Multiple Access Channels with Arbitrarily Correlated Sources

THOMAS M. COVER, FELLOW, IEEE, ABBAS EL GAMAL, MEMBER, IEEE, AND MASOUD SALEHI, MEMBER, IEE

Abstract—Let $\{(U_i,V_i)\}_{i=1}^n$ be a source of independent identically distributed (i.i.d.) discrete random variables with joint probability mass function $p(u,v)$ and common part $w=f(u)=g(v)$ in the sense of Witsenhausen, Gacs, and Körner. It is shown that such a source can be sent with arbitrarily small probability of error over a multiple access channel (MAC)

$$\{\mathcal{X}_1\times\mathcal{X}_2,\mathcal{Y},p(y|x_1,x_2)\},$$

with allowed codes $\{x_1(u),x_2(v)\}$ if there exist probability mass functions $p(s), p(x_1|s,u), p(x_2|s,v)$, such that

$$H(U|V)<I(X_1;Y|X_2,V,S),$$
$$H(V|U)<I(X_2;Y|X_1,U,S),$$
$$H(U,V|W)<I(X_1,X_2;Y|W,S),$$
$$H(U,V)<I(X_1,X_2;Y),$$

where

$$p(s,u,v,x_1,x_2,y)=p(s)p(u,v)p(x_1|u,s)p(x_2|v,s)p(y|x_1,x_2).$$

This region includes the multiple access channel region and the Slepian–Wolf data compression region as special cases.

I. Introduction

The MULTIPLE access channel (MAC) $p(y|x_1,x_2)$ has a capacity region [1], [2] given by the convex hull of all (R_1,R_2) satisfying, for some $p(x_1,x_2)=p(x_1)p(x_2)$, the inequalities

$$\begin{aligned}R_1&\leq I(X_1;Y|X_2),\\ R_2&\leq I(X_2;Y|X_1),\\ R_1+R_2&\leq I(X_1,X_2;Y).\end{aligned}\tag{1}$$

Suppose now that the source U for X_1 and V for X_2 are correlated according to $p(u,v)$. It follows easily that U and V can be sent over the multiple access channel if, for some $p(x_1,x_2)=p(x_1)p(x_2)$,

$$\begin{aligned}H(U)&<I(X_1;Y|X_2),\\ H(V)&<I(X_2;Y|X_1),\\ H(U)+H(V)&<I(X_1,X_2;Y).\end{aligned}\tag{2}$$

In this paper, we increase this achievable region in two ways: 1) the left side will be made smaller[1], and 2) the right side will be made larger by allowing X_1 and X_2 to depend on U and V and thereby increasing the set of mass distributions $p(x_1,x_2)$. It will be shown (see Theorem 1 for a precise and more general statement) that U and V can be sent with arbitrarily small error to Y if

$$\begin{aligned}H(U|V)&<I(X_1;Y|X_2,V),\\ H(V|U)&<I(X_2;Y|X_1,U),\\ H(U,V)&<I(X_1,X_2;Y),\end{aligned}\tag{3}$$

for some $p(u,v,x_1,x_2,y)=p(u,v)p(x_1|u)p(x_2|v)\cdot p(y|x_1,x_2)$. This result can be further generalized to sources (U,V) with a common part $W=f(U)=g(V)$. The following theorem is proved.

Theorem 1: A source $(U,V)\sim\prod_i p(u_i,v_i)$ can be sent with arbitrarily small probability of error over a multiple access channel $\{\mathcal{X}_1\times\mathcal{X}_2,\mathcal{Y},p(y|x_1,x_2)\}$, with allowed codes $\{\boldsymbol{x}_1(\boldsymbol{u}),\boldsymbol{x}_2(\boldsymbol{v})\}$ if there exist probability mass functions $p(s), p(x_1|s,u), p(x_2|s,v)$, such that

$$\begin{aligned}H(U|V)&<I(X_1;Y|X_2,V,S),\\ H(V|U)&<I(X_2;Y|X_1,U,S),\\ H(U,V|W)&<I(X_1,X_2;Y|W,S),\\ H(U,V)&<I(X_1,X_2;Y),\end{aligned}\tag{4}$$

where $p(s,u,v,x_1,x_2,y)=p(s)p(u,v)p(x_1|u,s)\cdot p(x_2|v,s)p(y|x_1,x_2)$.

Remark 1: The region described above is convex. Therefore no time sharing is necessary. The proof of the convexity is given in Appendix B.

Remark 2: It can be shown that if error-free transmission is possible, then in order to generate a random code for error-free transmission, it is enough to consider those auxiliary random variables S whose cardinality is bounded above by $\min\{\|X_1\|\cdot\|X_2\|,\|Y\|\}$.

[1]This improvement could be obtained from the results of Slepian and Wolf [3].

Manuscript received November 28, 1978; revised February 28, 1980. This work was supported in part by the National Science Foundation under Grant ENG 76-23334, in part by the Stanford Research Institute under International Contract DAHC-15-C-0187, and in part by the Joint Scientific Engineering Program under Contracts N00014-75-C-0601 and F44620-76-C-0601. This paper was presented at the 1979 IEEE International Symposium on Information Theory, Grignano, Italy, June 25–29, 1979.

T. M. Cover is with the Departments of Electrical Engineering and Statistics, Stanford University, Durand Building, Room 121, Stanford, CA 94305.

A. El Gamal was with the Department of Electrical Engineering, University of Southern California, University Park, Los Angeles, CA. He is now with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305.

M. Salehi was with the Department of Electrical Engineering, Stanford University, Stanford, CA. He is now with the Department of Electrical Engineering, University of Isfahan, Isfahan, Iran.

Example for Theorem 1: Consider the transmission of the correlated sources (U,V) with the joint distribution $p(u,v)$ given by

$\mathcal{V} \backslash \mathcal{U}$	0	1
0	1/3	1/3
1	0	1/3

over the multiple access channel defined by

$$\mathcal{X}_1=\mathcal{X}_2=\{0,1\}$$
$$\mathcal{Y}=\{0,1,2\},$$
$$Y=X_1+X_2.$$

Here $H(U,V)=\log 3=1.58$ bits. On the other hand, if X_1 and X_2 are independent,

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

Thus $H(U,V)>I(Y;X_1,X_2)$ for all $p(x_1)p(x_2)$. Consequently there is no way, even with the use of Slepian–Wolf data compression on U and V, to use the standard multiple access channel capacity region to send U and V reliably to Y. However, it is easy to see that with the choice $X_1\equiv U$, and $X_2\equiv V$, error-free transmission of the sources over the channel is possible. This example shows that the separate source and channel coding described above is not optimal—the partial information that each of the random variables U and V contains about the other is destroyed in this separation.

To allow partial cooperation between the two transmitters, we allow our codes to depend statistically on the source outputs. This induces dependence between codewords.

We note that, while there are $2^{nH(U)}$ $\boldsymbol{x}_1$ associated with the typical $\boldsymbol{u}$ and $2^{nH(V)}$ $\boldsymbol{x}_2$ associated with the typical $\boldsymbol{v}$, there are only $2^{nH(U,V)}$ pairs $(\boldsymbol{x}_1(\boldsymbol{u}),\boldsymbol{x}_2(\boldsymbol{v}))$ that are likely to occur jointly.

Applications of Theorem 1 yield the following known results as special cases.

Special Cases

a) Slepian and Wolf Data Compression [3]*:* Let (U,V) be correlated according to $p(u,v)$. To obtain the data compression rate region, we set up a noiseless dummy channel with $Y=(X_1,X_2)$. Let $p(u,v,x_1,x_2)=p(u,v)p(x_1)p(x_2)$. Then the right side of (3) collapses, yielding the known rate region

$$\begin{aligned}
&H(U|V)<I(X_1;Y|X_2,V)=H(X_1) && (=R_1)\\
&H(V|U)<I(X_2;Y|X_1,U)=H(X_2) && (=R_2)\\
&H(U,V)<I(X_1,X_2;Y)=H(X_1)+H(X_2) && (=R_1+R_2).
\end{aligned} \tag{5}$$

b) Multiple Access Channel (Ahlswede [1]*, Liao* [2]*):* Let U and V be independent dummy sources with rates R_1 and R_2, respectively. Choose $p(u,v,x_1,y)=p(u)p(v)p(x_1)p(x_2)p(y|x_1,x_2)$. Now both sides of (3) simplify to yield achievability of rates (R_1,R_2) for the multiple access channel to

$$\begin{aligned}
&H(U|V)=H(U)=R_1<I(X_1;Y|X_2),\\
&H(V|U)=H(V)=R_2<I(X_2;Y|X_1),\\
&H(U,V)=H(U)+H(V)=R_1+R_2<I(X_1,X_2;Y).
\end{aligned} \tag{6}$$

c) Cooperative Multiple Access Channel Capacity: If both X_1 and X_2 have access to the same source, we can find the cooperative capacity for the multiple access channel $p(y|x_1,x_2)$ as follows. Let U be a dummy source with rate R, and let $W=V=U$. Choose $p(u,s,x_1,x_2,y)=p(u)p(s)p(x_1|s)p(x_2|s)p(y|x_1,x_2)$. Eliminating the trivial inequalities, we then have the achievability of rate R if

$$R<I(X_1,X_2;Y), \tag{7}$$

for some joint probability mass function $p(x_1,x_2)$.

d) The Correlated Source Multiple Access Channel Capacity Region of Slepian and Wolf [4]*:* Following Slepian and Wolf [4] for the multiple access channel $p(y|x_1,x_2)$, suppose that x_1 sees a source of rate R_1, x_2 sees a source of rate R_2, and in addition, both x_1 and x_2 see a common source of rate R_0. All three sources are independent.

