0,1}(i)\}$ is bounded as

$$P\{E^*_{0,1}(i)\} \leq \exp\Big\{ -a\ln\left(\frac{a}{b}\right) + a - b \Big\}. \tag{114}$$

Note that $a = 2^{n(R_2 - I(X_1;X_2)+\delta(\epsilon)+\epsilon)}$ and $b = 2^{n(R_2 - I(X_1;X_2)+\delta(\epsilon))}$.

So,

$$P\{E^*_{0,1}(i)\} \leq \exp\Big\{ -2^{n(R_2 - I(X_1;X_2)+\delta(\epsilon))}\left[2^{n\epsilon}\ln(2^{n\epsilon}) - 2^{n\epsilon} + 1\right] \Big\} \tag{115}$$

$$= \exp\Big\{ -2^{n(R_2 - I(X_1;X_2)+\delta(\epsilon))}\eta \Big\} \tag{116}$$

where $\eta = 2^{n\epsilon}\ln(2^{n\epsilon}) - 2^{n\epsilon} + 1$. Here, $\eta > 0$ since $f(x) = x\ln(x) - x + 1$ is increasing function of x and $f(x) > 0$ for $x > 1$.

Therefore, for sufficiently large n, by applying the union bound,

$$P\{E^*_{0,1}\} = P\Big\{ \bigcup_{i=1}^{2^{nR_1}} E^*_{0,1}(i) \Big\} \tag{117}$$

$$\leq \sum_{i=1}^{2^{nR_1}} P\{E^*_{0,1}(i)\} \tag{118}$$

$$\leq 2^{nR_1}\exp\big\{ -2^{n(R_2 - I(X_1;X_2)+\delta(\epsilon))}\eta \big\} \tag{119}$$

$$= \exp\big\{ nR_1\ln 2 - 2^{n(R_2 - I(X_1;X_2)+\delta(\epsilon))}\eta \big\} \tag{120}$$

$$\overset{(a)}{<} \frac{\epsilon}{14} \tag{121}$$

provided $R_2 > I(X_1;X_2) - \delta(\epsilon)$, where (a) is from the fact that $nR_1\ln 2$ is linearly increasing but $2^{n(R_2 - I(X_1;X_2)+\delta(\epsilon))}$ is exponentially increasing as n increases.

In a similar way, we can also show that $P\{E^*_{0,2}\} < \frac{\epsilon}{14}$ for sufficiently large n. ■

E Proof of Lemma 4

Now let us calculate the probability $P(E \cap E_0^c \cap E_1^c \cap E_2^c)$. Without loss of generality, let us select an edge, $k \in E(\mathbb{G})$, assuming that the corresponding message pair is transmitted. In other words, a codeword pair $(X_1^n(i_k), X_2^n(j_k))$ is transmitted. For a fixed k, consider the following error events;

E_3 : $(X_1^n(i_k), X_2^n(j_k), Y^n) \notin A_\epsilon^{(n)}(X_1, X_2, Y)$,

E_4 : $\exists j' \neq j_k$ such that $(X_1^n(i_k), X_2^n(j'), Y^n) \in A_\epsilon^{(n)}(X_1, X_2, Y)$,

E_5 : $\exists i' \neq i_k$ such that $(X_1^n(i'), X_2^n(j_k), Y^n) \in A_\epsilon^{(n)}(X_1, X_2, Y)$,

E_6 : $\exists i' \neq i_k, j' \neq j_k$ such that $(X_1^n(i'), X_2^n(j'), Y^n) \in A_\epsilon^{(n)}(X_1, X_2, Y)$.

$$P(E \cap E_0^c \cap E_1^c \cap E_2^c) = P\left(\cup_{i=3}^6 E_i \cap E_0^c \cap E_1^c \cap E_2^c\right) \tag{122}$$
$$\leq \sum_{i=3}^{6} P\left(E_i \cap E_0^c \cap E_1^c \cap E_2^c\right) \tag{123}$$

By the property of jointly typical sequences [8], $P(E_3 \cap E_0^c \cap E_1^c \cap E_2^c) < \frac{\epsilon}{7}$ for sufficiently large n. For any $x_1^n \in A_\epsilon^{(n)}(X_1)$, let $A_\epsilon^{(n)}(X_2|x_1^n)$ denote the set of sequences in $\mathcal{X}_2^n$ that are jointly ϵ-typical with x_1^n. Now

