MULTIPLE ACCESS CHANNELS WITH ARBITRARILY CORRELATED SOURCES

BY

THOMAS M. COVER
ABBAS EL GAMAL
MASOUD SALEHI

TECHNICAL REPORT NO. 41
NOVEMBER 1979

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION
GRANT ENG 78-23334

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
MULITPLE ACCESS CHANNELS WITH ARBITRARILY CORRELATED SOURCES

by

Thomas M. Cover
Abbas El Gamal
Masoud Salehi

TECHNICAL REPORT NO. 41
NOVEMBER 1979

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION
GRANT ENG 78-23334

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
MULTIPLE ACCESS CHANNELS WITH ARBITRARILY CORRELATED SOURCES

Thomas M. Cover\textsuperscript{†}
Abbas El Gamal\textsuperscript{‡‡}
Masoud Salehi\textsuperscript{§§§}

Abstract

Let \( \{(U_i, V_i)\}_{i=1}^n \) be a source of 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. We prove that such a source can be sent with arbitrarily small probability of error over a multiple access channel

\[
\{ \overline{X}_1 \times \overline{X}_2, Y, p(y|x_1, x_2) \}
\]

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

\[
H(U|V) \leq I(X_1; Y|X_2, V, S)
\]

\[
H(V|U) \leq I(X_2; Y|X_1, U, S)
\]

\[
H(U,V|W) \leq I(X_1, X_2; Y|W, S)
\]

\[
H(U,V) \leq I(X_1, X_2; Y)
\]

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.

\textsuperscript{†}Stanford University, Stanford, CA. This work was partially supported by NSF Grant ENG 76-23334 and SRI International Contract DAHC-15-C-0187, and JSEP Contract N00014-75-C-0601.

\textsuperscript{‡‡}University of Southern California, Los Angeles, CA. This work was partially supported by JSEP Contract F44620-76-C-0061.

\textsuperscript{§§§}Stanford University, Stanford, CA.
1. Introduction

The multiple access channel \( p(y|x_1,x_2) \) 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

\[
R_1 \leq I(X_1;Y|X_2) \\
R_2 \leq I(X_2;Y|X_1) \\
R_1 + R_2 \leq I(X_1,X_2;Y).
\]

(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 \( p(x_1,x_2) = p(x_1)p(x_2) \),

\[
H(U) \leq I(X_1;Y|X_2) \\
H(V) \leq I(X_2;Y|X_1) \\
H(U) + H(V) \leq I(X_1,X_2;Y).
\]

(2)

In this paper, we increase this achievable region in 2 ways: 1) the left hand side will be made smaller\(^\dagger\), and 2) the right hand 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

\[^\dagger\text{This improvement could be obtained from the results of Slepian and Wolf [3].}\]
\[ H(U|V) \leq I(X_1; Y|X_2, V) \]
\[ H(V|U) \leq I(X_2; Y|X_1, U) \]
\[ H(U, V) \leq I(X_1, X_2; Y) \] (3)

for some \( p(u, v, x_1, x_2, y) = p(u, v)p(x_1|u)p(x_2|v)p(y|x_1, x_2) \). 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 \((U, V) \sim \Pi p(u_1, v_1)\) can be sent with arbitrarily small probability of error over a multiple access channel \(\{X_1 \times X_2, Y, p(y|x_1, x_2)\}\), 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

\[ H(U|V) \leq I(X_1; Y|X_2, V, S) \]
\[ H(V|U) \leq I(X_2; Y|X_1, U, S) \]
\[ H(U, V|W) \leq I(X_1, X_2; Y|W, S) \]
\[ H(U, V) \leq I(X_1, X_2; Y) \],

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

**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 \(\min\{||X_1||, ||X_2||, ||Y||\}\).

**Example for Theorem 1:**

Consider the transmission of the correlated sources \((U, V)\) with the joint distribution \(p(u, v)\) given by
over the multiple access channel defined by

\[ X_1 = X_2 = \{0, 1\} \]
\[ Y = \{0, 1, 2\}, \]
\[ Y = X_1 + X_2. \]

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 source 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 $2^{nH(U)}$ $X_1$'s associated with the typical $U$'s and $2^{nH(V)}$ $X_2$'s associated with the typical $V$'s, there are only $2^{nH(U,V)}$ pairs $(X_1(U), X_2(V))$ 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 hand side of (3) collapses, yielding the known rate region

$$H(U|V) = I(X_1;Y|X_2,V) = H(X_1) \quad (=R_1)$$

$$H(V|U) = I(X_2;Y|X_1,U) = H(X_2) \quad (=R_2)$$

$$H(U,V) = I(X_1, X_2; Y) = H(X_1) + H(X_2) \quad (=R_1 + R_2).$$

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

$$H(U|V) = H(U) = R_1 \leq I(X_1;Y|X_2)$$

$$H(V|U) = H(V) = R_2 \leq I(X_2;Y|X_1)$$

$$H(U,V) = H(U) + H(V) = R_1 + R_2 \leq I(X_1,X_2;Y).$$
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

$$R \leq I(X_1,X_2;Y),$$

(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 3 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 $p(u,v,s,x_1,x_2,y) = p(u)p(v)p(w)p(s)p(x_1|s)p(x_2|s)$ $p(y|x_1,x_2)$, where $u = (u',w)$, $v = (v',w)$. We then have achievability of $(R_0,R_1,R_2)$ if

$$H(U|V) = H(U') = R_1 \leq I(X_1;Y|X_2,S)$$

$$H(V|U) = H(V') = R_2 \leq I(X_2;Y|X_1,S)$$

(8)

$$H(U,V|W) = H(U') + H(V') = R_1 + R_2 \leq I(X_1,X_2;Y|S)$$

$$H(U,V) = H(U') + H(V') + H(W) = R_0 + R_1 + R_2 \leq I(X_1,X_2;Y).$$
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 3. An expression for source-channel capacity is given in Section 4, but does not satisfy the "single-letter" conditions that we seek.
2. Definition of the Problem

Assume we have two information sources \( U_1, U_2, \ldots \) and \( V_1, V_2, \ldots \) 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 2 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 \( g \)

\[
f: U \rightarrow \{1, 2, \ldots, k\} \\
g: V \rightarrow \{1, 2, \ldots, k\}
\]

with \( P\{f(U) = i\} > 0 \), \( P\{g(V) = i\} > 0 \), \( i = 1, 2, \ldots, k \), 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 \( k \geq 2 \).