To obtain the desired region, let U', V', W be independent dummy random variables with $R_1=H(U')$, $R_2=H(V')$, $R_0=H(W)$. Let $U=(U',W)$ and $V=(V',W)$. Choose $p(u,v,s,x_1,x_2,y)=p(u')p(v')p(w)p(s)p(x_1|s)\cdot p(x_2|s)p(y|x_1,x_2)$, where $u=(u',w)$, $v=(v',w)$. We then have achievability of (R_0,R_1,R_2) if

$$\begin{aligned}
H(U|V)&=H(U')=R_1<I(X_1;Y|X_2,S),\\
H(V|U)&=H(V')=R_2<I(X_2;Y|X_1,S),\\
H(U,V|W)&=H(U')+H(V')\\
&=R_1+R_2<I(X_1,X_2;Y|S),\\
H(U,V)&=H(U')+H(V')+H(W)\\
&=R_0+R_1+R_2<I(X_1,X_2;Y).
\end{aligned} \tag{8}$$

Theorem 1 shows that the multiple access channel capacity region and the Slepian and Wolf data compression region are special cases of a single theorem. Also, multiple source compression and multiple access channel coding do not seem to factor into separate source and channel coding problems. The work of Slepian and Wolf on correlated sources with common rate R_0 and conditionally independent rates R_1 and R_2 can be generalized to sources with common rate R_0 and conditionally dependent sources. Finally, as shown in Theorem 1, the dependence of U and V can be used to create the appearance of cooperation in the channel coding, even if U and V do not have a common part.

In the next section we shall give a formal definition of the problem and outline the proof for the simple achievability in (3). The proof of Theorem 1 is given in Section III. An expression for source-channel capacity is given in Section IV but does not satisfy the "single-letter" conditions that we seek.

II. Definition of the Problem

Assume we have two information sources $U_1, U_2, \cdots$ and $V_1, V_2, \cdots$ generated by repeated independent drawings of a pair of discrete random variables U and V from a given bivariate distribution $p(u,v)$. We shall require the following notion of the common part of two random variables.

Definition: The *common part* W of two random variables U and V is defined by finding the maximum integer k such that there exist functions f and g

$$f: \mathcal{U} \to \{1,2,\cdots,k\}$$
$$g: \mathcal{V} \to \{1,2,\cdots,k\}$$

with $P\{f(U)=i\}>0$, $P\{g(V)=i\}>0$, $i=1,2,\cdots,k$, such that $f(U)=g(V)$ with probability one and then defining $W=f(U)\ \ (=g(V))$.

With this definition, it is obvious that the observers of U and V can agree on the value of W with probability one. Note that any pair of sources (U,V) has a trivial common part $f(U)=g(V)=1$. Here $k=1$ in the construction that follows the definition. We shall say that U and V have a common part only if $k \geqslant 2$.

Also, it can be shown [7] that the common part of sequence (U_i, V_i) i.i.d.$\sim p(u,v)$ is the sequence of the common parts W_i. The concept of the common part of two random variables will be used in Section III.

We now define the communication problem over the multiple access channel in Fig. 1. This includes the definition of block codes for sources, the definition of probability of error, and the definition of reliable transmission of sources over the channel.

A *block code* for the channel consists of an integer n, two *encoding functions*

$$\boldsymbol{x}_1^n: \mathcal{U}^n \to \mathcal{X}_1^n,$$
$$\boldsymbol{x}_2^n: \mathcal{V}^n \to \mathcal{X}_2^n$$

assigning codewords to the source outputs, and a *decoding function*

$$d^n: \mathcal{Y}^n \to \mathcal{U}^n \times \mathcal{V}^n. \tag{9}$$

The *probability of error* is given by

$$\begin{aligned} P_n &= P\{(\boldsymbol{U}^n, \boldsymbol{V}^n) \neq d^n(\boldsymbol{Y}^n)\} \\ &= \sum_{(\boldsymbol{u},\boldsymbol{v}) \in \mathcal{U}^n \times \mathcal{V}^n} p(\boldsymbol{u}^n, \boldsymbol{v}^n) \\ &\quad \cdot P\{d^n(\boldsymbol{Y}^n) \neq (\boldsymbol{u}^n, \boldsymbol{v}^n) | (\boldsymbol{U}^n, \boldsymbol{V}^n) = (\boldsymbol{u}^n, \boldsymbol{v}^n)\}. \end{aligned} \tag{10}$$

where the joint probability mass function is given, for a code assignment $\{\boldsymbol{x}_1(\boldsymbol{u}^n), \boldsymbol{x}_2(\boldsymbol{v}^n)\}$, by

$$p(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{y}) = \prod_{i=1}^{n} p(u_i, v_i) p(y_i | x_{1i}(\boldsymbol{u}^n), x_{2i}(\boldsymbol{v}^n)). \tag{11}$$

Definition: The source $(U,V) \sim \Pi p(u_i, v_i)$ can be *reliably transmitted* over the multiple access channel $(\mathcal{X}_1 \times \mathcal{X}_2, \mathcal{Y}, p(y|x_1,x_2))$ if there exists a sequence of block codes $\{\boldsymbol{x}_1^n(\boldsymbol{u}^n), \boldsymbol{x}_2^n(\boldsymbol{v}^n)\}$, $d^n(\boldsymbol{y}^n)$ such that

$$P_n = P\{d^n(\boldsymbol{Y}^n) \neq (\boldsymbol{U}^n, \boldsymbol{V}^n)\} \to 0.$$

The notions of jointly ϵ-typical sequences and the asymptotic equipartition property as defined in [5] and [6] will be used throughout this paper.

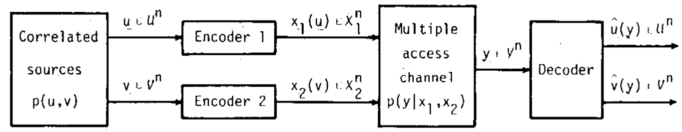

Fig. 1. Multiple access channel with arbitrarily correlated sources.

Since the proof of Theorem 1, given in the next section, is rather long and technical we shall outline here a proof of the simpler case in which U and V have no common part. In this case, we must show that U and V can be reliably sent to Y if, for $p(u,v)p(x_1|u)p(x_2|v)p(y|x_1,x_2)$,

$$\begin{aligned} H(U|V) &< I(X_1;Y|X_2,V), \\ H(V|U) &< I(X_2;Y|X_1,U), \\ H(U,V) &< I(X_1,X_2;Y). \end{aligned} \tag{12}$$

The proof will employ random coding. We first describe the random code generation and encoding–decoding schemes and then analyze the probability of error.

Generating Random Codes: Fix $p(x_1|u)$ and $p(x_2|v)$; for each $\boldsymbol{u} \in \mathcal{U}^n$ generate *one* $\boldsymbol{x}_1$ sequence drawn according to $\Pi_{i=1}^n p(x_{1i}|u_i)$ and for each $\boldsymbol{v} \in \mathcal{V}^n$ generate *one* $\boldsymbol{x}_2$ sequence drawn according to $\Pi_{i=1}^n p(x_{2i}|v_i)$. Call these sequences $\boldsymbol{x}_1(\boldsymbol{u})$ and $\boldsymbol{x}_2(\boldsymbol{v})$, respectively.

Encoding: Transmitter 1, upon observing $\boldsymbol{u}$ at the output of source 1, transmits $\boldsymbol{x}_1(\boldsymbol{u})$, and transmitter 2, after observing $\boldsymbol{v}$ at the output of source 2, transmits $\boldsymbol{x}_2(\boldsymbol{v})$. Assume the maps $\boldsymbol{x}_1(\cdot)$, $\boldsymbol{x}_2(\cdot)$ are known to the receiver.

Decoding: Upon receiving $\boldsymbol{y}$, the decoder finds the only $(\boldsymbol{u},\boldsymbol{v})$ pair such that $(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{x}_1(\boldsymbol{u}), \boldsymbol{x}_2(\boldsymbol{v}), \boldsymbol{y}) \in A_\epsilon$, where A_ϵ is the set of jointly ϵ-typical sequences. If there is no such $(\boldsymbol{u},\boldsymbol{v})$ pair, or there exists more than one such pair, the decoder declares an error. A helpful picture is given in Fig. 2.

Error: Suppose $(\boldsymbol{u}_0, \boldsymbol{v}_0)$ is the source output. Then an error is made if

i) $(\boldsymbol{u}_0, \boldsymbol{v}_0, \boldsymbol{x}_1(\boldsymbol{u}_0), \boldsymbol{x}_2(\boldsymbol{v}_0), \boldsymbol{y}) \notin A_\epsilon$,

or

ii) There exists some $(\boldsymbol{u},\boldsymbol{v}) \neq (\boldsymbol{u}_0, \boldsymbol{v}_0)$ such that $(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{x}_1(\boldsymbol{u}), \boldsymbol{x}_2(\boldsymbol{v}), \boldsymbol{y}) \in A_\epsilon$.