$$P(E_4 \cap E_0^c \cap E_1^c \cap E_2^c) \leq \sum_{j' \neq j_k} P((X_1^n(i_k), X_2^n(j'), Y^n) \in A_\epsilon^{(n)}) \tag{124}$$
$$= \sum_{j' \neq j_k} \sum_{(x_1^n, x_2^n, y^n) \in A_\epsilon^{(n)}(X_1, X_2, Y)} |A_\epsilon^{(n)}(X_2)|^{-1} |A_\epsilon^{(n)}(X_1)|^{-1} Pr(y^n|x_1^n) \tag{125}$$
$$= \sum_{j' \neq j_k} \sum_{(x_1^n, x_2^n, y^n) \in A_\epsilon^{(n)}(X_1, X_2, Y)} \frac{|A_\epsilon^{(n)}(X_2)|^{-1} |A_\epsilon^{(n)}(X_1)|^{-1}}{|A_\epsilon^{(n)}(X_2|x_1^n)|} \sum_{x_2^n \in A_\epsilon^{(n)}(X_2|x_1^n)} p^n_{Y|X_1,X_2}(y^n|x_1^n, x_2^n) \tag{126}$$
$$\leq \sum_{j' \neq j_k} \sum_{(x_1^n, x_2^n, y^n) \in A_\epsilon^{(n)}(X_1, X_2, Y)} |A_\epsilon^{(n)}(X_2)|^{-1} |A_\epsilon^{(n)}(X_1)|^{-1} \sum_{x_2^n \in A_\epsilon^{(n)}(X_2|x_1^n)} 2^{2n\epsilon_1} p^n_{Y,X_2|X_1}(y^n, x_2^n|x_1^n) \tag{127}$$
$$\leq \sum_{j' \neq j_k} \sum_{(x_1^n, x_2^n, y^n) \in A_\epsilon^{(n)}(X_1, X_2, Y)} |A_\epsilon^{(n)}(X_2)|^{-1} |A_\epsilon^{(n)}(X_1)|^{-1} p^n_{Y|X_1}(y^n|x_1^n) 2^{2n\epsilon_1} \tag{128}$$
$$\leq 2^{nR_2} |A_\epsilon^{(n)}(X_1, X_2, Y)| 2^{-n(H(X_2)-\epsilon_1)} 2^{-n(H(X_1)-\epsilon_1)} 2^{-n(H(Y|X_1)-2\epsilon_1)} 2^{2n\epsilon_1} \tag{129}$$
$$\leq 2^{nR_2} 2^{-n(H(X_2)+H(X_1,Y)-H(X_1,X_2,Y)-7\epsilon_1)} \tag{130}$$
$$= 2^{n(R_2 - I(X_2;X_1,Y)+7\epsilon_1)} \tag{131}$$

where we have used the fact that $\forall x_2^n \in A_\epsilon^{(n)}(X_2|x_1^n)$, we have

$$|A_\epsilon^{(n)}(X_2|x_1^n)|^{-1} \leq 2^{2n\epsilon_1} p^n_{X_2|X_1}(x_2^n|x_1^n), \tag{132}$$

and $\epsilon_1(\epsilon) \to 0$ as $\epsilon \to 0$. Similarly,

$$P(E_5 \cap E_0^c \cap E_1^c \cap E_2^c) \leq 2^{n(R_1 - I(X_1;X_2,Y)+7\epsilon_1)}, \tag{133}$$

$$P(E_6 \cap E_0^c \cap E_1^c \cap E_2^c) = \sum_{i' \neq i_k} \sum_{j' \neq j_k} P((X_1^n(i'), X_2^n(j'), Y^n) \in A_\epsilon^{(n)}) \tag{134}$$
$$\leq 2^{nR_1} 2^{nR_2} |A_\epsilon^{(n)}| 2^{-n(H(X_1)-\epsilon_1)} 2^{-n(H(X_2)-\epsilon_1)} 2^{-n(H(Y)-3\epsilon_1)} \tag{135}$$
$$\leq 2^{n(R_1+R_2)} 2^{-n(H(X_1)+H(X_2)+H(Y)-H(X_1,X_2,Y)-6\epsilon_1)} \tag{136}$$
$$= 2^{n(R_1+R_2-I(X_1,X_2;Y)-I(X_1;X_2)+6\epsilon_1)} \tag{137}$$

So,

$$P(E_4 \cap E_0^c \cap E_1^c \cap E_2^c) \leq 2^{n(R_2 - I(X_1,Y;X_2)+7\epsilon_1)} < \frac{\epsilon}{7}, \tag{138}$$

$$P(E_5 \cap E_0^c \cap E_1^c \cap E_2^c) \leq 2^{n(R_1 - I(X_2,Y;X_1)+7\epsilon_1)} < \frac{\epsilon}{7}, \tag{139}$$

$$P(E_6 \cap E_0^c \cap E_1^c \cap E_2^c) \leq 2^{n(R_1+R_2 - I(X_1,X_2;Y)-I(X_1;X_2)+7\epsilon_1)} < \frac{\epsilon}{7}, \tag{140}$$

for sufficiently large n, if R_1 and R_2 satisfies the conditions given in (37), (38) and (39). Hence, the proof of Lemma 4 has been completed. ■