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 3.

Assume given a discrete memoryless multiple access channel \((X_1 \times X_2, Y, p(y|x_1,x_2))\), such that the first transmitter has access to the \( U \) process and the second transmitter has access to the \( V \) process and each transmitter encodes the process that it can observe and transmits it over the channel.

See Figure 1. The receiver wants to estimate the source outputs. We are interested in the conditions under which error-free transmission is possible.
A block code for the channel consists of an integer \( n \), and two encoding functions

\[ x_1^n : u^n + x_1^n \]

\[ x_2^n : v^n + x_2^n \]

assigning codewords to the source outputs, and a decoding function

\[ d^n : y^n + u^n \times v^n . \]  \hspace{1cm} (9)

The probability of error is given by

\[ p_n = P \{ (u^n, v^n) \neq d^n(y^n) \} \]

\[ = \sum_{(u,v) \in u^n \times v^n} p(u^n, v^n) p \{ d^n(y^n) \neq (u^n, v^n) | (u^n, v^n) = (u^n, v^n) \} . \]  \hspace{1cm} (10)

where the joint probability mass function is given, for a code assignment \{\( x_1(u^n), x_2(v^n) \)\}, by

\[ p(u, v, y) = \prod_{i=1}^{n} p(u_i, v_i) p(y_i | x_1(u^n), x_2(v^n)) . \]  \hspace{1cm} (11)
Definition: The source \( (U,V) \sim \prod p(u_i,v_i) \) can be reliably transmitted over the multiple access channel \( (X_1 \times X_2, Y, p(y|x_1,x_2)) \) if there exists a sequence of block codes \( \{ x_1^n(u^n), x_2^n(v^n) \} \), \( d^n(y^n) \) such that

\[
P_n = p\{ d^n(y^n) = (u^n,v^n) \} \to 0 .
\]

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

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 \( p(u,v)p(x_1|u)p(x_2|v)p(y|x_1,x_2) \),

\[
H(U|V) < I(X_1;Y|X_2,V) \tag{12}
\]

\[
H(V|U) < I(X_2;Y|X_1,U)
\]

\[
H(U,V) < I(X_1,X_2;Y)
\]

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 \( u \in U^n \) generate one \( x_1 \) sequence drawn according to \( \prod_{i=1}^n p(x_1|u_i) \) and for each \( v \in V^n \) generate one \( x_2 \) sequence drawn according to \( \prod_{i=1}^n p(x_2|v_i) \). Call these sequences \( x_1(u) \) and \( x_2(v) \) respectively.

Encoding: Transmitter 1, upon observing \( u \) at the output of source 1, transmits \( x_1(u) \) and transmitter 2, after observing \( v \) at the output of
source 2, transmits $x_2(v)$. Assume the maps $x_1(\cdot), x_2(\cdot)$ are known to the receiver.

**Decoding:** Upon receiving $y$, the decoder finds the only $(u, v)$ pair such that $(u, v, x_1(u), x_2(v), y) \in A_\epsilon$, where $A_\epsilon$ is the set of jointly $\epsilon$-typical sequences. If there is no such $(u, v)$ pair, or there exists more than one such pair, the decoder declares an error. A helpful picture is given in Figure 2.

![Figure 2: Picture of Joint Typicality for Multiple Access Channel](image)

The 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.

**Error:** Suppose $(u_0, v_0)$ is the source output. Then an error is made if

1) $(u_0, v_0, x_1(u_0), x_2(v_0), y) \notin A_\epsilon$,

or 2) There exists some $(u, v) \neq (u_0, v_0)$ such that $(u, v, x_1(u), x_2(v), y) \in A_\epsilon$.
Then the probability of error $P_n$ can be bounded as follows:

$$
P_n = P \{ (U_0, V_0) \in A_\varepsilon \} + P \{ \exists (u', v') \neq (U_0, V_0): (u', v', X_1(u'), X_2(v')) \in A_\varepsilon, (U_0, V_0) \in A_\varepsilon \}$$

$$
\leq \varepsilon + \sum_{(u_0, v_0) \in A_\varepsilon} \sum_{\substack{u' \neq u_0 \atop v' = v_0}} P(\cdot) + \sum_{(u_0, v_0) \in A_\varepsilon} \sum_{\substack{u' \neq u_0 \atop v' \neq v_0}} P(\cdot)
$$

$$
\leq \varepsilon + 2^{2H(U|V)+\varepsilon} - n(I(X_1; Y|X_2, V) - \varepsilon) + 2^{H(V|U)+\varepsilon} - n(I(X_2; Y|X_1, U) - \varepsilon)
$$

$$
+ 2^{nH(U, V) - n(I(X_1, X_2; Y) - \varepsilon)}.
$$

Consequently, $P_n \to 0$ if the conditions in (12) are satisfied.
3. 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 \( w = f(u) = (f(u_1), \ldots, f(u_n)) \)
and independently generate one \( x_1 \) sequence according to \( \prod_{i=1}^{n} p(x_1|u_i,s_i(w)) \).
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 \( \prod_{i=1}^{n} p(x_2|v_i,s_i(w)) \), is repeated for the
\( v \) sequence. These sequences are indexed by \( x_2(v|s(g(v))) \), or for simpli-
city 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
\( x_1^n \) and \( x_2^n \) 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 \( v \), the decoder de-
clare \((\hat{u},\hat{v})\) to be the transmitted source sequence pair if \((\hat{u},\hat{v})\) is the
unique pair \((u,v)\) such that
\[(u, v, w, z(w), x_1(u|s), x_2(v|s), y) \in A_\varepsilon,\]

where \( w = f(u) \).