Then the probability of error P_n can be bounded as:

$$\begin{aligned} P_n &= P\{(\boldsymbol{U}_0, \boldsymbol{V}_0, \boldsymbol{X}_1(\boldsymbol{U}_0), \boldsymbol{X}_2(\boldsymbol{V}_0), \boldsymbol{Y}) \notin A_\epsilon\} \\ &\quad + P\{\exists (\boldsymbol{u},\boldsymbol{v}) \neq (\boldsymbol{U}_0, \boldsymbol{V}_0) : (\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{X}_1(\boldsymbol{u}), \boldsymbol{X}_2(\boldsymbol{v}), \boldsymbol{Y}) \in A_\epsilon, \\ &\quad (\boldsymbol{U}_0, \boldsymbol{V}_0) \in A_\epsilon\} \\ &\leqslant \epsilon + \sum_{(\boldsymbol{u}_0,\boldsymbol{v}_0) \in A_\epsilon} p(\boldsymbol{u}_0, \boldsymbol{v}_0) \\ &\quad \cdot \sum_{\substack{\boldsymbol{u} \neq \boldsymbol{u}_0, \\ \boldsymbol{v} = \boldsymbol{v}_0}} P\{(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{X}_1(\boldsymbol{u}), \boldsymbol{X}_2(\boldsymbol{v}), \boldsymbol{Y}) \in A_\epsilon | (\boldsymbol{u}_0, \boldsymbol{v}_0)\} \\ &\quad \cdot \sum_{(\boldsymbol{u}_0,\boldsymbol{v}_0) \in A_\epsilon} p(\boldsymbol{u}_0, \boldsymbol{v}_0) \sum_{\substack{\boldsymbol{u} = \boldsymbol{u}_0, \\ \boldsymbol{v} \neq \boldsymbol{v}_0}} P\{\cdot\} \\ &\quad + \sum_{(\boldsymbol{u}_0,\boldsymbol{v}_0) \in A_\epsilon} p(\boldsymbol{u}_0, \boldsymbol{v}_0) \sum_{\substack{\boldsymbol{u} \neq \boldsymbol{u}_0, \\ \boldsymbol{v} \neq \boldsymbol{v}_0}} P\{\cdot\} \\ &\leqslant \epsilon + 2^{n(H(U|V)+\epsilon)} 2^{-n(I(X_1;Y|X_2,V)-\epsilon)} \\ &\quad + 2^{n(H(V|U)+\epsilon)} 2^{-n(I(X_2;Y|X_1,U)-\epsilon)} \\ &\quad + 2^{nH(U,V)} 2^{-n(I(X_1,X_2;Y)-\epsilon)}. \end{aligned} \tag{13}$$

Consequently $P_n \to 0$ if the conditions in (12) are satisfied.

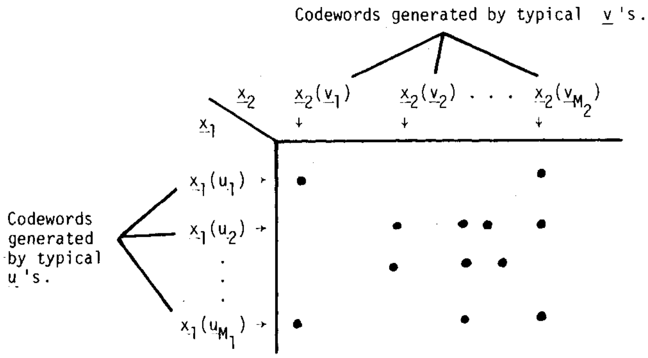

Fig. 2. Picture of joint typicality for multiple access channel. Dots correspond to jointly typical (X_1, X_2) pairs. Note that only $2^{nH(U,V)}(x_1(u), x_2(v))$ pairs are likely to occur.

III. Proof of Theorem 1

The encoding and decoding schemes for Theorem 1 will be described; then the probability of error will be analyzed.

Generation of Random Codes: Fix the probability mass functions $p(s)$, $p(x_1|s,u)$, $p(x_2|s,v)$.

i) For each $w \in \mathcal{W}^n$, independently generate one s sequence according to $\prod_{i=1}^n p(s_i)$. Index them by $s(w)$, $w \in \mathcal{W}^n$.

ii) For each $u \in \mathcal{U}^n$ find the corresponding $w = f(u) = (f(u_1), \cdots, f(u_n))$ and independently generate one x_1 sequence according to $\prod_{i=1}^n p(x_{1i}|u_i, s_i(w))$. Index the x_1 sequences by $x_1(u|s(f(u)))$ or for simplicity by $x_1(u|s)$, $u \in \mathcal{U}^n$, $s \in \mathcal{S}^n$, where u and s are such that $s = s(f(u))$, as generated in i). The same procedure, using $\prod_{i=1}^n p(x_{2i}|v_i, s_i(w))$, is repeated for the v sequences. These sequences are indexed by $x_2(v|s(g(v)))$ or for simplicity by $x_2(v|s)$, $v \in \mathcal{V}^n$, $s \in \mathcal{S}^n$, where v and s are such that $s = s(g(v))$.

Encoding: Upon observing the output u of the source, transmitter 1 finds $s(f(u))$ and sends $x_1(u|s)$. Similarly, transmitter 2 sends $x_2(v|s)$, where $s = s(g(v))$.

Note that every $u \in \mathcal{U}^n$ and every $v \in \mathcal{V}^n$ is mapped into a codeword in $\mathcal{X}_1^n$ and $\mathcal{X}_2^n$, respectively. However, with high probability only $2^{nH(U,V)}$ codeword pairs (x_1, x_2) can simultaneously occur. This fact is crucial in the proof of achievability.

Decoding: Upon observing the received sequence y, the decoder declares $(\hat{u}, \hat{v})$ to be the transmitted source sequence pair if $(\hat{u}, \hat{v})$ is the unique pair (u, v) such that

$$(u, v, w, s(w), x_1(u|s), x_2(v|s), y) \in A_\epsilon,$$

where $w = f(u)$.

Error: Suppose (u_0, v_0) was the source output pair, then an error is made if

i) $(u_0, v_0, w_0, s(w_0), x_1(u_0|s), x_2(v_0|s), y) \in A_\epsilon$,

or

ii) there exists some $(u, v) \neq (u_0, v_0)$ such that $(u, v, w, s(w), x_1(u|s), x_2(v|s), y) \in A_\epsilon$.

Analysis of the Probability of Error: Letting A_ϵ denote the appropriate set of jointly ϵ-typical sequences (see [5] and [6]), we have

$$\bar{P}_n = \sum_{(u,v) \in \mathcal{U}^n \times \mathcal{V}^n} p(u,v) P\{\text{error made at decoder } |(u,v) \text{ is the output of the source}\}, \quad (14)$$

or

$$\bar{P}_n \leq \sum_{(u,v) \in A_\epsilon} p(u,v) P\{\text{error made at decoder } |(u,v) \text{ is the output of the source}\} + \sum_{(u,v,w) \notin A_\epsilon} p(u,v). \quad (15)$$

From the asymptotic equipartition property (AEP), for sufficiently large n,

$$\bar{P}_n \leq \sum_{(u,v,w) \in A_\epsilon} p(u,v) P\{\text{error made at decoder } |(u,v) \text{ is the output of the source}\} + \epsilon. \quad (16)$$

Now we show that as long as $(u, v, w) \in A_\epsilon$, there exists an upper bound independent of (u, v) for the terms in the summation. To show this, we assume that $(u_0, v_0, w_0) \in A_\epsilon$ and let $\mathfrak{B}$ denote the event that this special triple is the output of the source. We are interested in an upper bound for $P\{\text{error made at decoder}|\mathfrak{B}\}$.

The event E that an error is made at decoder is the union of two events E_1 and E_2,

$$E = E_1 \cup E_2, \quad (17)$$

where

E_1: the event that $(u_0, v_0, w_0, S_0, X_1(u|S_0), X_2(v|S_0), Y) \notin A_\epsilon$;

E_2: the event that there exists some $(u, v) \neq (u_0, v_0)$ such that

$$(u, v, w, S(w), X_1(u|S), X_2(v|S), Y) \in A_\epsilon.$$

Note: Since we have generated our code randomly and we are averaging the probability of error over all coding schemes generated this way, S, X_1, X_2, and Y are the only random variables in the event E.

It follows from the AEP that n can be chosen large enough such that

$$P\{E_1|\mathfrak{B}\} \leq \epsilon, \quad (18)$$

and therefore by the union bound

$$P\{E|\mathfrak{B}\} \leq P\{E_2|\mathfrak{B}\} + \epsilon. \quad (19)$$

Using (16) and (19) and the definition of the event E we have

$$\bar{P}_n \leq P\{E_2|\mathfrak{B}\} + 2\epsilon. \quad (20)$$

We decompose the event E_2 into

$$E_2 = E_{21} \cup E_{22} \cup E_{23} \cup E_{24} \cup E_{25}, \quad (21)$$

where

E_{21}: the event that there exists a $u \neq u_0$ such that

$$(u, v_0, w_0, S_0, X_1(u|S_0), X_2(v_0|S_0), Y) \in A_\epsilon;$$

E_{22}: the event that there exists a $v \neq v_0$ such that

$$(u_0, v, w_0, S_0, X_1(u_0|S_0), X_2(v|S_0), Y) \in A_\epsilon;$$

E_{23}: the event that there exists a $u \neq u_0$ and a $v \neq v_0$ such that

$$f(u) = g(v) = w_0$$

and

$$(u, v, w_0, S_0, X_1(u|S_0), X_2(v|S_0), Y) \in A_\epsilon.$$

E_{24}: the event that there exists a $\boldsymbol{u}\neq\boldsymbol{u}_0$ and a $\boldsymbol{v}\neq\boldsymbol{v}_0$ such that

$$\boldsymbol{w}=\boldsymbol{f}(\boldsymbol{u})=\boldsymbol{g}(\boldsymbol{v})\neq\boldsymbol{w}_0, \qquad \boldsymbol{S}(\boldsymbol{f}(\boldsymbol{u}))\neq\boldsymbol{S}_0$$

and

$$(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w},\boldsymbol{S}(\boldsymbol{w}),\boldsymbol{X}_1(\boldsymbol{u}|\boldsymbol{S}),\boldsymbol{X}_2(\boldsymbol{v}|\boldsymbol{s}),\boldsymbol{Y})\in A_\epsilon;$$

E_{25}: the event that there exists a $\boldsymbol{u}\neq\boldsymbol{u}_0$ and a $\boldsymbol{v}\neq\boldsymbol{v}_0$ such that

$$\boldsymbol{w}=\boldsymbol{f}(\boldsymbol{u})=\boldsymbol{g}(\boldsymbol{v})\neq\boldsymbol{w}_0, \qquad \boldsymbol{S}(\boldsymbol{f}(\boldsymbol{u}))=\boldsymbol{S}_0$$

and

$$(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w},\boldsymbol{S}(\boldsymbol{w}),\boldsymbol{X}_1(\boldsymbol{u}|\boldsymbol{S}),\boldsymbol{X}_2(\boldsymbol{v}|\boldsymbol{s}),\boldsymbol{Y})\in A_\epsilon.$$

By the union bound, we have

$$P\{E_2|\mathfrak{B}\}\leq\sum_{i=1}^{5}P\{E_{2i}|\mathfrak{B}\}. \tag{22}$$

Now it remains to bound $P\{E_{2i}|\mathfrak{B}\}$ for $i=1,2,3,4,5$.

