Non-Asymptotic Achievability Bounds in Multiuser Information Theory

Sergio Verdú
Princeton University, NJ 08544, USA

Abstract—Invoking random coding, but not typical sequences, we give non-asymptotic achievability results for the major setups in multiuser information theory. No limitations, such as memorylessness or discreteness, on sources/channels are imposed. All the bounds given are powerful enough to yield the constructive side of the (asymptotic) capacity regions in the memoryless case. The approach relies on simple non-asymptotic counterparts of the packing and covering lemmas conventionally used in conjunction with the typical sequence approach.

Index Terms—Shannon theory, achievability, finite blocklength regime, random coding, multiple-access channels, Wyner-Ziv compression, broadcast channels, data transmission with encoder side information, almost-lossless compression with a helper.

I. INTRODUCTION

Coding theorems express the fundamental limits of data transmission and/or compression in terms of the statistical laws of channels and sources. Since the inception of information theory [1] the constructive side of those theorems relies on two central notions: random coding, and typical sequences. A proof technique which is an instance of the probabilistic method, random coding shows that the average error probability attained by the set of all codes (under some auxiliary distribution on that set) is no worse than a certain threshold. This guarantees the existence of a code whose own error probability satisfies that property. In [1], Shannon used random coding to show that memoryless channel capacity is lower bounded by the maximal mutual information, while in [2] he used a random selection of the digital-to-analog converter (or, more generally, the codebook of reproduction points) in order to show that the rate-distortion function of a memoryless source is upper bounded by the minimal mutual information. When used to show existence of lossless codes, random coding is known as random binning, and is also powerful enough to prove the achievability part of Shannon’s lossless coding theorem. Introduced by Cover [3], random binning saw the light in the multiuser setting of separate compression of dependent sources due to Slepian and Wolf [4]. Notable exceptions to achievability proofs that rely on random coding are Shannon’s almost-lossless compression coding theorem [1], the transmission code construction in Feinstein’s doctoral dissertation [5], and Wolfowitz’s lossy compression construction [6] based on his earlier transmission converse bound [7] and Feinstein’s construction, and the explicit construction of optimum lossless variable-length codes without prefix constraints (e.g. [8], [9]). Those alternatives to random coding have proven to be difficult to generalize to multiuser information theory settings, where random coding is indeed the existence method of choice.

Also conceived by Shannon in [1], the typical sequence approach neglects the possibility that the source or channel may deviate appreciably from the average behavior predicted by their statistical laws. The law of large numbers (or the ergodic theorem in the case of systems with memory) guarantees that such an optimistic view is validated in the long run. As in the case of random coding, finite-alphabet embodiments of the typical sequence achievability approach reign supreme in multiuser information theory (e.g. Csiszár and Körner [10] and El Gamal and Kim [11]). In that context, key tools for the typical sequence approach are the so-called packing and covering lemmas.

In a series of recent works [9], [12]–[15] we have investigated the fundamental limits of various single-user data transmission and data compression problems as a function of the blocklength. With the single exception identified in [9], the achievability and converse bounds do not coincide in the non-asymptotic regime. However, they do provide an excellent approximation to the backoff from the Shannon limits for blocklengths as low as a few hundred.

This paper obtains non-asymptotic achievability random-coding bounds in multiuser information theory without invoking the method of typical sequences. We do not impose any assumptions on sources or channels such as memorylessness, ergodicity, etc. Instead, time plays no role in the analysis and all the bounds are given in the fundamental one-shot version involving a single source observation and a single channel use. When applying the results, those symbols can be viewed as vectors with \( n \) components where \( n \) is the blocklength. Taking limits and invoking the law of large numbers the bounds we give readily recover the known asymptotic achievability results for memoryless systems. Equally easily they can be applied to systems with memory that behave ergodically. Finite alphabet assumptions are superfluous except where noted. Beyond the worthy goal of assessing the backoff from the fundamental limits due to limited delay, our main motivations here are simplification, generalization and insight. Therefore, in this paper we pursue multiuser counterparts to the simplest achievability bounds, rather than the more sophisticated single-user bounds developed in [12], [14].

This work was supported in part by the National Science Foundation (NSF) under Grant CCF-1016625 and by the Center for Science of Information (CSoI), an NSF Science and Technology Center, under Grant CCF-0939370.
The key information quantities whose averages are the familiar information measures are listed in Section II. In Sections III and IV we give non-asymptotic versions of the packing and covering lemmas, which play a key role in our proofs. We deal with the problems of almost lossless compression with encoded side information at the decompressor (Section V); lossy compression with side information at the decompressor (Section VI); data transmission with side information at the encoder (Section VII); multiaccess (Section VIII); broadcast without public messages (Section IX).