F Proof of Theorem 2

Since the message pairs are equally likely, and because of the conditions imposed on the degrees of the vertexes of the message-graph, we have,

$$n(R_1 + R_2') \leq H(W_1, W_2) + \log\mu \tag{141}$$

$$= I(W_1, W_2; Y^n) + H(W_1, W_2|Y^n) + \log\mu \tag{142}$$

$$\overset{(a)}{\leq} I(W_1, W_2; Y^n) + n\epsilon_n + \log\mu \tag{143}$$

$$\overset{(b)}{\leq} I(X_1^n(W_1), X_2^n(W_2); Y^n) + n\epsilon_n + \log\mu \tag{144}$$

$$= H(Y^n) - H(Y^n|X_1^n(W_1), X_2^n(W_2)) + n\epsilon_n + \log\mu \tag{145}$$

$$\overset{(c)}{=} H(Y^n) - \sum_{i=1}^{n} H(Y_i|Y^{i-1}, X_1^n(W_1), X_2^n(W_2)) + n\epsilon_n + \log\mu \tag{146}$$

$$\overset{(d)}{=} H(Y^n) - \sum_{i=1}^{n} H(Y_i|X_{1i}, X_{2i}) + n\epsilon_n + \log\mu \tag{147}$$

$$\overset{(e)}{\leq} \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|X_{1i}, X_{2i}) + n\epsilon_n + \log\mu \tag{148}$$

$$= \sum_{i=1}^{n} I(X_{1i}, X_{2i}; Y_i) + n\epsilon_n + \log\mu, \tag{149}$$

where

(a) follows from Fano's inequality,

(b) from the data processing inequality,

(c) from the chain rule,

(d) from the fact that Y_i depends only on X_{1i} and X_{2i} and is conditionally independent of everything else, and

(e) is obtained from the chain rule and removing conditioning.

Hence we have

$$R_1 + R_2' \leq \frac{1}{n}\sum_{i=1}^{n} I(X_{1i}, X_{2i}; Y_i) + \epsilon_n + \frac{1}{n}\log\mu \tag{150}$$

$$= \frac{1}{n}\sum_{q=1}^{n} I(X_{1q}, X_{2q}; Y_q|Q=q) + \epsilon_n + \frac{1}{n}\log\mu \tag{151}$$

$$= I(X_{1Q}, X_{2Q}; Y_Q) + \epsilon_n + \frac{1}{n}\log\mu \tag{152}$$

$$= I(X_1, X_2; Y) + \epsilon_n + \frac{1}{n}\log\mu, \tag{153}$$

where $Q = i \in \{1, 2, \ldots, n\}$ with probability $\frac{1}{n}$, and $X_1 \triangleq X_{1Q}$, $X_2 \triangleq X_{2Q}$, $Y \triangleq Y_{1Q}$ are new random variables whose distribution depend on Q.

So, for sufficiently large n,

$$R_1 + R_2' \leq I(X_1, X_2; Y) \tag{154}$$

for some distribution $p(q)p(x_1, x_2|q)p(y|x_1, x_2)$. Similarly, we can prove

$$R_1' + R_2 \leq I(X_1, X_2; Y) \tag{155}$$

for some distribution $p(q)p(x_1, x_2|q)p(y|x_1, x_2)$. ■

G Proof of Lemma 6

The event $E_{0,1}$ can be considered as

$$E_{0,1} = \bigcup_{i=1}^{2^{nR_1}} E_{0,1}(i) \tag{156}$$

where $E_{0,1}(i)$ is the event that for a particular $i \in V_1(G)$, $\deg_{G,1}(i) < 2^{n(H(S,T)-R_1-\epsilon')}$.

Let us define super-bins $\tilde{B}_p$ and $\tilde{C}_q$ for $p = \{1, 2, \ldots, 2^{n(H(S,T)-R_2-\epsilon')}\}$ and $q = \{1, 2, \ldots, 2^{n(H(S,T)-R_1-\epsilon')}\}$, which is a union of $2^{n(R_2+R_1-H(S,T)+\epsilon')}$ consecutive B_i and C_j bins, respectively. The size of each super-bin $\tilde{B}_p$ and $\tilde{C}_q$ is $2^{n(R_2-H(T|S)+\gamma+\epsilon')}$ and $2^{n(R_1-H(S|T)+\gamma+\epsilon')}$, respectively.

Then, the event $E_{0,1}(i)$ can be expressed as

$$E_{0,1}(i) \subset \bigcup_{q=1}^{2^{n(H(S,T)-R_1-\epsilon')}} E_{0,1}(i, q) \tag{157}$$

where $E_{0,1}(i, q)$ is the event that $\forall S^n \in B_i$ and $\forall T^n \in \tilde{C}_q$, $(S^n, T^n) \notin A_\epsilon^{(n)}(S, T)$. So, by using the union bound the probability of this event $P(E_{0,1}(i))$ can be bounded as:

$$P(E_{0,1}(i)) \leq P\left(\bigcup_{q=1}^{2^{n(H(S,T)-R_1-\epsilon')}} E_{0,1}(i, q)\right) \tag{158}$$

$$\leq \sum_{q=1}^{2^{n(H(S,T)-R_1-\epsilon')}} P(E_{0,1}(i, q)) \tag{159}$$

$$\overset{(a)}{\leq} 2^{n(H(S,T)-R_1-\epsilon')} 2^{-nM} \tag{160}$$

(a) is from the Lemma 5, and from the fact $M > 0$ is a sufficiently large number satisfying $M > H(S, T) - \epsilon'$. This is true because $\frac{1}{n} \log |B_i||\tilde{C}_q| = I(S; T) + 2\gamma + \epsilon'$.