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

i) \[(u_0, v_0, w_0, z(w_0), x_1(u_0|s), x_2(v_0|s), y) \in A_\varepsilon,\]

or ii) There exists some \((u, v) \neq (u_0, v_0)\) such that

\[(u, v, w, z(w), x_1(u|s), x_2(v|s), y) \in A_\varepsilon.\]

**Analysis of the Probability of Error**: Letting \( A_\varepsilon \) denote the appropriate set of jointly \(\varepsilon\)-typical sequences (see \([5, 6]\)), we have

\[
\overline{p}_n = \sum_{(u, v) \in U^n \times V^n} p(u, v) P\{\text{error made at decoder} | (u, v) \text{ is the output of the source}\}. \tag{14}
\]

or

\[
\overline{p}_n \leq \sum_{(u, v) \in A_\varepsilon} p(u, v) P\{\text{error made at decoder} | (u, v) \text{ is the output of the source}\} + \sum_{(u, v, w) \notin A_\varepsilon} p(u, v). \tag{15}
\]

From the AEP, for sufficiently large \( n \),

\[
\overline{p}_n \leq \sum_{(u, v, w) \in A_\varepsilon} p(u, v) P\{\text{error made at decoder} | (u, v) \text{ is the output of the source}\} + \varepsilon. \tag{16}
\]

Now we show that as long as \((u, v, w) \in A_\varepsilon\), there exists an upper bound independent of \((u, v)\) for the terms in the summation. To show this, we assume that \((u_0, v_0, w_0) \in A_\varepsilon\) and let \( B \) denote the event that this special pair is the output of the source. We are interested in an upper bound for

\[P\{\text{error made at decoder} | B\}.
\]

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

$$E = E_1 \cup E_2,$$

(17)

where $E_1$: The event that $(u_0, v_0, w_0, S_0, X_1(u|S_0), X_2(v|S_0), Y) \notin A_\varepsilon$,

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

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

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

$$P\{E_1|B\} \leq \varepsilon,$$

(18)

and therefore by the union bound

$$P\{E|B\} \leq P\{E_2|B\} + \varepsilon.$$  

(19)

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

$$P_n \leq P\{E_2|B\} + 2\varepsilon.$$  

(20)

We decompose the event $E_2$ by

$$E_2 = E_{21} \cup E_{22} \cup E_{23} \cup E_{24} \cup E_{25}$$

(21)

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

$$(u, v_0, w_0, S_0, X_1(u|S_0), X_2(v_0|S_0), Y) \in A_\varepsilon.$$

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

$$(u_0, v, w_0, S_0, X_1(u_0|S_0), X_2(v|S_0), Y) \in A_\varepsilon.$$
\( E_{23} \): The event that there exists a \( u \neq u_0 \) and a \( v \neq v_0 \) such that 
\[
f(u) = g(v) = w_0 \quad \text{and} \quad (u,v,w_0,S_0,X_1(u|S_0),X_2(v|S_0),Y) \in A_\epsilon.
\]

\( E_{24} \): The event that there exists a \( u \neq u_0 \) and a \( v \neq v_0 \) such that 
\[
w = f(u) = g(v) \neq w_0, S(f(u)) \neq S_0
\]
and
\[
(u,v,w,S(w),X_1(u|S),X_2(v|S),Y) \in A_\epsilon.
\]

\( E_{25} \): The event that there exists a \( u \neq u_0 \) and a \( v \neq v_0 \) such that 
\[
w = f(u) = g(v) \neq w_0, S(f(u)) = S_0
\]
and
\[
(u,v,w,S(w),X_1(u|S),X_2(v|S),Y) \in A_\epsilon.
\]

By the union bound, we have
\[
P\{E_2|B\} \leq \sum_{i=1}^{5} P\{E_{2i}|B\}. \tag{22}
\]

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

**Bound for \( P\{E_{21}|B\} \):** We have
\[
P\{E_{21}|B\} = P\{ \exists \ u \neq u_0 : (u,v_0,w_0,S_0,X_1(u|S_0),X_2(v_0|S_0),Y) \in A_\epsilon|B \}. \tag{23}
\]

Therefore,
\[
P\{E_{21}|B\} = \sum_{u \neq u_0 : (u,v_0,w_0) \in A_\epsilon} P\{ (u,v_0,w_0,S_0,X_1(u|S_0),X_2(v_0|S_0),Y) \in A_\epsilon|B \}. \tag{24}
\]

From Appendix A (A-13) we have for \( (u,v_0,w_0) \in A_\epsilon \),
\[
P\{ (u,v_0,w_0,S_0,X_1(u|S_0),X_2(v_0|S_0),Y) \in A_\epsilon|B \} \leq \frac{1}{2} \cdot \frac{1}{n}[I(X_1;Y|X_2,V,S)-8\epsilon]. \tag{25}
\]
Notice that this bound is independent of \( u \) as long as \( (u, y_0) \in A_\varepsilon \). Substituting (25) into (24), we have

\[
P(E_{21} | B) \leq \sum_{u \neq u_0}^{(u, y_0, w_0) \in A_\varepsilon} -n[I(X_1; Y | X_2, V, S) - 8\varepsilon] \cdot \| \{ u : (u, y_0, w_0) \in A_\varepsilon \} \|.
\]

(26)

or

\[
P(E_{21} | B) \leq 2^{-n[I(X_1; Y | X_2, V, S) - 8\varepsilon]} \cdot \| \{ u : (u, y_0, w_0) \in A_\varepsilon \} \|.
\]

(27)

But typicality yields

\[
\| \{ u : (u, y_0, w_0) \in A_\varepsilon \} \| \leq 2^{-n[H(U | V, W) + 2\varepsilon]}.
\]

(28)

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

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

(29)