Bound for $P\{E_{21}|\mathfrak{B}\}$: We have

$$P\{E_{21}|\mathfrak{B}\}=P\{\exists\boldsymbol{u}\neq\boldsymbol{u}_0\colon(\boldsymbol{u},\boldsymbol{v}_0,\boldsymbol{w}_0,\boldsymbol{S}_0,\boldsymbol{X}_1(\boldsymbol{u}|\boldsymbol{S}_0),\boldsymbol{X}_2(\boldsymbol{v}_0|\boldsymbol{S}_0),\boldsymbol{Y})\in A_\epsilon|\mathfrak{B}\}. \tag{23}$$

Therefore,

$$P\{E_{21}|\mathfrak{B}\}=\sum_{\substack{\boldsymbol{u}\neq\boldsymbol{u}_0:\\(\boldsymbol{u},\boldsymbol{v}_0,\boldsymbol{w}_0)\in A_\epsilon}}P\{(\boldsymbol{u},\boldsymbol{v}_0,\boldsymbol{w}_0,\boldsymbol{S}_0,\boldsymbol{X}_1(\boldsymbol{u}|\boldsymbol{S}_0),\boldsymbol{X}_2(\boldsymbol{v}_0|\boldsymbol{S}_0),\boldsymbol{Y})\in A_\epsilon|\mathfrak{B}\}. \tag{24}$$

From Appendix A (A13) we have for $(\boldsymbol{u},\boldsymbol{v}_0,\boldsymbol{w}_0)\in A_\epsilon$,

$$P\{(\boldsymbol{u},\boldsymbol{v}_0,\boldsymbol{w}_0,\boldsymbol{S}_0,\boldsymbol{X}_1(\boldsymbol{u}|\boldsymbol{S}_0),\boldsymbol{X}_2(\boldsymbol{v}_0|\boldsymbol{S}_0),\boldsymbol{Y})\in A_\epsilon|\mathfrak{B}\}\leq 2^{-n[I(X_1;Y|X_2,V,S)-8\epsilon]}. \tag{25}$$

Notice that this bound is independent of $\boldsymbol{u}$ as long as $(\boldsymbol{u},\boldsymbol{v}_0)\in A_\epsilon$. Substituting (25) into (24), we have

$$P\{E_{21}|\mathfrak{B}\}\leq\sum_{\substack{\boldsymbol{u}\neq\boldsymbol{u}_0:\\(\boldsymbol{u},\boldsymbol{v}_0,\boldsymbol{w}_0)\in A_\epsilon}}2^{-n[I(X_1;Y|X_2,V,S)-8\epsilon]}, \tag{26}$$

or

$$P\{E_{21}|\mathfrak{B}\}\leq 2^{-n[I(X_1;Y|X_2,V,S)-8\epsilon]}\cdot\|\{\boldsymbol{u}\colon(\boldsymbol{u},\boldsymbol{v}_0,\boldsymbol{w}_0)\in A_\epsilon\}\|, \tag{27}$$

but typicality yields

$$\|\{\boldsymbol{u}\colon(\boldsymbol{u},\boldsymbol{v}_0,\boldsymbol{w}_0)\in A_\epsilon\}\|\leq 2^{n[H(U|V,W)+2\epsilon]}. \tag{28}$$

From (27) and (28) and using the fact that $H(U|V,W)=H(U|V)$, we have

$$P\{E_{21}|\mathfrak{B}\}\leq 2^{n[H(U|V)-I(X_1;Y|X_2,V,S)+10\epsilon]}. \tag{29}$$

Thus if

$$H(U|V)<I(X_1;Y|X_2,V,S)-10\epsilon, \tag{30}$$

then for large enough n, we have

$$P\{E_{21}|\mathfrak{B}\}\leq\epsilon. \tag{31}$$

Bound for $P\{E_{22}|\mathfrak{B}\}$: This case is parallel to the previous case and it can be shown similarly that if

$$H(V|U)<I(X_2;Y|X_1,U,S)-10\epsilon, \tag{32}$$

then by choosing n sufficiently large, we have

$$P\{E_{22}|\mathfrak{B}\}\leq\epsilon. \tag{33}$$

Bound for $P\{E_{23}|\mathfrak{B}\}$: Here we have

$$P\{E_{23}|\mathfrak{B}\}=P\{\exists\boldsymbol{u}\neq\boldsymbol{u}_0,\boldsymbol{v}\neq\boldsymbol{v}_0\colon\boldsymbol{f}(\boldsymbol{u})=\boldsymbol{g}(\boldsymbol{v})=\boldsymbol{w}_0 \text{ and } (\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}_0,\boldsymbol{S}_0,\boldsymbol{X}_1(\boldsymbol{u}|\boldsymbol{S}_0),\boldsymbol{x}_2(\boldsymbol{v}|\boldsymbol{S}_0),\boldsymbol{Y})\in A_\epsilon|\mathfrak{B}\}. \tag{34}$$

Therefore,

$$P\{E_{23}|\mathfrak{B}\}=\sum_{\substack{\boldsymbol{u}\neq\boldsymbol{u}_0,\,\boldsymbol{v}\neq\boldsymbol{v}_0\\(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}_0)\in A_\epsilon}}P\{(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}_0,\boldsymbol{S}_0,\boldsymbol{X}_1(\boldsymbol{u}|\boldsymbol{S}_0),\boldsymbol{X}_2(\boldsymbol{v}|\boldsymbol{S}_0),\boldsymbol{Y})\in A_\epsilon|\mathfrak{B}\} \tag{35}$$

Again, note that $\boldsymbol{u}$, $\boldsymbol{v}$, and $\boldsymbol{w}_0$ are fixed and $\boldsymbol{S}_0$, $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, and $\boldsymbol{Y}$ are random variables. Using Appendix A (A17) we have

$$P\{(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}_0,\boldsymbol{S}_0,\boldsymbol{X}_1(\boldsymbol{u}|\boldsymbol{S}_0),\boldsymbol{X}_2(\boldsymbol{v}|\boldsymbol{S}_0),\boldsymbol{Y})\in A_\epsilon|\mathfrak{B}\}\leq 2^{-n[I(X_1,X_2;Y|W,S)-8\epsilon]}. \tag{36}$$

Substituting this bound into (35), and noting that this bound is independent of $(\boldsymbol{u},\boldsymbol{v})$, we have

$$P\{E_{23}|\mathfrak{B}\}\leq\sum_{\substack{\boldsymbol{u}\neq\boldsymbol{u}_0,\,\boldsymbol{v}\neq\boldsymbol{v}_0:\\(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}_0)\in A_\epsilon}}2^{-n[I(X_1,X_2;Y|W,S)-8\epsilon]}, \tag{37}$$

or

$$P\{E_{23}|\mathfrak{B}\}\leq 2^{-n[I(X_1,X_2;Y|W,S)-8\epsilon]}\cdot\|\{(\boldsymbol{u},\boldsymbol{v})\colon(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}_0)\in A_\epsilon,\boldsymbol{u}\neq\boldsymbol{u}_0,\boldsymbol{v}\neq\boldsymbol{v}_0\}\|. \tag{38}$$

On the other hand, we have

$$\{(\boldsymbol{u},\boldsymbol{v})\colon(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}_0)\in A_\epsilon,\boldsymbol{u}\neq\boldsymbol{u}_0,\boldsymbol{v}\neq\boldsymbol{v}_0\}\subset\{(\boldsymbol{u},\boldsymbol{v})\colon(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}_0)\in A_\epsilon\}, \tag{39}$$

and

$$\|\{(\boldsymbol{u},\boldsymbol{v})\colon(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}_0)\in A_\epsilon\}\|\leq 2^{n[H(U,V|W)+2\epsilon]}. \tag{40}$$

Using (38)–(40), we obtain

$$P\{E_{23}|\mathfrak{B}\}\leq 2^{n[H(U,V|W)-I(X_1,X_2;Y|W,S)+10\epsilon]}. \tag{41}$$

Thus if

$$H(U,V,W)<I(X_1,X_2;Y|W,S)-10\epsilon, \tag{42}$$

then by choosing n large enough, we can make

$$P\{E_{23}|\mathfrak{B}\}\leq\epsilon. \tag{43}$$

Bound for $P\{E_{23}|\mathfrak{B}\}$: Recall from the definition of E_{24} that

$$P\{E_{24}|\mathfrak{B}\}=P\{\exists\boldsymbol{u}\neq\boldsymbol{u}_0,\boldsymbol{v}\neq\boldsymbol{v}_0\colon \boldsymbol{w}=\boldsymbol{f}(\boldsymbol{u})=\boldsymbol{g}(\boldsymbol{v})\neq\boldsymbol{w}_0,\boldsymbol{S}(\boldsymbol{f}(\boldsymbol{u}))\neq\boldsymbol{S}_0 \text{ and } (\boldsymbol{u},\boldsymbol{v},\boldsymbol{w},\boldsymbol{S}(\boldsymbol{w}),\boldsymbol{S}(\boldsymbol{f}(\boldsymbol{u})),\boldsymbol{X}_1(\boldsymbol{u}|\boldsymbol{S}),\boldsymbol{X}_2(\boldsymbol{v}|\boldsymbol{S}),\boldsymbol{Y})\in A_\epsilon|\mathfrak{B}\}, \tag{44}$$

from which we have

$$P\{E_{24}|\mathfrak{B}\}=\sum_{\substack{\boldsymbol{u}\neq\boldsymbol{u}_0,\,\boldsymbol{v}\neq\boldsymbol{v}_0:\\(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w})\in A_\epsilon,\,\boldsymbol{w}\neq\boldsymbol{w}_0}}P\{\boldsymbol{S}(\boldsymbol{w})\neq\boldsymbol{S}_0 \text{ and } (\boldsymbol{u},\boldsymbol{v},\boldsymbol{w},\boldsymbol{S}(\boldsymbol{w}),\boldsymbol{X}_1(\boldsymbol{u}|\boldsymbol{S}),\boldsymbol{X}_2(\boldsymbol{v}|\boldsymbol{S}),\boldsymbol{Y})\in A_\epsilon|\mathfrak{B}\}. \tag{45}$$