Therefore, for sufficiently large n, the probability of the event $E_{0,1}$ can be bounded by applying the union bound:

$$P(E_{0,1}) \leq \sum_{i=1}^{2^{nR_1}} P(E_{0,1}(i)) \tag{161}$$

$$\leq 2^{nR_1} 2^{n(H(S,T)-R_1-\epsilon')} 2^{-nM} \tag{162}$$

$$= 2^{n(H(S,T)-\epsilon'-M)} \tag{163}$$

$$\stackrel{(b)}{\leq} \frac{\epsilon}{12}. \tag{164}$$

In a similar way, we can also show that $P(E_{0,2}) < \frac{\epsilon}{12}$ for sufficiently large n. ■

H Proof of Lemma 7

The event $E^*_{0,1}$ can be expressed as

$$E^*_{0,1} = \bigcup_{i=1}^{2^{nR_1}} E^*_{0,1}(i) \tag{165}$$

where $E^*_{0,1}(i)$ is the event that for a particular $i \in V_1(G)$, $\deg_{G,1}(i) > 2^{n(H(S,T)-R_1+\epsilon')}$.

For a particular random bin B_i, let the random set $\mathcal{D} = \{j : \exists (S^n, T^n) \in A_\epsilon^{(n)}(S,T) \cap (B_i \times C_j)\}$. Then,

$$|\mathcal{D}| = \sum_{j=1}^{2^{nR_2}} \psi(j), \text{ where } \psi(j) = \begin{cases} 1, & \text{if } \exists (S^n, T^n) \in A_\epsilon^{(n)}(S,T) \cap (B_i \times C_j), \\ 0, & \text{otherwise.} \end{cases} \tag{166}$$

In particular,

$$P\{\psi(j) = 1\} = P\{\ |\{(S^n, T^n) : (S^n, T^n) \in A_\epsilon^{(n)}(S,T) \cap (B_i \times C_j)\}| \neq 0\} \tag{167}$$

$$\stackrel{(c)}{\leq} 2^{n(H(S)-R_1+\gamma)} 2^{n(H(T)-R_2+\gamma)} 2^{-n(I(S;T)-3\epsilon_1)} \tag{168}$$

$$\leq 2^{n(H(S,T)-R_1-R_2+2\gamma+3\epsilon_1)} \tag{169}$$

where (c) is obtained by applying the union bound, and from the property of strongly jointly ϵ-typical sequences [8]: for a randomly and independently chosen $S^n \in A_\epsilon^{(n)}(S)$ and $T^n \in A_\epsilon^{(n)}(T)$, for sufficiently large n, the probability that $(S^n, T^n) \in A_\epsilon^{(n)}(S,T)$ is bounded by

$$2^{-n(I(S;T)+3\epsilon_1)} \leq P\{(S^n, T^n) \in A_\epsilon^{(n)}\} \leq 2^{-n(I(S;T)-3\epsilon_1)} \tag{170}$$

where $\epsilon_1 \to 0$ as $\epsilon \to 0$, since $2^{n(H(S,T)-\epsilon_1)} \leq |A_\epsilon^{(n)}(S,T)| \leq 2^{n(H(S,T)+\epsilon_1)}$, $2^{n(H(S)-\epsilon_1)} \leq |A_\epsilon^{(n)}(S)| \leq 2^{n(H(S)+\epsilon_1)}$, and $2^{n(H(T)-\epsilon_1)} \leq |A_\epsilon^{(n)}(T)| \leq 2^{n(H(T)+\epsilon_1)}$.

So, the expectation of $|\mathcal{D}|$ can be bounded as follows.

$$E|\mathcal{D}| = \sum_{j=1}^{2^{nR_2}} P\{\psi(j) = 1\} \tag{171}$$

$$\leq 2^{nR_2} 2^{n(H(S,T)-R_1-R_2+2\gamma+3\epsilon_1)} \tag{172}$$

$$\leq 2^{n(H(S,T)-R_1+2\gamma+3\epsilon_1)} \tag{173}$$

Now,

$$P\{E^*_{0,1}(i)\} = P\{|\mathcal{D}| > \underbrace{2^{n(H(S,T)-R_1+\epsilon')}}_{\triangleq a}\} \tag{174}$$

$$\stackrel{(a)}{<} e^{-at} E\{e^{t|\mathcal{D}|}\} \tag{175}$$

for any $t > 0$ where (a) follows from the Chernoff bound [8].