Thus if

\[
H(U | V) < I(X_1; Y | X_2, V, S) - 10\varepsilon,
\]

then for large enough \( n \), we have

\[
P(E_{21} | B) \leq \varepsilon.
\]

(31)

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

\[
H(V | U) < I(X_2; Y | X_1, U, S) - 10\varepsilon,
\]

then by choosing \( n \) sufficiently large, we have

-17-
\[ P\{E_{22}|B\} \leq \varepsilon. \] \hspace{1cm} (33)

**Bound for** \( P\{E_{23}|B\} : \) Here we have

\[
P\{E_{23}|B\} = P\{ \exists \ u \neq u_0, v \neq v_0 : f(u) = g(v) = w_0 \text{ and } \]
\[(u, v, w_0, S_0, X_1(u|S_0), X_2(v|S_0), y) \in A_{\varepsilon}|B \}. \] \hspace{1cm} (34)

Therefore,

\[
P\{E_{23}|B\} = \sum_{u \neq u_0, \ v \neq v_0} \sum_{(u, v, w_0) \in A_{\varepsilon}} P\{ (u, v, w_0, S_0, X_1(u|S_0), X_2(v|S_0), y) \in A_{\varepsilon}|B \}. \] \hspace{1cm} (35)

Again, note that \( u, v, \) and \( w_0 \) are fixed and \( S_0, X_1, X_2 \) and \( y \) are random variables. Using Appendix A (A-17) we have

\[
P\{ (u, v, w_0, S_0, X_1(u|S_0), X_2(v|S_0), y) \in A_{\varepsilon}|B \} \leq 2^{-n[I(X_1, X_2; Y|W, S)-8\varepsilon]} \] \hspace{1cm} (36)

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

\[
P\{E_{23}|B\} \leq \sum_{u \neq u_0, \ v \neq v_0} 2^{-n[I(X_1, X_2; Y|W, S)-8\varepsilon]}, \] \hspace{1cm} (37)

or

\[
P\{E_{23}|B\} \leq 2^{-n[I(X_1, X_2; Y|W, S)-8\varepsilon]} \|\{(u, v)(u, v, w_0) \in A_{\varepsilon} : u \neq u_0, v \neq v_0\}\|. \] \hspace{1cm} (38)

On the other hand, we have

\[
\{(u, v)(u, v, w_0) \in A_{\varepsilon} : u \neq u_0, v \neq v_0\} = \{(u, v)(u, v, w_0) \in A_{\varepsilon}\}, \] \hspace{1cm} (39)

and
\[ \left\| \{ (u,v):(u,v,w) \in A_\varepsilon \} \right\| \leq 2^{n[H(U,V|W)+2\varepsilon]} . \]  

(40)

Using (38), (39), and (40), we obtain

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

(41)

Thus if

\[ H(U,V,W) < I(X_1,X_2;Y|W,S)-10\varepsilon , \]  

(42)

then by choosing \( n \) large enough, we can make

\[ P \{ E_{23} | B \} \leq \varepsilon . \]  

(43)

**Bound for \( P \{ E_{24} | B \} \):** Recall from the definition of \( E_{24} \) that

\[ P \{ E_{24} | B \} = P \{ \exists u \neq u_0, v \neq v_0 : w = f(u) = g(v) = w_0, S(f(u)) = S_0 \text{ and} \]  

\[ (u,v,w,S(w),S(f(u)),X_1(u|S),X_2(v|S),Y) \in A_\varepsilon | B \}. \]  

(44)

from which we have

\[ P \{ E_{24} | B \} = \sum_{u \neq u_0, v \neq v_0 : (u,v,w) \in A_\varepsilon, w = w_0} P \{ S(w) = S_0 \text{ and} (u,v,w,S(w),X_1(u|S),X_2(v|S),Y) \in A_\varepsilon | B \} \]  

(45)

But, by the chain rule,

\[ P \{ S(w) = S_0 \text{ and} (u,v,w,S(w),X_1(u|S),X_2(v|S),Y) \in A_\varepsilon | B \} \]  

\[ = P \{ S(w) = S_0 | B \} P \{ (u,v,w,S(w),X_1(u|S),X_2(v|S),Y) \in A_\varepsilon | S(w) = S_0, B \} . \]  

(46)

Therefore
\[
P\{ S(w) \neq S_0 \text{ and } (u,v,w,S(w),X_1(u|S),X_2(v|S),Y) \in A_\epsilon | B \} \\
\leq P\{ (u,v,w,S(w),X_1(u|S),X_2(v|S),Y) \in A_\epsilon | S(w) \neq S_0, B \}. \quad (47)
\]

But

\[
P\{ (u,v,w,s',S(w),X_1(u|S),X_2(v|S),Y) \in A_\epsilon | S(w) \neq S_0, B \}
\]

\[
= \sum_{s' \in S^n} P\{ (u,v,w,s',X_1(u|s'),X_2(v|s'),Y) \in A_\epsilon | s_0 \neq s', B \} \cdot P\{ S(w) = s'|B \}
\]

\[
= \sum_{s' \in A_\epsilon} P\{ (u,v,w,s',X_1(u|s'),X_2(v|s'),Y) \in A_\epsilon | s_0 \neq s', B \} \cdot P\{ S(w) = s'|B \},
\]

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

\[
P\{ (u,v,w,s',X_1(u|s'),X_2(v|s'),Y) \in A_\epsilon | s_0 \neq s', B \} = 0.
\]

From Appendix A (A-20) for \( s' \in A_\epsilon \), we have

\[
P\{ (u,v,w,s',X_1(u|s'),X_2(v|s'),Y) \in A_\epsilon | s_0 \neq s', B \} \leq 2^{-n[I(X_1,X_2;Y)-8\epsilon]}.
\]

Therefore

\[
P\{ (u,v,w,S(w),X_1(u|S),X_2(v|S),Y) \in A_\epsilon | S(w) \neq S_0, B \}
\]

\[
\leq \sum_{s' \in A_\epsilon} 2^{-n[I(X_1,X_2;Y)-8\epsilon]} \cdot 2^{-n[H(S)+\epsilon]}.
\]