But, by the chain rule,

$$\begin{aligned}&P\{S(w)\neq S_0 \text{ and } (u,v,w,S(w),\\&\qquad X_1(u|S),X_2(v|S),Y)\in A_\epsilon|\mathfrak{B}\}\\&\quad=P\{S(w)\neq S_0|\mathfrak{B}\}P\{(u,v,w,S(w),X_1(u|S),\\&\qquad X_2(v|S),Y)\in A_\epsilon|S(w)\neq S_0,\mathfrak{B}\}.\end{aligned} \tag{46}$$

Therefore

$$\begin{aligned}&P\{S(w)\neq S_0 \text{ and } (u,v,w,S(w),\\&\qquad X_1(u|S),X_2(v|S),Y)\in A_\epsilon|\mathfrak{B}\}\\&\quad\leq P\{(u,v,w,S(w),\\&\qquad X_1(u|S),X_2(v|S),Y)\in A_\epsilon|S(w)\neq S_0,\mathfrak{B}\}.\end{aligned} \tag{47}$$

But

$$\begin{aligned}&P\{(u,v,w,S(w),\\&\quad X_1(u|S),X_2(v|S),Y)\in A_\epsilon|S(w)\neq S_0,\mathfrak{B}\}\\&=\sum_{s'\in\mathbb{S}^n}P\{(u,v,w,s',\\&\quad X_1(u|s'),X_2(v|s'),Y)\in A_\epsilon|S_0\neq s',\mathfrak{B}\}\cdot P\{S(w)=s'|\mathfrak{B}\}\\&=\sum_{s'\in A_\epsilon}P\{(u,v,w,s',\\&\quad X_1(u|s'),X_2(v|s'),Y)\in A_\epsilon|S_0\neq s',\mathfrak{B}\}\cdot P\{S(w)=s'|\mathfrak{B}\},\end{aligned} \tag{48}$$

where the last equality follows from the fact that for $s'\notin A_\epsilon$,

$$P\{(u,v,w,s',X_1(u|s'),X_2(v|s'),Y)\in A_\epsilon|S_0\neq s',\mathfrak{B}\}=0.$$

From Appendix A (A20) for $s'\in A_\epsilon$, we have

$$P\{(u,v,w,s',X_1(u|s'),X_2(v|s'),Y)\in A_\epsilon|S_0\neq s',\mathfrak{B}\}\leq 2^{-n[I(X_1,X_2;Y)-8\epsilon]}. \tag{49}$$

Therefore

$$\begin{aligned}&P\{(u,v,w,S(w),\\&\quad X_1(u|S),X_2(v|S),Y)\in A_\epsilon|S(w)\neq S_0,\mathfrak{B}\}\\&\qquad\leq\sum_{s'\in A_\epsilon}2^{-n[I(X_1,X_2;Y)-8\epsilon]}2^{-n[H(S)+\epsilon]}.\end{aligned} \tag{50}$$

Using the fact that

$$\|\{s': s'\in A_\epsilon\}\|\leq 2^{n[H(S)+\epsilon]}, \tag{51}$$

we have

$$\begin{aligned}&P\{(u,v,w,S(w),\\&\quad X_1(u|S),X_2(v|S),Y)\in A_\epsilon|S(w)\neq S_0,\mathfrak{B}\}\\&\qquad\leq 2^{-n[I(X_1,X_2;Y)-8\epsilon]}.\end{aligned} \tag{52}$$

Substituting this result into (46) and then into (49) we have

$$P\{E_{24}|\mathfrak{B}\}\leq\sum_{\substack{u\neq u_0,\,v\neq v_0\\(u,v,w)\in A_\epsilon,\,w\neq w_0}}2^{-n[I(X_1,X_2;Y)-8\epsilon]} \tag{53}$$

or

$$P\{E_{24}|\mathfrak{B}\}\leq 2^{-n[I(X_1,X_2;Y)-8\epsilon]}\cdot\|\{(u,v):(u,v)\in A_\epsilon\}\|, \tag{54}$$

but

$$\|\{(u,v):(u,v)\in A_\epsilon\}\|\leq 2^{n[H(U,V)+\epsilon]}. \tag{55}$$

Hence

$$P\{E_{24}|\mathfrak{B}\}<2^{n[H(U,V)-I(X_1,X_2;Y)+9\epsilon]}. \tag{56}$$

From this inequality it follows that if

$$H(U,V)<I(X_1,X_2;Y)-9\epsilon, \tag{57}$$

then we can choose n sufficiently large that

$$P\{E_{24}|\mathfrak{B}\}<\epsilon. \tag{58}$$

Bound for $P\{E_{25}|\mathfrak{B}\}$: Recall from the definition of E_{25} that

$$\begin{aligned}P\{E_{25}|\mathfrak{B}\}=P\{&\exists u\neq u_0,v\neq v_0:\\&w=f(u)=g(v)\neq w_0,S(w)=S_0,\\&(u,v,w,S(w),X_1(u|S),X_2(v|S),Y)\in A_\epsilon|\mathfrak{B}\}.\end{aligned} \tag{59}$$

Here, as in the previous cases, we can write,

$$\begin{aligned}P\{E_{25}|\mathfrak{B}\}=&\sum_{\substack{u\neq u_0,\,v\neq v_0:\\(u,v,w)\in A_\epsilon,\,w\neq w_0}}P\{S(w)=S_0 \text{ and}\\&(u,v,w,S(w),X_1(u|S),X_2(v|s),Y)\in A_\epsilon|\mathfrak{B}\},\end{aligned} \tag{60}$$

but by the chain rule we have

$$\begin{aligned}&P\{S(w)=S_0 \text{ and } (u,v,w,S(w),X_1(u|S),\\&X_2(v|S),Y)\in A_\epsilon|\mathfrak{B}\}=P\{S(w)=S_0|\mathfrak{B}\}\\&\qquad\cdot P\{(u,v,w,S(w),X_1(u|S),\\&\qquad X_2(v|S),Y)\in A_\epsilon|S(w)=S_0,\mathfrak{B}\}.\end{aligned} \tag{61}$$

It can be easily seen that

$$\begin{aligned}&P\{S(w)=S_0|\mathfrak{B}\}P\{(u,v,w,S(w),X_1(u|S),\\&\quad X_2(v|S),Y)\in A_\epsilon|S(w)=S_0,\mathfrak{B}\}\\&=\sum_{s'\in\mathbb{S}^n}P\{S(w)=s'|\mathfrak{B}\}P\{S_0=s'|\mathfrak{B}\}\\&\quad\cdot P\{(u,v,w,s',X_1(u|s'),X_2(v|s'),Y)\in A_\epsilon|S_0=s',\mathfrak{B}\},\end{aligned} \tag{62}$$

but since $s'\notin A_\epsilon$ we have

$$P\{(u,v,w,s',X_1(u|s'),X_2(v|s'),Y)\in A_\epsilon|S_0=s',\mathfrak{B}\}=0. \tag{63}$$

Therefore, using this and (60)–(62), we have

$$P\{E_{25}|\mathfrak{B}\}=\sum_{\substack{u\neq u_0,\,v\neq v_0:\\(u,v,w)\in A_\epsilon,\,w\neq w_0}}\ \sum_{s':\,s'\in A_\epsilon}P\{S(w)=s'|\mathfrak{B}\}\cdot P\{S_0=s'|\mathfrak{B}\}P'_{25} \tag{64}$$

where

$$P'_{25} = P\{(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}, \boldsymbol{s}', \boldsymbol{X}_1(\boldsymbol{u}|\boldsymbol{s}'), \boldsymbol{X}_2(\boldsymbol{v}|\boldsymbol{s}'), \boldsymbol{Y}) \in A_\epsilon | \boldsymbol{S}_0 = \boldsymbol{s}', \mathcal{B}\}. \quad (65)$$

By using Appendix A (A23), we can bound P'_{25} by

$$P'_{25} \leq 2^{-n[(X_1, X_2; Y|S) - 8\epsilon]}. \quad (66)$$

On the other hand for $\boldsymbol{s}' \in A_\epsilon$ we have

$$P\{\boldsymbol{S}(\boldsymbol{w}) = \boldsymbol{s}' | \mathcal{B}\} \leq 2^{-n[H(S) - \epsilon]}, \quad (67)$$

and

$$P\{\boldsymbol{S}_0 = \boldsymbol{s}' | \mathcal{B}\} \leq 2^{-n[H(S) - \epsilon]}. \quad (68)$$

Substituting this result in (64), we have

$$P\{E_{25}|\mathcal{B}\} \leq \sum_{\substack{\boldsymbol{u} \neq \boldsymbol{u}_0,\, \boldsymbol{v} \neq \boldsymbol{v}_0: \\ (\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) \in A_\epsilon,\, \boldsymbol{w} \neq \boldsymbol{w}_0}} \sum_{\boldsymbol{s}': \boldsymbol{s}' \in A_\epsilon} 2^{-n[2H(S) - 2\epsilon]} \cdot 2^{-n[I(X_1, X_2; Y|S) - 8\epsilon]}. \quad (69)$$

or

$$P\{E_{25}|\mathcal{B}\} < 2^{-n[I(X_1, X_2; Y|S) + 2H(S) - 10\epsilon]} \cdot \|\{(\boldsymbol{u}, \boldsymbol{v}): (\boldsymbol{u}, \boldsymbol{v}) \in A_\epsilon\}\| \cdot \|\{\boldsymbol{s}': \boldsymbol{s}' \in A_\epsilon\}\|. \quad (70)$$