Now we calculate an upper bound of $P\{E^*_{0,1}(i)\}$. Since $|B_i| = \alpha$, let us denote the sequences in B_i by $S^n(i_1)$, $S^n(i_2)$, ..., $S^n(i_\alpha)$. Also, let $s^n[l]$ be the l-th sequence (using some ordering) in $A^{(n)}_\epsilon(S)$.

$$E\{e^{t|\mathcal{D}|}\} = E\Big\{\exp\Big(t\sum_{j=1}^{2^{nR_2}}\psi(j)\Big)\Big\} \tag{176}$$

$$= E\Big\{\prod_{j=1}^{2^{nR_2}} e^{t\psi(j)}\Big\} \tag{177}$$

$$= \sum_{l_1=1}^{|A^{(n)}_\epsilon(S)|} p\{S^n(i_1) = s^n[l_1]\} \sum_{l_2=1}^{|A^{(n)}_\epsilon(S)|} p\{S^n(i_2) = s^n[l_2]\} \ldots \sum_{l_\alpha=1}^{|A^{(n)}_\epsilon(S)|} p\{S^n(i_\alpha) = s^n[l_\alpha]\}$$
$$E\Big\{\prod_{j=1}^{2^{nR_2}} e^{t\psi(j)}\Big|\ S^n(i_\theta) = s^n[l_\theta], \text{ for } \theta = 1, 2, \ldots, \alpha\ \Big\} \tag{178}$$

$$= \sum_{l_1=1}^{|A^{(n)}_\epsilon(S)|}\sum_{l_2=1}^{|A^{(n)}_\epsilon(S)|}\cdots\sum_{l_\alpha=1}^{|A^{(n)}_\epsilon(S)|} \frac{1}{|A^{(n)}_\epsilon(S)|^\alpha}$$
$$E\Big\{\prod_{j=1}^{2^{nR_2}} e^{t\psi(j)}\Big|\ S^n(i_\theta) = s^n[l_\theta], \text{ for } \theta = 1, 2, \ldots, \alpha\ \Big\} \tag{179}$$

$$\stackrel{(*)}{=} \sum_{l_1=1}^{|A^{(n)}_\epsilon(S)|}\sum_{l_2=1}^{|A^{(n)}_\epsilon(S)|}\cdots\sum_{l_\alpha=1}^{|A^{(n)}_\epsilon(S)|} \frac{1}{|A^{(n)}_\epsilon(S)|^\alpha}$$
$$\prod_{j=1}^{2^{nR_2}} E\Big\{e^{t\psi(j)}\Big|\ S^n(i_\theta) = s^n[l_\theta], \text{ for } \theta = 1, 2, \ldots, \alpha\ \Big\} \tag{180}$$

where $(*)$ is from the fact that $\psi(j)$'s are independent when the outcomes of $S^n(i_1), S^n(i_2), \ldots, S^n(i_\alpha)$ are fixed. Let us denote $p_j = P\{\psi(j) = 1\big|\ S^n(i_\theta) = s^n[l_\theta], \text{ for } \theta = 1, 2, \ldots, \alpha\ \}$.

Then,

$$E\{e^{t\psi(j)}\big|\ S^n(i_\theta) = s^n[l_\theta], \text{ for } \theta = 1, 2, \ldots, \alpha\ \} = e^t p_j + 1 \cdot (1 - p_j) \tag{181}$$
$$= 1 - p_j(1 - e^t) \tag{182}$$
$$\le e^{-p_j(1-e^t)} \quad \text{since } 1 - x \le e^{-x}. \tag{183}$$

So,

$$\prod_{j=1}^{2^{nR_2}} E\{e^{t\psi(j)}|\ S^n(i_\theta) = s^n[l_\theta], \text{ for } \theta = 1, 2, \ldots, \alpha\ \} \le \prod_{j=1}^{2^{nR_2}} e^{p_j(e^t-1)} \tag{184}$$
$$= \exp\Big\{(e^t - 1)\sum_{j=1}^{2^{nR_2}} p_j\Big\} \tag{185}$$
$$= \exp\Big\{(e^t - 1)E\left\{|\mathcal{D}|\big|\ S^n(i_\theta) = s^n[l_\theta], \text{ for } \theta = 1, 2, \ldots, \alpha\ \right\}\Big\} \tag{186}$$

Then,

$$E\{e^{t|\mathcal{D}|}\} \leq \sum_{l_1=1}^{|A_\epsilon^{(n)}(S)|} \cdots \sum_{l_\alpha=1}^{|A_\epsilon^{(n)}(S)|} \frac{1}{|A_\epsilon^{(n)}(S)|^\alpha} \exp\Big\{(e^t-1)E\left\{|\mathcal{D}|\big|S^n(i_\theta)=s^n[l_\theta], \text{ for } \theta=1,2,\ldots,\alpha\ \right\}\Big\} \tag{187}$$

$$\overset{(a)}{\leq} \exp\Big\{(e^t-1)2^{n(H(S,T)-R_1+2\gamma+3\epsilon_1)}\Big\} \tag{188}$$

where (a) is obtained because $E\left\{|\mathcal{D}|\big|\ S^n(i_\theta)=s^n[l_\theta], \text{ for } \theta=1,2,\ldots,\alpha\ \right\}$ is bounded by the inequality (173) regardless of the particular sequences $s^n[l_1], s^n[l_2], \ldots, s^n[l_\alpha]$.