Using the fact that

\[
\|\{ s' : s' \in A_\epsilon \} \| \leq 2^{n[H(S)+\epsilon]}, \quad (51)
\]
we have
\[ p \{ (u, v, w, S(w), X_{1}(u|S), X_{2}(v|S), Y) \in A_{\varepsilon} | S(w) \neq S_{0}, B \} \leq 2^{-n[I(X_{1}, X_{2}; Y) - 8\varepsilon]} . \]  
(52)

Substituting this result into (46) and then into (49) we have
\[ P \{ E_{24}|B \leq \sum_{u \neq u_{0}, v \neq v_{0}} 2^{-n[I(X_{1}, X_{2}; Y) - 8\varepsilon]} \] 
(53)

or
\[ P \{ E_{24}|B \} < 2^{-n[I(X_{1}, X_{2}; Y) - 8\varepsilon]} \cdot ||\{ (u, v): (u, v) \in A_{\varepsilon} \}|| . \]  
(54)

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

Hence,
\[ P \{ E_{24}|B \} < 2^{-n[H(U, V) - I(X_{1}, X_{2}; Y) + 9\varepsilon]} . \]  
(56)

From this inequality, it follows that if
\[ H(U, V) < I(X_{1}, X_{2}; Y) - 9\varepsilon , \]  
(57)

then we can choose \( n \) sufficiently large that
\[ P \{ E_{24}|B \} < \varepsilon . \]  
(58)

Bound for \( P \{ E_{25}|B \} \): Recall from the definition of \( E_{25} \) that
\[ P \{ E_{25}|B \} = P \{ \exists u \neq u_{0}, v \neq v_{0}; w = f(u) = g(v) \neq w_{0}, S(w) = S_{0}, (u, v, w, S(w), X_{1}(u|S), X_{2}(v|S), Y) \in A_{\varepsilon}|B \} . \]  
(59)
Here, as in the previous cases, we can write,

\[
\mathbb{P}\{E_{25}|B\} = \sum_{u \neq u_0, v \neq v_0: (u,v,w) \in A_{\epsilon}, w \neq w_0} \mathbb{P}\{S(w) = S_0 \text{ and } (u,v,w,S(w),X_1(u|S),X_2(v|S),y) \in A_{\epsilon}|B\}.
\]  \hspace{1cm} (60)

But by the chain rule, we have

\[
\mathbb{P}\{S(w) = S_0 \text{ and } (u,v,w,S(w),X_1(u|S),X_2(v|S),y) \in A_{\epsilon}|B\} = \mathbb{P}\{S(w) = S_0|B\} \cdot \mathbb{P}\{(u,v,w,S(w),X_1(u|S),X_2(v|S),y) \in A_{\epsilon}|S(w) = S_0,B\}.
\]  \hspace{1cm} (61)

It can be easily seen that

\[
\mathbb{P}\{S(w) = S_0|B\} \cdot \mathbb{P}\{(u,v,w,S(w),X_1(u|S),X_2(v|S),y) \in A_{\epsilon}|S(w) = S_0,B\} = \sum_{s \in S^n} \mathbb{P}\{S(w) = s'|B\} \cdot \mathbb{P}\{S_0 = s'|B\} \cdot \mathbb{P}\{(u,v,w,s',X_1(u|s'),X_2(v|s'),y) \in A_{\epsilon}|S_0 = s',B\}.
\]  \hspace{1cm} (62)

But since \( s' \notin A_{\epsilon} \) we have

\[
\mathbb{P}\{(u,v,w,s',X_1(u|s'),X_2(v|s'),y) \in A_{\epsilon}|S_0 = s',B\} = 0.
\]  \hspace{1cm} (63)

Therefore, using this and (60), (61), and (62), we have

\[
\mathbb{P}\{E_{25}|B\} = \sum_{u \neq u_0, v \neq v_0: (u,v,w) \in A_{\epsilon}, w \neq w_0} \sum_{s':s' \in A_{\epsilon}} \mathbb{P}\{S(w) = s'|B\} \cdot \mathbb{P}\{S_0 = s'|B\} \cdot \mathbb{P}_{25}.
\]  \hspace{1cm} (64)

where

\[
\mathbb{P}_{25} = \mathbb{P}\{(u,v,w,s',X_1(u|s'),X_2(v|s'),y) \in A_{\epsilon}|S_0 = s',B\}.
\]  \hspace{1cm} (65)
By using Appendix A (A-23), we can bound $P_{25}$ by

$$P_{25} \leq 2^{-n[(X_1, X_2; Y|S) - 8\varepsilon]}$$

(66)

On the other hand, for $s' \in A_\varepsilon$, we have

$$P\{ S(w) = s' | B \} \leq 2^{-n[H(S) - \varepsilon]}$$

(67)

and

$$P\{ S_0 = s' | B \} \leq 2^{-n[H(S) - \varepsilon]}$$

(68)

Substituting this result in (64), we have

$$P\{ E_{25} | B \} \leq \sum_{u \neq u_0, v \neq v_0, s' : s' \in A_\varepsilon} \sum_{(u, v, w) \in A_\varepsilon, w \neq w_0} 2^{-n[2H(S) - 2\varepsilon]} \cdot 2^{-n[I(X_1, X_2; Y|S) - 8\varepsilon]}$$

(69)

or

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

(70)

Substituting

$$\| \{ (u, v) : (u, v) \in A_\varepsilon \} \| \leq 2^{-n[H(U, V) + \varepsilon]}$$

(71)

$$\| \{ s' : s' \in A_\varepsilon \} \| \leq 2^{-n[H(S) + \varepsilon]}$$

(72)

into (70), we have

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

(73)
This shows that if
\[ H(U,V) < I(X_1,X_2;Y|S) + H(S) - 12 \varepsilon \ , \quad (74) \]
then by choosing a sufficiently large \( n \)
\[ P\{ E_{25} | B \} < \varepsilon . \quad (75) \]

Now we prove that inequality (57) dominates inequality (74), thus establishing the redundancy of condition (74). Expand the right hand side of (74):
\[
I(X_1,X_2;Y|S) + H(S) - 12 \varepsilon = H(Y|S) + H(S) - H(Y|X_1,X_2,S) - 12 \varepsilon \\
\geq H(Y) - H(Y|X_1,X_2) - 12 \varepsilon \\
= I(X_1,X_2;Y) - 12 \varepsilon ,\quad (76)
\]
where in step 1, we have use the fact that \( S \) and \( Y \) are independent given \( (X_1,X_2) \). Using the fact that \( \varepsilon \) is arbitrary, this shows that if (57) is satisfied, then (74) is also satisfied.