Substituting

$$\|\{(\boldsymbol{u}, \boldsymbol{v}): (\boldsymbol{u}, \boldsymbol{v}) \in A_\epsilon\}\| \leq 2^{n[H(U,V) + \epsilon]}, \quad (71)$$

$$\|\{\boldsymbol{s}': \boldsymbol{s}' \in A_\epsilon\}\| \leq 2^{n[H(S) + \epsilon]}. \quad (72)$$

into (70), we have

$$P\{E_{25}|\mathcal{B}\} < 2^{n[H(U,V) - I(X_1, X_2; Y|S) - H(S) + 12\epsilon]}. \quad (73)$$

This shows that if

$$H(U,V) < I(X_1, X_2; Y|S) + H(S) - 12\epsilon, \quad (74)$$

then by choosing a sufficiently large n

$$P\{E_{25}|\mathcal{B}\} < \epsilon. \quad (75)$$

Now we prove that inequality (57) dominates inequality (74), thus establishing the redundancy of condition (74). Expand the right side of (74):

$$\begin{aligned} I(X_1, X_2; Y|S) &+ H(S) - 12\epsilon \\ &= H(Y|S) + H(S) - H(Y|X_1, X_2, S) - 12\epsilon \\ &\overset{①}{=} H(Y, S) - H(Y|X_1, X_2) - 12\epsilon \\ &\geq H(Y) - H(Y|X_1, X_2) - 12\epsilon \\ &= I(X_1, X_2; Y) - 12\epsilon, \end{aligned} \quad (76)$$

where in step ①, we have used the fact that S and Y are independent given (X_1, X_2). Using the fact that ϵ is arbitrary, this shows that if (57) is satisfied, then (74) is also satisfied.

The bounds on $P\{E_{2i}|\mathcal{B}\}$ for $i = 1,2,3,4,5$ show that if conditions (30), (32), (42), and (57) are satisfied, we will have (see (22)),

$$P\{E_2|\mathcal{B}\} < 5\epsilon. \quad (77)$$

Finally from (20) we see that

$$\overline{P}_n < 7\epsilon, \quad (78)$$

if the conditions of Theorem 1 are satisfied. This completes the proof of Theorem 1.

IV. An Uncomputable Expression for the Capacity Region

The previous theorem develops so-called single letter characterizations of an achievable rate region for correlated sources sent over a multiple access channel. This region is computable in the sense that it can be calculated to any desired accuracy in finite time. The following theorem exhibits the capacity region but does not lead to a finite computation.

Theorem 2 (Capacity Region): The correlated sources (U, V) can be communicated reliably over the discrete memoryless multiple access channel $(\mathcal{X}_1 \times \mathcal{X}_2, \mathcal{Y}, p(y|x_1, x_2))$ if and only if

$$(H(U|V), H(V|U), H(U,V)) \in \bigcup_{k=1}^{\infty} C_k,$$

where

$$C_k = \left\{(R_1, R_2, R_3): \quad \begin{aligned} R_1 &< \frac{1}{k} I(\boldsymbol{X}_1^k; \boldsymbol{Y}^k | \boldsymbol{U}^k, \boldsymbol{X}_2^k) \\ R_2 &< \frac{1}{k} I(\boldsymbol{X}_2^k; \boldsymbol{Y}^k | \boldsymbol{V}^k, \boldsymbol{X}_1^k) \\ R_3 &< \frac{1}{k} I(\boldsymbol{X}_1^k, \boldsymbol{X}_2^k; \boldsymbol{Y}^k) \end{aligned}\right. \quad (79)$$

for some

$$\left.\prod_{i=1}^{k} p(u_i, v_i) p(\boldsymbol{x}^k | \boldsymbol{u}^k) p(\boldsymbol{x}^k | \boldsymbol{v}^k) \prod_{i=1}^{k} p(y_i | x_{1i}, x_{2i})\right\}.$$

Remark 1: It is easily seen that $C_k \subseteq C_{2k} \subseteq C_{3k} \subseteq \cdots$. In fact, $C_{n+m} \supseteq (m/(m+n))C_m \cup (n/(m+n))C_n$, for all m, n. Also, the sets C_k are uniformly bounded above. Thus, from Gallager [1, Appendix 4A], $\cup_{k=1}^{\infty} C_k = \lim_{k \to \infty} C_k$.

Remark 2: The existence of $C = \lim_{k \to \infty} C_k$ suggests that C is computable. However, there are no evident bounds on the computation error, so, while we know $C \supseteq C_k$, we do not have an upper bound $\overline{C}_k$, $C \subseteq \overline{C}_k$, and hence do not know when C has been defined to sufficient accuracy to terminate the computation.

Proof of Theorem 2:

1) Achievability: Reliable transmission for $\boldsymbol{H}$ in C_k follows immediately from Theorem 1 if we replace the channel by its kth extension.
2) Converse: Given the two correlated sources

$$(\boldsymbol{U}, \boldsymbol{V}) \sim \prod_{i=1}^{n} p(u_i, v_i)$$

and a code book

$$\mathcal{C} = \{(\boldsymbol{x}_1(\boldsymbol{u}), \boldsymbol{x}_2(\boldsymbol{v})): \boldsymbol{u} \in \mathcal{U}^n, \boldsymbol{v} \in \mathcal{V}^n\},$$

we construct the empirical probability mass function on the set $\mathcal{U}^n \times \mathcal{V}^n \times \mathcal{X}_1^n \times \mathcal{X}_2^n \times \mathcal{Y}^n$ defined by

$$p(\boldsymbol{u},\boldsymbol{v},\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{y})=\prod_{i=1}^{n}\left(p(u_i,v_i)p(\boldsymbol{x}_1|\boldsymbol{u})p(\boldsymbol{x}_2|\boldsymbol{v})\right.$$
$$\cdot\prod_{i=1}^{n}p(y_i|x_{1i},x_{2i}). \quad (80)$$

Now, applying Fano's inequality, we obtain

$$(1/n)H(\boldsymbol{U},\boldsymbol{V}|\boldsymbol{Y})\leqslant P_n(1/n)\log\|\mathcal{U}^n\times\mathcal{V}^n\|+1/n$$
$$=P_n(\log\|U\|+\log\|V\|)+1/n\triangleq\lambda_n, \quad (81)$$

where $\|U\|$ and $\|V\|$ are the respective alphabet sizes (assumed finite) of U and V. Thus if $P_n\to 0$, λ_n must converge to zero. Standard inequalities yield

$$\begin{aligned}\text{i)}\quad (1/n)H(\boldsymbol{U}|\boldsymbol{V})&=H(U|V)\\&=(1/n)H(\boldsymbol{U}|\boldsymbol{V},\boldsymbol{X}_2)\\&=(1/n)I(\boldsymbol{U};\boldsymbol{Y}|\boldsymbol{V},\boldsymbol{X}_2)\\&\quad+(1/n)H(\boldsymbol{U}|\boldsymbol{V},\boldsymbol{Y},\boldsymbol{X}_2)\\&\leqslant(1/n)I(\boldsymbol{X}_1;\boldsymbol{Y}|\boldsymbol{V},\boldsymbol{X}_2)+\lambda_n.\end{aligned} \quad (82)$$

Similarly,

$$\text{ii)}\quad H(V|U)\leqslant(1/n)I(\boldsymbol{X}_2;\boldsymbol{Y}|\boldsymbol{U},\boldsymbol{X}_1)+\lambda_n. \quad (83)$$

Finally,

$$\begin{aligned}\text{iii)}\quad H(U,V)&\leqslant(1/n)I(\boldsymbol{U},\boldsymbol{V};\boldsymbol{Y})+\lambda_n\\&\leqslant(1/n)I(\boldsymbol{X}_1,\boldsymbol{X}_2;\boldsymbol{Y})+\lambda_n.\end{aligned} \quad (84)$$

Now, if (U,V) is to be transmitted reliably, then $\lambda_n\to 0$ as $n\to\infty$. It follows from (82), (83), and (84), that

$$(H(U|V),H(V|U),H(U,V))\in\lim_{n\to\infty}C_n,$$

which proves the converse.

Finally, for m correlated sources, we have the following result.

Theorem 3: The correlated sources $\{U_1,U_2,\cdots,U_m\}$ can be communicated reliably over the MAC $(\mathcal{X}_1\times\mathcal{X}_2\times\cdots\times\mathcal{X}_m,\mathcal{Y},p(y|x_1,x_2,\cdots,x_m))$ if and only if there exists some k such that

$$H(\boldsymbol{U}(S)|\boldsymbol{U}(S^c))<(1/k)I(\boldsymbol{X}(S);\boldsymbol{Y}|\boldsymbol{X}(S^c),\boldsymbol{U}(S^c)), \quad (85)$$

for all subsets $S\subseteq\{1,2,\cdots,m\}$.

In Theorem 2, as well as in the previous sections, we assumed that the observed number of source symbols per unit time was equal to the number of channel transmissions per unit time.

We now generalize the problem to allow the observation of R source symbols per channel transmission.

Theorem 4: The correlated sources $\{(U_i,V_i)\}_{i=1}^{\infty}$, arriving at the channel at the rate R symbols per channel use, can be communicated reliably over the discrete memoryless multiple access channel if and only if

$$(H(U|V),H(V|U),H(U,V))\in\bigcup_{n=1}^{\infty}C_n,$$

where

$$C_n=\left\{(R_1,R_2,R_3):\quad \begin{aligned}R_1&<\frac{1}{\lfloor nR\rfloor}I\left(X_1^n;Y^n,U^{[nR]},X_2^n\right)\\R_2&<\frac{1}{\lfloor nR\rfloor}I\left(X_2^n;Y^n,V^{[nR]},X_1^n\right)\\R_3&<\frac{1}{\lfloor nR\rfloor}I(X_1^n,X_2^n;Y^n)\end{aligned}\right. \quad (86)$$

for some

$$\prod_{i=1}^{[Rn]}p(u_i,v_i)p\left(x_1^n(\boldsymbol{u}^{[Rn]})\right)p\left(x_2^n(v^{[Rn]})\right)\prod_{i=1}^{n}p(y_i|x_{1i},x_{2i})\Big\}. \quad (87)$$

Proof: The proof follows easily from that of Theorem 2 by choosing a sequence of integers p_i,q_i such that $p_i/q_i\to R$ and breaking the $(\boldsymbol{U},\boldsymbol{V})$ sequences into blocks of superletters of length p_i and breaking the $\boldsymbol{X}$ sequence into blocks of superletters of length q_i.