Therefore, for $t>0$,

$$P\{E_{0,1}^*(i)\} \leq e^{-at}\exp\Big\{(e^t-1)2^{n(H(S,T)-R_1+2\gamma+3\epsilon_1)}\Big\} \tag{189}$$

$$= \exp\Big\{-at+(e^t-1)\underbrace{2^{n(H(S,T)-R_1+2\gamma+3\epsilon_1)}}_{\triangleq\ b}\Big\} \tag{190}$$

To get a tighter upper bound, let us denote $f(t)=-at+b(e^t-1)$, for $t>0$. As mentioned in the proof of Lemma 3, $f(t)$ has the minimum value when $t=\ln\left(\frac{a}{b}\right)$.

Thus, $P\{E_{0,1}^*(i)\}$ is bounded as

$$P\{E_{0,1}^*(i)\} \leq \exp\Big\{-a\ln\left(\frac{a}{b}\right)+a-b\Big\} \tag{191}$$

where $a=2^{n(H(S,T)-R_1+\epsilon')}$ and $b=2^{n(H(S,T)-R_1+2\gamma+3\epsilon_1)}$.

So,

$$P\{E_{0,1}^*(i)\} \leq \exp\left\{-2^{n(H(S,T)-R_1+2\gamma+3\epsilon_1)}[2^{n(\epsilon'-2\gamma-3\epsilon_1)}\ln(2^{n(\epsilon'-2\gamma-3\epsilon_1)})-2^{n(\epsilon'-2\gamma-3\epsilon_1)}+1]\right\} \tag{192}$$

$$= \exp\big\{-2^{n(H(S,T)-R_1+2\gamma+3\epsilon_1)}\eta\big\} \tag{193}$$

where $\eta = 2^{n(\epsilon'-2\gamma-3\epsilon_1)}\ln(2^{n(\epsilon'-2\gamma-3\epsilon_1)})-2^{n(\epsilon'-2\gamma-3\epsilon_1)}+1$. As before, $\eta>0$ since $\epsilon'>2\gamma+3\epsilon_1$ and $f(x)=x\ln(x)-x+1$ is increasing function of x and $f(x)>0$ for $x>1$.

Therefore, for sufficiently large n, by applying the union bound,

$$P\{E_{0,1}^*\} = P\Big\{\bigcup_{i=1}^{2^{nR_1}} E_{0,1}^*(i)\Big\} \tag{194}$$

$$\leq \sum_{i=1}^{2^{nR_1}} P\{E_{0,1}^*(i)\} \tag{195}$$

$$\leq 2^{nR_1}\exp\big\{-2^{n(H(S,T)-R_1+2\gamma+3\epsilon_1)}\eta\big\} \tag{196}$$

$$= \exp\big\{nR_1\ln 2-2^{n(H(S,T)-R_1+2\gamma+3\epsilon_1)}\eta\big\} \tag{197}$$

$$\overset{(a)}{<} \frac{\epsilon}{12} \tag{198}$$

provided $R_1 < H(S,T)+2\gamma+3\epsilon_1$, where (a) is from the fact that $nR_1\ln 2$ is linearly increasing but $2^{n(H(S,T)-R_1+2\gamma+3\epsilon_1)}$ is exponentially increasing as n increases.

In a similar way, we can also show that $P\{E_{0,2}^*\} < \frac{\epsilon}{12}$ for sufficiently large n. ■

I Proof of Lemma 8

Now let us calculate the probability $P(E \cap E_1^c \cap E_2^c)$. If previous error events do not occur, we define other error events as follows.

E_3 : $(S^n, T^n) \notin A_\epsilon^{(n)}$,

E_4 : $\exists S^n \in A_\epsilon^{(n)}(S)$ such that $f_1(S^n) = 0$,

E_5 : $\exists T^n \in A_\epsilon^{(n)}(T)$ such that $f_2(T^n) = 0$,

E_6 : $\exists(\bar{S}^n, \bar{T}^n) \in A_\epsilon^{(n)}(S,T) \cap (B_{f_1(S^n)} \times C_{f_2(T^n)})$ such that $(\bar{S}^n, \bar{T}^n) \neq (S^n, T^n)$.

Then,

$$P(E \cap E_1^c \cap E_2^c) = P\left(\cup_{i=3}^{6} E_i \cap E_1^c \cap E_2^c\right) \tag{199}$$

$$\leq \sum_{i=3}^{6} P\left(E_i \cap E_1^c \cap E_2^c\right) \tag{200}$$

By the property of jointly typical sequences [8], $P(E_3 \cap E_1^c \cap E_2^c) < \frac{\epsilon}{6}$ for sufficiently large n.

Following [28], for sufficiently large n

$$P(E_4 \cap E_1^c \cap E_2^c) < \frac{\epsilon}{6} \tag{201}$$

$$P(E_5 \cap E_1^c \cap E_2^c) < \frac{\epsilon}{6} \tag{202}$$

if $\gamma > \epsilon_1$.