The bounds on \( P\{ E_{2i} | B \} \) for \( i = 1,2,3,4,5 \) show that if conditions (30), (32), (42), and (57) are satisfied, we will have (see (22)),
\[ P\{ E_2 | B \} < 5 \varepsilon . \quad (77) \]

Finally from (20) we see that
\[ \bar{p}_n < 7 \varepsilon , \quad (78) \]
if the conditions of Theorem 1 are satisfied. This completes the proof of Theorem 1.
4. 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 \((X_1 \times X_2, Y, p(y|x_1, x_2))\) if and only if

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

where

\[
C_k = \{ (R_1, R_2, R_3) : R_1 < \frac{1}{k} I(x_1^k; y^k | u^k, x_2^k) \\
R_2 < \frac{1}{k} I(x_2^k; y^k | y^k, x_1^k) \\
R_3 < \frac{1}{k} I(x_1^k, x_2^k; y^k) \}
\]

for some \(p(u_i, v_i)p(x_i^k | u_i^k)p(x_i^k | v_i^k)\) \(i=1, \ldots, k\) \(p(y_i | x_1^i, x_2^i)\).

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

**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 \( \mathcal{C} \supseteq \mathcal{C}_k \), we do not have an upper bound \( \mathcal{C}_k \), \( \mathcal{C} \subseteq \mathcal{C}_k \), and hence do not know when \( C \) has been defined to sufficient accuracy to terminate the computation.

Proof of Theorem 2:

1. Achievability: The achievability of each \( \mathcal{C}_k \) follows immediately from Theorem 1 if we replace the channel by its \( k \)th extension.

2. Converse:

Given the two correlated sources

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

and a code book

\[
\mathcal{C} = \{ (x_1^u, x_2^v) : u \in \mathcal{U}^n, \ v \in \mathcal{V}^n \}
\]

we construct the empirical probability mass function on the set \( \mathcal{U}^n \times \mathcal{V}^n \times x_1^n \times x_2^n \times \mathcal{Y}^n \) defined by

\[
p(u, v, x_1, x_2, y) = \prod_{i=1}^{n} p(u_i, v_i) p(x_1 | x_i) p(x_2 | y) p(y_i | x_1, x_2, y).
\]  

(80)

Now, applying Fano's inequality, we obtain

\[
1/n \ H(U, V | Y) \leq P_n \frac{1}{n} \log ||U^n \times V^n|| + 1/n
\]

\[
= P_n (\log ||U|| + \log ||V||) + 1/n \frac{\Delta}{\lambda_n},
\]

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

Standard inequalities yield

\begin{align*}
(i) \quad (1/n)H(U|V) &= H(U|V) \\
&= (1/n) H(U|V, X_2) \\
&= (1/n) I(U; Y|V, X_2) + (1/n)H(U|V, X_2) \\
&\leq (1/n) I(X_1; Y|V, X_2) + \lambda_n. \quad (82)
\end{align*}

Similarly,

\begin{align*}
(ii) \quad H(V|U) &\leq (1/n) I(X_2; Y|U, X_1) + \lambda_n. \quad (83)
\end{align*}

Finally,

\begin{align*}
(iii) \quad H(U, V) &\leq (1/n) I(U, V; Y) + \lambda_n \\
&\leq (1/n) I(X_1, X_2; Y) + \lambda_n. \quad (84)
\end{align*}

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

\begin{align*}
(H(U|V), H(V|U), H(U, V)) &\in \lim_{n \to \infty} C_n,
\end{align*}

which proves the converse.

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

**Theorem 3:**

The correlated sources \{ $U_1, U_2, \ldots, U_m$ \} can be communicated reliably over the MAC \( (X_1 \times X_2 \times \ldots \times X_m; Y, p(y|x_1, x_2, \ldots, x_m)) \) if and only if there exists some $k$ such that
\[ H(U(S)|U(S^C)) < \frac{1}{k}I(X(S);Y|X(S^C),U(S^C)) , \]

for all subsets \( S \subseteq \{1,2,\ldots,m\} \).

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

\[ C_n = \{ (R_1,R_2,R_3): R_1 < \frac{1}{\lfloor nR \rfloor} I(X_1^n;Y^n|U_1^nR_1^n,X_2^n) \]

\[ R_2 < \frac{1}{\lfloor nR \rfloor} I(X_2^n;Y^n|V_1^nR_1^n,X_1^n) \]  \hspace{1cm} \text{(86)}

\[ R_3 < \frac{1}{\lfloor nR \rfloor} I(X_1^n,X_2^n;Y^n) \]

for some \( \prod_{i=1}^{\lfloor nR \rfloor} p(u_i,v_i)p(x_1^n|u_1^{\lfloor nR \rfloor})p(x_2^n|y_1^{\lfloor nR \rfloor}) \prod_{i=1}^{\lfloor nR \rfloor} p(y_i|x_1i,x_2i) \).

\[ \text{Proof: } \] 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 \((U,V)\) sequences into blocks of superletters of length \( p_i \), and breaking the \( X \) sequence into blocks of superletters of length \( q_i \).
REFERENCES


APPENDIX A

In this appendix, we shall bound

\[ P \{ (u, v, w, s(w), x_1(u|s), x_2(v|s), y) \in A_\varepsilon | B \}, \]

under the various assumptions of independence on \( u, v, w, s, x_1, x_2, \) and \( y \) that arise in the proof of Theorem 1. Recall that \((u_0, v_0, w_0) \in A_\varepsilon \), where \( A_\varepsilon \) denotes the set of all jointly typical \((u, v, w)\) sequences, and \( B \) denotes the event that this particular \((u_0, v_0)\) is the output of the source. Our bound will hold uniformly for each \((u_0, v_0) \in A_\varepsilon \).