Appendix A

In this appendix, we shall bound

$$P\{(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w},\boldsymbol{S}(\boldsymbol{w}),\boldsymbol{X}_1(\boldsymbol{u}|\boldsymbol{S}),\boldsymbol{X}_2(\boldsymbol{v}|\boldsymbol{S}),\boldsymbol{Y})\in A_\epsilon|\mathcal{B}\},$$

under the various assumptions of independence on $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{w}$, $\boldsymbol{s}$, $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, and $\boldsymbol{Y}$ that arise in the proof of Theorem 1. Recall that $(\boldsymbol{u}_0,\boldsymbol{v}_0,\boldsymbol{w}_0)\in A_\epsilon$, where A_ϵ denotes the set of all jointly typical $(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w})$ sequences, and $\mathcal{B}$ denotes the event that this particular $(\boldsymbol{u}_0,\boldsymbol{v}_0)$ is the output of the source. Our bound will hold uniformly for each $(\boldsymbol{u}_0,\boldsymbol{v}_0)\in A_\epsilon$.

First we prove a lemma which is used repeatedly in the proof.

Lemma: Let (Z_1,Z_2,Z_3,Z_4,Z_5) be random variables with joint distribution $p(z_1,z_2,z_3,z_4,z_5)$. Fix $(z_1,z_2)\in A_\epsilon$, and let $\boldsymbol{Z}_3,\boldsymbol{Z}_4,\boldsymbol{Z}_5$ be drawn according to

$$P(\boldsymbol{Z}_3=\boldsymbol{z}_3,\boldsymbol{Z}_4=\boldsymbol{z}_4,\boldsymbol{Z}_5=\boldsymbol{z}_5|\boldsymbol{z}_1,\boldsymbol{z}_2)$$
$$=\prod_{i=1}^{n}p(z_{3i}|z_{1i},z_{2i})p(z_{4i}|z_{3i},z_{2i})p(z_{5i}|z_{3i},z_{1i}). \quad (A1)$$

In other words $\boldsymbol{Z}_3$ depends only on $\boldsymbol{Z}_1,\boldsymbol{Z}_2$; $\boldsymbol{Z}_4$ depends only on $\boldsymbol{Z}_3,\boldsymbol{Z}_2$; and $\boldsymbol{Z}_5$ depends only on $\boldsymbol{Z}_3,\boldsymbol{Z}_1$. Then

$$P\{(\boldsymbol{z}_1,\boldsymbol{z}_2,\boldsymbol{Z}_3,\boldsymbol{Z}_4,\boldsymbol{Z}_5)\in A_\epsilon\}$$
$$\leqslant 2^{-n[I(Z_1;Z_4|Z_2,Z_3)+I(Z_5;Z_2,Z_4|Z_1,Z_3)-8\epsilon]}. \quad (A2)$$

Proof: Since $(\boldsymbol{z}_1,\boldsymbol{z}_2)\in A_\epsilon$, we have

$$P\{(\boldsymbol{z}_1,\boldsymbol{z}_2,\boldsymbol{Z}_3,\boldsymbol{Z}_4,\boldsymbol{Z}_5)\in A_\epsilon\}$$
$$=\sum_{\substack{(\boldsymbol{z}_3,\boldsymbol{z}_4,\boldsymbol{z}_5):\\(\boldsymbol{z}_1,\boldsymbol{z}_2,\boldsymbol{z}_3,\boldsymbol{z}_4,\boldsymbol{z}_5)\in A_\epsilon}}P\{(\boldsymbol{Z}_3,\boldsymbol{Z}_4,\boldsymbol{Z}_5)=(\boldsymbol{z}_3,\boldsymbol{z}_4,\boldsymbol{z}_5)|\boldsymbol{z}_1,\boldsymbol{z}_2\}. \quad (A3)$$

But from (A1)

$$P\{(Z_3,Z_4,Z_5)=(z_3,z_4,z_5)|z_1,z_2\}$$
$$=P\{Z_3=z_3|z_1,z_2\}\cdot P\{Z_4=z_4|z_3,z_2\}\cdot P\{Z_5=z_5|z_3,z_1\}, \quad (A4)$$

and since $(z_1,z_2,z_3,z_4,z_5)\in A_\epsilon$, we have from the AEP

$$P\{Z_3=z_3|z_1,z_2\}\leqslant 2^{-n[H(Z_3|Z_1,Z_2)+2\epsilon]}, \quad (A5)$$
$$P\{Z_4=z_4|z_3,z_2\}\leqslant 2^{-n[H(Z_4|Z_3,Z_2)+2\epsilon]}, \quad (A6)$$
$$P\{Z_5=z_5|z_3,z_1\}\leqslant 2^{-n[H(Z_5|Z_3,Z_1)+2\epsilon]}, \quad (A7)$$

Using (A5)–(A7) and the bound on the cardinality of the set $\{(z_3,z_4,z_5)\colon (z_1,z_2,z_3,z_4,z_5)\in A_\epsilon\}$, we have

$$P\{(z_1,z_2,Z_3,Z_4,Z_5)\in A_\epsilon\}\leqslant 2^{n[H(Z_3,Z_4,Z_5|Z_1,Z_2)-2\epsilon]}$$
$$\cdot 2^{-n[H(Z_3|Z_1,Z_2)+2\epsilon]}2^{-n[H(Z_4|Z_3,Z_2)+2\epsilon]}2^{-n[H(Z_5|Z_3,Z_1)+2\epsilon]}. \quad (A8)$$

Substituting

$$H(Z_3,Z_4,Z_5|Z_1,Z_2)=H(Z_3|Z_1,Z_2)+H(Z_4|Z_1,Z_2,Z_3)+H(Z_5|Z_1,Z_2,Z_3,Z_4) \quad (A9)$$

into (A8) we have

$$P\{(z_1,z_2,Z_3,Z_4,Z_5)\in A_\epsilon\}\leqslant 2^{-n[I(Z_4;Z_1|Z_2,Z_3)+I(Z_5;Z_2,Z_4|Z_1,Z_3)-8\epsilon]}. \quad (A10)$$

This completes the proof.

Now we bound $P\{(\boldsymbol{u},\boldsymbol{v},f(\boldsymbol{u}),S(f(\boldsymbol{u})),X_1(\boldsymbol{u}|S),X_2(\boldsymbol{v}|S),Y)\in A_\epsilon|\mathfrak{B}\}$ in different cases. Note that in all cases we are assuming $(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w})\in A_\epsilon$. We now consider specific conditions.

1) $\boldsymbol{u}\neq\boldsymbol{u}_0$, $\boldsymbol{v}=\boldsymbol{v}_0$ (*therefore* $\boldsymbol{w}=\boldsymbol{w}_0$, $S=S_0$).

Here $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{w}_0$ are fixed and S_0, $X_1(\boldsymbol{u}|S_0)$, $X_2(\boldsymbol{v}|S_0)$, Y are random variables. We use Lemma 1 with $Z_1=(V_0,\boldsymbol{w}_0)$, $z_2=\boldsymbol{u}$, $Z_3=S_0$, $Z_4=X_1(\boldsymbol{u}|S_0)$, $Z_5=(X_2(\boldsymbol{v}_0|S_0),Y)$. Note that the assumption of the lemma on the conditional distribution of Z_3,Z_4,Z_5 given z_1,z_2 are satisfied. In (A10), we have

$$\begin{aligned}I(Z_4;Z_1|Z_3,Z_2)&=I(X_1;V,W|U,S)\\&=H(X_1|U,S)-H(X_1|U,V,W,S)\\&=H(X_1|U,S)-H(X_1|U,S)=0, \quad (A11)\end{aligned}$$

where the last step follows from the fact that X_1 and (V,W) are conditionally independent given (U,S).

We also have

$$\begin{aligned}I(Z_5;Z_2,Z_4|Z_1,Z_3)&=I(X_2,Y;U,X_1|V,W,S)\\&\overset{①}{=}I(X_2,Y;U,X_1|V,S)\\&=H(X_2,Y|V,S)-H(X_2,Y|U,V,X_1,S)\\&\overset{②}{=}H(X_2|V,S)+H(Y|X_2,V,S)\\&\quad-H(X_2|U,V,X_1,S)-H(Y|X_1,X_2)\\&\overset{③}{=}H(X_2|V,S)+H(Y|X_2,V,S)\\&\quad-H(X_2|V,S)-H(Y|X_1,X_2)\\&\overset{④}{=}H(Y|X_2,V,S)-H(Y|X_1,X_2,V,S)\\&=I(Y;X_1|X_2,V,S), \quad (A12)\end{aligned}$$

where each equality is justified by the following reasoning:

1) because W is a deterministic function of V;
2) from the chain rule for conditional entropy and the fact that Y and (U,V,S) are conditionally independent given (X_1,X_2);
3) from the fact that X_2 and (U,X_1) are conditionally independent given (V,S);
4) from the fact that Y and (V,S) are conditionally independent given (X_1,X_2).