Also, for sufficiently large n, by following [28] it can be shown that

$$P(E_6 \cap E_1^c \cap E_2^c) < \frac{\epsilon}{6} \tag{203}$$

provided $R_1 + R_2 > H(S,T) + 2\gamma + 3\epsilon_1$.

Therefore,

$$P(E) \leq P(E_1) + P(E_2) + P(E \cap E_0^c \cap E_1^c \cap E_2^c) < \epsilon. \tag{204}$$

In addition, by combining the conditions $\alpha \geq 1$, $\beta \geq 1$, and $R_1 + R_2 > H(S,T)$, it is easy to show that the following inequalities need to hold.

$$H(S|T) \leq R_1, \quad \text{and} \quad H(T|S) \leq R_2. \tag{205}$$

■

J Proof of converse part of Theorem 3

Let f_1, f_2, g be fixed. Also, let $f_1(S^n) = W_1,\ f_2(T^n) = W_2$. Then, from the Fano's inequality, we can have

$$H(S^n, T^n|W_1, W_2) \le P_e n(\log|\mathcal{S}||\mathcal{T}|) + 1 \le n\epsilon_n, \tag{206}$$

$$H(S^n|T^n, W_1, W_2) \le H(S^n, T^n|W_1, W_2) \le n\epsilon_n, \tag{207}$$

$$H(T^n|S^n, W_1, W_2) \le H(S^n, T^n|W_1, W_2) \le n\epsilon_n \tag{208}$$

where $\epsilon_n \to 0$ as $n \to \infty$. Using the constraints on the degrees of the vertexes in the message-graph, we have

$$\begin{aligned} n(R_1 + R_2') &\overset{(a)}{\ge} H(W_1, W_2) - \log\mu && (209)\\ &= I(S^n; T^n|W_1, W_2) + H(W_1, W_2|S^n, T^n) - \log\mu && (210)\\ &\overset{(b)}{=} I(S^n; T^n|W_1, W_2) - \log\mu && (211)\\ &= H(S^n, T^n) - H(S^n, T^n|W_1, W_2) - \log\mu && (212)\\ &\overset{(c)}{\ge} H(S^n, T^n) - n\epsilon_n - \log\mu && (213)\\ &\overset{(d)}{=} nH(S, T) - n\epsilon_n - \log\mu, && (214) \end{aligned}$$

where

(a) is from the fact that $(W_1, W_2) \in E(G)$ with $|E(G)| \le \mu 2^{n(R_1+R_2')}$,

(b) is obtained since W_1 and W_2 is a function of S^n and T^n, respectively,

(c) follows from Fano's inequality (206), and

(d) is obtained by using the chain rule and the fact that (S_i, T_i) are i.i.d.

Similarly, we can prove

$$n(R_1' + R_2) \ge nH(S, T) - n\epsilon_n - \log\mu. \tag{215}$$

Also, by using (207), we can have

$$\begin{aligned} nR_1 &\overset{(a)}{\ge} H(W_1) && (216)\\ &\ge H(W_1|T^n) && (217)\\ &= I(S^n; W_1|T^n) + H(W_1|S^n, T^n) && (218)\\ &\overset{(b)}{=} I(S^n; W_1|T^n) && (219)\\ &\overset{(c)}{=} H(S^n|T^n) - H(S^n|T^n, W_1, W_2) && (220)\\ &\overset{(d)}{\ge} H(S^n|T^n) - n\epsilon_n && (221)\\ &\overset{(e)}{=} nH(S|T) - n\epsilon_n, && (222) \end{aligned}$$

where

(a) is from the fact that $W_1 \in \{1, 2, \ldots, 2^{nR_1}\}$,

(b) is obtained since $W_1 = f_1(S^n)$,

(c) follows from $H(S^n|T^n, W_1) = H(S^n|T^n, W_1, W_2) + \underbrace{I(S^n; W_2|T^n, W_1)}_{=0}$ due to the Markov chain $S^n \to T^n \to W_2$,

(d) follows from Fano's inequality (207), and

(e) is obtained by using the chain rule and the fact that (S_i, T_i) are i.i.d.

Similarly, by using (208), we also can obtain

$$nR_2 \geq nH(T|S) - n\epsilon_n. \tag{223}$$

Therefore, we can have the converse proof by dividing the three inequalities (215), (222), and (223) by n, and taking the limit as $n \to \infty$. ■

References

[1] R. Ahlswede, "Multi-way communication channels," in *2nd Int. Symp. Inform. Theory*, Tsahkadsor, S.S.R. Armenia, 1971, pp. 23–52, Publishing House of the Hungarian Academy of Science, 1973.

[2] H. Liao, "A coding theorem for multiple access communications," in *Proc. Int. Symp. Inform. Theory*, Asilomar, CA, 1972, : also "Multiple Access Channels," Ph.D. dissertation, Dept. of Elec. Eng., Univ. of Hawaii, 1972.