First we are going to prove the following lemma which is used repeatedly in our 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 \((z_1, z_2) \in A_\varepsilon\), and let \(z_3, z_4, z_5\) be drawn according to

\[ P(z_3 = z_3, z_4 = z_4, z_5 = z_5 | z_1, z_2) = \prod_{i=1}^{n} p(z_3_i | z_1_i, z_2_i) p(z_4_i | z_3_i, z_2_i) p(z_5_i | z_3_i, z_1_i). \]

In other words \(z_3\) depends only on \(z_1, z_2\); \(z_4\) depends only on \(z_3, z_2\); and \(z_5\) depends only on \(z_3, z_1\). Then we have

\[ -n[I(z_1; z_4 | z_2, z_3) + I(z_5; z_2, z_4 | z_1, z_3) - 8 \varepsilon] \]

\[ P \{ (z_1, z_2, z_3, z_4, z_5) \in A_\varepsilon \} \leq 2 \tag{A-2} \]

**Proof:** Since \((z_1, z_2) \in A_\varepsilon\), we have

\[ P \{ (z_1, z_2, z_3, z_4, z_5) \in A_\varepsilon \} = \sum_{(z_3, z_4, z_5)} P \{ (z_3, z_4, z_5) = (z_3, z_4, z_5) | z_1, z_2 \} \]

\[ P \{ (z_1, z_2, z_3, z_4, z_5) \in A_\varepsilon \} \]

\[ (z_1, z_2, z_3, z_4, z_5) \in A_\varepsilon \]
But from (A-1)

\[ P \{ (Z_3, Z_4, Z_5) = (Z_3, Z_4, Z_5) | Z_1, Z_2 \} = \]

\[ P \{ Z_3 = z_3 | Z_1, Z_2 \} \cdot P \{ Z_4 = z_4 | Z_3, Z_2 \} \cdot P \{ Z_5 = z_5 | Z_3, Z_1 \} \]  \hspace{1cm} (A-4)

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

\[ -n [ H(Z_3 | Z_1, Z_2) + 2 \varepsilon ] \]

\[ P \{ Z_3 = z_3 | Z_1, Z_2 \} \leq 2 \]  \hspace{1cm} (A-5)

\[ -n [ H(Z_4 | Z_3, Z_2) + 2 \varepsilon ] \]

\[ P \{ Z_4 = z_4 | Z_3, Z_2 \} \leq 2 \]  \hspace{1cm} (A-6)

\[ -n [ H(Z_5 | Z_3, Z_1) + 2 \varepsilon ] \]

\[ P \{ Z_5 = z_5 | Z_3, Z_1 \} \leq 2 \]  \hspace{1cm} (A-7)

Using (A-5)-(A-7) and the bound on the cardinality of the set

\[ \{ (z_3, z_4, z_5) : (z_1, z_2, z_3, z_4, z_5) \in A_\varepsilon \} \], we have

\[ n [ H(Z_3, Z_4, Z_5 | Z_1, Z_2) ] - n [ H(Z_3 | Z_1, Z_2) + 2 \varepsilon ] \]

\[ P \{ (z_1, z_2, z_3, z_4, z_5) \in A_\varepsilon \} \leq 2 \]

\[ -n [ H(Z_4 | Z_3, Z_2) + 2 \varepsilon ] \]

\[ -n [ H(Z_5 | Z_3, Z_1) + 2 \varepsilon ] \]  \hspace{1cm} (A-8)

Substituting

\[ H(Z_3, Z_4, Z_5 | Z_1, Z_2) = H(Z_3 | Z_1, Z_2) + H(Z_4 | Z_1, Z_2, Z_3) + H(Z_5 | Z_1, Z_2, Z_3, Z_4) \]  \hspace{1cm} (A-9)

into (A-8) we have

\[ n [ I(Z_4; Z_1 | Z_2, Z_3) + I(Z_5; Z_2, Z_4 | Z_1, Z_3) - 8 \varepsilon ] \]

\[ P \{ (z_1, z_2, z_3, z_4, z_5) \in A_\varepsilon \} \leq 2 \]  \hspace{1cm} (A-10)
This completes the proof.

Now we bound \{ (u, v, f(u), S(f(u)), X_1(u | S), X_2(v | S), Y) \in A \epsilon | B \} in different cases. Note that in all cases we are assuming \((u, v, w) \in A \epsilon\).

**A-1:** \( u = u_0, v = v_0 \) (Therefore \( w = w_0, S = S_0 \)):

Here \( u, v, w_0 \) are fixed and \( S_0, X_1(u | S_0), X_2(v | S_0), Y \) are random variables. We use Lemma 1 with \( Z_1 = (v_0, w_0), Z_2 = u, Z_3 = S_0, Z_4 = X_1(u | S_0), Z_5 = (X_2(v_0 | S_0), Y) \).