From (A10)–(A12) it follows that

$$P\{(\boldsymbol{u},\boldsymbol{v}_0,\boldsymbol{w}_0,S_0,X_1(\boldsymbol{u}|S_0),X_2(\boldsymbol{v}_0|S_0),Y)\in A_\epsilon|\mathfrak{B}\}\leqslant 2^{-n[I(Y;X_1|X_2,V,S)-8\epsilon]}. \quad (A13)$$

2) $\boldsymbol{v}\neq\boldsymbol{v}_0$, $\boldsymbol{u}=\boldsymbol{u}_0$ (*therefore* $\boldsymbol{w}=\boldsymbol{w}_0$, $S=S_0$).

Again we assume $(\boldsymbol{u}_0,\boldsymbol{v},\boldsymbol{w}_0)\in A_\epsilon$. This case is similar to case (A1), and we obtain

$$P\{(\boldsymbol{v},\boldsymbol{u}_0,\boldsymbol{w}_0,S_0,X_1(\boldsymbol{u}_0|S_0),X_2(\boldsymbol{v}|S_0),Y)\in A_\epsilon|\mathfrak{B}\}\leqslant 2^{-n[I(Y;X_2|X_1,U,S)-8\epsilon]}. \quad (A14)$$

3) $\boldsymbol{u}\neq\boldsymbol{u}_0$, $\boldsymbol{v}\neq\boldsymbol{v}_0$ *but* $\boldsymbol{w}=\boldsymbol{w}_0$ (*hence* $S=S_0$).

As usual we are assuming $(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}_0)\in A_\epsilon$. Here $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{w}_0$ are fixed and S_0, $X_1(\boldsymbol{u}|S_0)$, $X_2(\boldsymbol{v}|S_0)$, and Y are random variables. We apply the lemma with $z_1=\boldsymbol{w}_0$, $z_2=(\boldsymbol{u},\boldsymbol{v})$, $Z_3=S_0$, $Z_4=(X_1(\boldsymbol{u}|S_0),X_2(\boldsymbol{v}|S_0))$, $Z_5=Y$. Again, with this choice, the conditions of the lemma on the joint distribution function of Z_3, Z_4, Z_5 given z_1,z_2 are satisfied, and we can apply inequality (A10). We have

$$I(Z_4;Z_1|Z_2,Z_3)=I(X_1,X_2;W|U,V,S)=0, \quad (A15)$$

because W is a deterministic function of U and V. Also

$$\begin{aligned}I(Z_5;Z_2,Z_4|Z_1,Z_3)&=I(Y;U,V,X_1,X_2|W,S)\\&\overset{①}{=}H(Y|W,S)-H(Y|X_1,X_2,W,S)\\&=I(Y;X_1,X_2|W,S), \quad (A16)\end{aligned}$$

where ① follows from the conditional independence of Y and (U,V) given (X_1,X_2). From (A10), (A15), and (A16) it follows that

$$P\{(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}_0,S_0,X_1(\boldsymbol{u}|S_0),X_2(\boldsymbol{v}|S_0),Y)\in A_\epsilon|\mathfrak{B}\}\leqslant 2^{-n[I(X_1,X_2;Y|W,S)-8\epsilon]}. \quad (A17)$$

4) $\boldsymbol{u}\neq\boldsymbol{u}_0$, $\boldsymbol{v}\neq\boldsymbol{v}_0$, $\boldsymbol{w}\neq\boldsymbol{w}_0$, $S_0\neq\boldsymbol{s}'$.

Here $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{w}$, $\boldsymbol{s}'$ are fixed, X_1, X_2, and Y are random variables, and we wish to bound $P\{(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w},\boldsymbol{s}',X_1(\boldsymbol{u}|\boldsymbol{s}'),X_2(\boldsymbol{v}|\boldsymbol{s}'),Y)\in A_\epsilon|S_0\neq\boldsymbol{s}',\mathfrak{B}\}$. It is assumed that $(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w})\in A_\epsilon$ and $\boldsymbol{s}'\in A_\epsilon$. Therefore by the independence of S from U,V,W it follows that $(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w},\boldsymbol{s}')\in A_\epsilon$. In the lemma, let

$$z_1=\varnothing,\quad z_2=(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w},\boldsymbol{s}'),\quad Z_3=\varnothing,$$
$$Z_4=(X_1(\boldsymbol{u}|\boldsymbol{s}'),X_2(\boldsymbol{v}|\boldsymbol{s}'),\quad Z_5=Y.$$

From the lemma, we have

$$I(Z_4;Z_1|Z_2,Z_3)=I(X_1,X_2;\varnothing|U,V,W,S)=0 \quad (A18)$$

and

$$I(Z_5;Z_2,Z_4|Z_1,Z_3)=I(Y;U,V,W,S,X_1,X_2)=I(Y;X_1,X_2). \quad (A19)$$

Hence

$$P\{(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w},\boldsymbol{s}',\boldsymbol{X}_1(\boldsymbol{u}|\boldsymbol{s}'),\boldsymbol{X}_2(\boldsymbol{v}|\boldsymbol{s}'),\boldsymbol{Y})\in A_\epsilon|\boldsymbol{S}_0\neq\boldsymbol{s}',\mathfrak{B}\} \leq 2^{-n[I(X_1,X_2;Y)-8\epsilon]} \quad \text{(A20)}$$

5) $\boldsymbol{u}\neq\boldsymbol{u}_0$, $\boldsymbol{v}\neq\boldsymbol{v}_0$, $\boldsymbol{w}\neq\boldsymbol{w}_0$, $\boldsymbol{S}_0=\boldsymbol{s}'$.

Here, as in (A4), $(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w},\boldsymbol{s}')\in A_\epsilon$ are fixed and $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, and $\boldsymbol{Y}$ are random variables, and we wish to bound

$$P\{(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w},\boldsymbol{s}',\boldsymbol{X}_1(\boldsymbol{u}|\boldsymbol{s}'),\boldsymbol{X}_2(\boldsymbol{v}|\boldsymbol{s}'),\boldsymbol{Y})\in A_\epsilon|\boldsymbol{S}_0=\boldsymbol{s}',\mathfrak{B}\}.$$

In the lemma, set

$$z_1=\boldsymbol{s}', \qquad z_2=(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w},\boldsymbol{s}'), \qquad \boldsymbol{Z}_3=\varnothing,$$
$$\boldsymbol{Z}_4=(\boldsymbol{X}_1(\boldsymbol{u}|\boldsymbol{s}'),\boldsymbol{X}_2(\boldsymbol{v}|\boldsymbol{s}')), \qquad \boldsymbol{Z}_5=\boldsymbol{Y},$$

thus obtaining

$$I(Z_4;Z_1|Z_2,Z_3)=I(X_1,X_2;S|U,V,W,S)=0 \quad \text{(A21)}$$

and

$$\begin{aligned} I(Z_5;Z_2,Z_4|Z_1,Z_3) &= I(Y;U,V,W,S,X_1,X_2|S) \\ &\overset{①}{=} H(Y|S)-H(Y|X_1,X_2,S) \\ &= I(Y;X_1,X_2|S), \end{aligned} \quad \text{(A22)}$$

where step ① follows from the conditional independence of Y and (U,V,W) given (X_1,X_2). Again, from the lemma, we obtain the bound

$$P\{(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w},\boldsymbol{s}',\boldsymbol{X}_1(\boldsymbol{v}|\boldsymbol{s}'),\boldsymbol{X}_2(\boldsymbol{v}|\boldsymbol{s}'),\boldsymbol{Y})\in A_\epsilon|\boldsymbol{S}_0=\boldsymbol{s}',\mathfrak{B}\} \leq 2^{-n[I(X_1,X_2;Y|S)-8\epsilon]}. \quad \text{(A23)}$$

Appendix B
Proof of Convexity in Theorem 1

Let $p_1(s)p_1(x_1|u,s)p_1(x_2|v,s)$ and $p_2(s)p_2(x_1|u,s)\cdot p_2(x_2|v,s)$ be two arbitrary conditional mass functions on $\mathcal{S}\times\mathcal{X}_1\times\mathcal{X}_2$. To show convexity, it suffices to show that for any $\alpha\in[0,1]$, there exists a conditional mass function $p(s')p(x_1|u,s')p(x_2|v,s')$ such that

$$\alpha I_1(X_1;Y|X_2,V,S)+(1-\alpha)I_2(X_1;Y|X_2,V,S) \leq I(X_1;Y|X_2,V,S'), \quad \text{(B1)}$$

$$\alpha I_1(X_2;Y|X_1,U,S)+(1-\alpha)I_2(X_2;Y|X_1,U,S) \leq I(X_2;Y|X_1,U,S'), \quad \text{(B2)}$$

and

$$\alpha I_1(X_1,X_2;Y)+(1-\alpha)I_2(X_1,X_2;Y) \leq I(X_1,X_2;Y), \quad \text{(B3)}$$

where the subscripts on the I refer to the conditional mass function used.

Define the independent random variable T, taking the value 1 with probability α and 2 with probability $1-\alpha$. let $S'=(S,T)$ and observe that

$$(\text{B1})=I(X_1;Y|X_2,V,S'),$$
$$(\text{B2})=I(X_2;Y|X_1,U,S'),$$

and

$$(\text{B3})=I(X_1,X_2;Y|T)\leq I(X_1,X_2;Y),$$

thus establishing convexity.

References

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

[2] H. Liao, "A coding theorem for multiple access communications," presented at the International Symp. on Information Theory, Asilomar, 1972. Also Ph.D. dissertation, "Multiple access channels," Dept. of Electrical Engineering, University of Hawaii, 1972.

[3] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 471–480, July 1973. Reprinted in *Key Papers in the Development of Information Theory*, D. Slepian, Ed. New York: IEEE, 1974, pp. 450–459.

[4] 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, Sept. 1973.

[5] T. Cover, "An achievable rate region for the broadcast channel," *IEEE Trans. Inform. Theory*, vol. IT-21, no. 4, pp. 399–404, July 1975.

[6] T. 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. Reprinted in *Ergodic and Information Theory*, L. Davisson and R. Gray eds.), Benchmark Papers in Electrical Engineering and Computer Science, Dowden, Hutchinson, and Ross, Penn.

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