[3] E. C. Van der Meulen, "The discrete memoryless channel with two senders and one receiver," in *2nd Int. Symp. Inform. Theory*, Tsahkadsor, S.S.R. Armenia, 1971, pp. 103–135, Publishing House of the Hungarian Academy of Science, 1973.

[4] R. Ahlswede, "The capacity region of a channel with two senders and two receivers," *Ann. Prob.*, vol. 2, pp. 805–814, Oct. 1974.

[5] ——, "An elementary proof of the strong converse theorem for the multiple-access channel," *J. Comb. Inform. Syst. Sci.*, vol. 7, pp. 216–230, 1982.

[6] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 471–480, Jul. 1973.

[7] T. M. Cover, "A proof of the data compression theorem of slepian and wolf for ergodic sources," *IEEE Trans. Inform. Theory*, vol. IT-21, no. 2, pp. 226–228, Mar. 1975.

[8] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* New York:Wiley, 1991.

[9] D. Slepian and J. K. Wolf, "A coding theorem for multiple access channels with correlated sources," *Bell Syst. Tech. J.*, vol. 52, pp. 1037–1076, Sep. 1973.

[10] T. M. Cover, A. El Gamal, and M. Salehi, "Multiple-access channel with arbitrarily correlated sources," *IEEE Trans. Inform. Theory*, vol. IT-26, no. 6, pp. 648–657, Nov. 1980.

[11] P. Gács and J. Körner, "Common information is much less than mutual information," *Problems of Control and Information Theory*, vol. 2, pp. 149–162, 1973.

[12] H. S. Witsenhausen, "On sequences of pairs of dependent random variables," *SIAM J. Appl. Math.*, vol. 28, pp. 100–113, Jan. 1975.

[13] A. D. Wyner, "The common information of two dependent random variables," *IEEE Trans. Inform. Theory*, vol. IT-21, no. 2, pp. 163–179, Mar. 1975.

[14] S. Vembu, S. Verdú, and Y. Steinberg, "The source-channel separation theorem revisited," *IEEE Trans. Inform. Theory*, vol. 41, no. 1, pp. 44–54, Jan. 1995.

[15] G. Dueck, "A note on the multiple access channel with correlated sources," *IEEE Trans. Inform. Theory*, vol. IT-27, no. 2, pp. 232–235, Mar. 1981.

[16] T. S. Han, "The capacity region of general multiple access channel with certain correlated sources," *Inform. Contr.*, vol. 40, pp. 37–60, 1979.

[17] F. Willems, "The discrete memoryless multiple access channel with partially cooperating encoders," *IEEE Trans. Inform. Theory*, vol. IT-29, no. 3, pp. 441–445, May 1983.

[18] K. De Bruyn, V. V. Prelov, and E. C. Van der Meulen, "Reliable transmission of two correlated sources over an asymmetric multiple access channel," *IEEE Trans. Inform. Theory*, vol. IT-33, no. 5, pp. 716–718, Sep. 1987.

[19] W. Kang and S. Ulukus, "A single-letter upper bound for the sum rate of multiple access channels with correlated sources," in *39th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, 2005.

[20] J. H. van Lint and R. M. Wilson, *A course in combinatorics.* Cambrigde University Press, 2003.

[21] R. P. Stanley, *Enumerative combinatorics.* Cambrigde University Press, 2002.

[22] R. Ahlswede and T. S. Han, "On source coding with side information via a multiple-access channel and related problems in multi-user information theory," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 396–412, May 1983.

[23] B. Wang and F. Zhang, "On the precise number of (0,1)-matrices in U(R,S)," *Discrete mathematics*, vol. 187, pp. 211–220, 1998.

[24] B. R. Perez-Salvador, S. de-los Cobos-Silva, M. A. Gutierrez-Andrade, and A. Torres-Chazaro, "A reduced formula for the precise number of (0,1)-matrices in A(R,S)," *Discrete mathematics*, vol. 256, pp. 361–372, 2002.

[25] B. McKay, "Practical graph isomorphism," *Congressus Numerantium*, vol. 30, pp. 45–87, 1981.

[26] D. G. Corneil and C. C. Gottlieb, "An efficient algorithm for graph isomorphism," *Journal of the ACM*, vol. 17, pp. 51–64, 1970.

[27] E. M. Luks, "Isomorphism of graphs of bounded valence can be tested in polynomial time," *Journal of computer and systems sciences*, vol. 25, pp. 42–65, 1982.

[28] T. Berger, *Multiterminal Source Coding. In: The Information Theory Approach to Communications (ed. G. Longo), CISM Courses and Lecture Notes No. 229.* Springer, Wien-New York, 1977.

[29] R. G. Gallager, *Information Theory and Reliable Communication.* New York:Wiley, 1968.

[30] A. D. Wyner, "The rate-distortion function for source coding with side information at the decoder-II: General sources," *Inform. Contr.*, vol. 38, pp. 60–80, 1979.