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 (A-10), we have

\[
I(Z_4; Z_1 | Z_3, Z_2) = I(X_1; V, W | U, S) = H(X_1 | U, S) - H(X_1 | U, V, W, S)
\]

\[
= H(X_1 | U, S) - H(X_1 | U, S) = 0 , \quad (A-11)
\]

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

Also we have

\[
I(Z_5; Z_2, Z_4 | Z_1, Z_3) = I(X_2; Y | U, X_1 | V, W, S)
\]

\[
\overset{1}{=} I(X_2; Y | U, X_1 | V, S)
\]

\[
= H(X_2, Y | V, S) - H(X_2, Y | U, V, X_1, S)
\]

\[
\overset{2}{=} H(X_2 | V, S) + H(Y | X_2, V, S) - H(X_2 | U, V, X_1, S) - H(Y | X_1, X_2)
\]

\[
\overset{3}{=} H(X_2 | V, S) + H(Y | X_2, V, S) - H(X_2 | V, S) - H(Y | X_1, X_2)
\]

\[
\overset{4}{=} H(Y | X_2, V, S) - H(Y | X_1, X_2, V, S)
\]

\[
= I(Y; X_1 | X_2, V, S) , \quad (A-12)
\]
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 (A-10), (A-11), and (A-12) it follows that

$$P\{ (u,v,w_0,S_0,X_1(u|S_0),X_2(v|S_0),Y) \in A_\varepsilon | B \} \leq 2$$

(A-13)

**A-2: $v \neq v_0$, $u = u_0$ (Therefore $w = w_0$, $S = S_0$):**

Again we assume $(u_0,v,w_0) \in A_\varepsilon$. This case is similar to case A-1, and we obtain

$$P\{ (v,u_0,w_0,S_0,X_1(u_0|S_0),X_2(v|S_0),Y) \in A_\varepsilon | B \} \leq 2$$

(A-14)

**A-3: $u \neq u_0$, $v \neq v_0$ but $w = w_0$ (Hence $S = S_0$):**

As usual we are assuming $(u,v,w_0) \in A_\varepsilon$. Here $u$, $v$, $w_0$ are fixed and $S_0$, $X_1(u|S_0)$, $X_2(v|S_0)$ and $Y$ are random variables. We apply the lemma with $z_1 = w_0$, $z_2 = (u,v)$, $z_3 = S_0$, $z_4 = (X_1(u|S_0), X_2(v|S_0))$, $z_5 = Y$.

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 (A-10). We have

-33-
\[ I(Z_4;Z_1|Z_2,Z_3) = I(X_1,X_2;W|U,V,S) = 0, \quad (A-15) \]

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

\[
I(Z_5;Z_2,Z_4|Z_1,Z_3) = I(Y;U,V,X_1,X_2|W,S)
\]

\[ = H(Y|W,S) - H(Y|X_1,X_2,W,S) \]

\[ = I(Y;X_1,X_2|W,S), \quad (A-16) \]

where \( 1 \) follows from the conditional independence of \( Y \) and \((U,V)\) given \((X_1,X_2)\). From \((A-10)\), \((A-15)\), \((A-16)\) it follows that

\[ -n[I(X_1,X_2;Y|W,S)-8 \varepsilon] \]

\[ P\{ (u,v,w_0,S_0,X_1(u|S_0),X_2(v|S_0),Y) \in A \in \mathcal{B} \} \leq 2 \quad (A-17) \]

**A-4: \( u \neq u_0, \ v \neq v_0, \ w \neq w_0, \ s_0 \neq s' \):**

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

\[ Z_1 = \phi, \ Z_2 = (u,v,w,s'), \ Z_3 = \phi, \]

\[ Z_4 = (X_1(u|s'),X_2(v|s'),Y), \ Z_5 = Y. \]

From the lemma, we have

\[ I(Z_4;Z_1|Z_2,Z_3) = I(X_1,X_2;\phi|U,V,W,S) = 0 \quad (A-18) \]

and

\[ I(Z_5;Z_2,Z_4|Z_1,Z_3) = I(Y;U,V,W,S,X_1,X_2) = I(Y;X_1,X_2). \quad (A-19) \]
Hence

\[ P\{ (u,v,w,s',X_1(u|s'),X_2(v|s'),Y) \in A_{s'} | S_0 = s', B \} \leq 2^{-n[I(X_1,X_2;Y)-8 \epsilon]} \]  
(A-20)

A-5: \( u \neq u_0, \ v \neq v_0, \ w \neq w_0, \ S_0 = s' \)

Here, as in A-4, \((u,v,w,s') \in A_{s'}\) are fixed and \(X_1, X_2\) and \(Y\) are random variables, and we wish to bound

\[ P\{ (u,v,w,s',X_1(u|s'),X_2(v|s'),Y) \in A_{s'} | S_0 = s', B \} . \]

In the lemma, set

\[ z_1 = s', \ z_2 = (u,v,w,s'), \ z_3 = \phi, \ z_4 = (X_1(u|s'),X_2(v|s')), \ z_5 = Y, \]

thus obtaining

\[ I(z_4;z_1|z_2,z_3) = I(X_1,X_2;S|U,V,W,S) = 0 \]  
(A-21)

and

\[ I(z_5;z_2,z_4|z_1,z_3) = I(Y;U,V,W,S,X_1,X_2|S) \]

\[ = H(Y|S) - H(Y|X_1,X_2,S) \]

\[ = I(Y;X_1,X_2|S), \]  
(A-22)

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

\[ P\{ (u,v,w,s',X_1(v|s'),X_2(v|s'),Y) \in A_{s'} | S_0 = s', B \} \leq 2^{-n[I(X_1,X_2;Y|S)-8 \epsilon]} \]  
(A-23)
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 $p_2(s)p_2(x_1|u,s)p_2(x_2|v,s)$ be two arbitrary conditional mass functions on $S \times X_1 \times X_2$ as defined in ( ).

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

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

$$\alpha I_1(X_2;Y|X_1,U,S) + (1-\alpha)I_2(X_2;Y|X_1,U,S) \leq I(X_2;Y|X_1,U,S'), \quad (B-2)$$

and

$$\alpha I_1(X_1,X_2;Y) + (1-\alpha)I_2(X_1,X_2;Y) \leq I(X_1,X_2;Y), \quad (B-3)$$

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

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

$$(B-1) = I(X_1;Y|X_2,V,S')$$

$$(B-2) = I(X_2;Y|X_1,U,S')$$

and

$$(B-3) = I(X_1,X_2;Y|T) \leq I(X_1,X_2;Y),$$

thus establishing convexity.