II. PRELIMINARIES

Entropy and mutual information are the expectations of the random variables we refer to as information and information density, respectively. In contrast to the conventional packing and covering lemmas [11, Lemma 3.1 and 3.3] and [10, Lemma 10.1] as well as in contrast to the asymptotic fundamental limits, the non-asymptotic versions do not involve the expectations but the distribution functions of the information random variables.

Definition 1. If \( A \) is discrete, the conditional information of \( a \in A \) given \( b \in B \), according to \( P_{X|Y} \), is

\[
i_{X|Y}(a|b) = \log \frac{1}{P_{X|Y}(a|b)}
\]

It easily follows that

\[
|a \in A : i_{X|Y}(a|b) \leq \log \theta| \leq \theta
\]

Definition 2. The relative information of \( a \in A \) corresponding to probability measures \( P \) and \( Q \) on \( (A, \mathcal{F}) \) is

\[
i_{P||Q}(a) = \frac{dP}{dQ}(a)
\]

We often abbreviate \( i_{P||P_Y} = i_{X||Y} \). A simple property satisfied by the relative information is

Lemma 1 (\( \beta, P_X, P_Y, \gamma \)).

\[
\beta \mathbb{E} [g(Y)1\{i_{X||Y}(Y) \geq \log \beta\}] \\
\leq \mathbb{E} [g(Y)1\{i_{X||Y}(X) \geq \log \beta\}]
\]

Definition 3. Given a distribution \( P_Y \) on \( B \) and a random transformation \( P_{Y|X}: A \rightarrow B \) the information density of \((x, y) \in A \times B \) is

\[
i_{X,Y}(x; y) = i_{P_{Y|X} = 1||P_Y}(y)
\]

III. NON-ASYMPTOTIC PACKING LEMMA

Lemma 2 (\( P_{XY}, N, \eta \)).

\[
\mathbb{P} \left[ \max_{i=1, \ldots ,N} i_{X,Y}(X_i; Y) \geq \eta \right] \leq N \exp(-\eta)
\]

where for all \( i = 1, \ldots , N \) \( P_{X,Y} = P_X \times P_Y \)

Proof: Note that in the special case of \( N = 1 \), (6) follows from Lemma 1 (\( \exp(-\eta), P_{XY}, P_X \times P_Y \), 1):

\[
\mathbb{P} [i_{X,Y}(X_1; Y) \geq \eta] \leq \exp(-\eta)
\]
• (14) & (15) $\iff (V_1, \ldots, V_m)$ are independent identically distributed (also conditioned on $Z$);

• (16) $\iff$ Lemma 1 $\left( \frac{\exp(\gamma)}{M} \right) P_V|Z = z, P_V, 1 \{ (z, \cdot) \in S \} ;$

• (17) $\iff (1 - \frac{\epsilon}{M})^M < 1 - p + e^{-\alpha}$

Averaging (15)-(17) with respect to $P_Z$, we obtain

$$\mathbb{P} \left[ \bigcap_{m=1}^M \{ (Z, V_m) \notin \mathcal{S} \} \right] \leq \mathbb{P} \left[ (Z, V) \notin \mathcal{S} \right] + e^{-\exp(\gamma)} \quad (18)$$

which is the desired result.

Using the same notation as in Lemma 4, we have the following useful generalization:

**Lemma 5** $(P_{ZV}, M, \gamma, \mathcal{F})$. For any event $\mathcal{F}$,

$$\mathbb{P} \left[ \bigcap_{m=1}^M \{ (Z, V_m) \notin \mathcal{F} \} \right] \leq \mathbb{P} \left[(Z, V) \notin \mathcal{F} \right] + e^{-\exp(\gamma)} + \mathbb{P} \left[ |Z, V| > \log M - \gamma \right] \quad (19)$$

**Proof:** It is easy to check that the bound in (18) still holds if the event $\mathcal{S}$ is a subset of the right side of (13). Taking

$$\mathcal{S} = \mathcal{F} \cap \{ (z, v) \in Z \times V, i_{Z,V}(z; v) \leq \log M - \gamma \} \quad (20)$$

we get

$$\mathbb{P} \left[ \bigcap_{m=1}^M \{ (Z, V_m) \notin \mathcal{F} \} \right] \leq \mathbb{P} \left[ \bigcap_{m=1}^M \{ (Z, V_m) \notin \mathcal{S} \} \right] \leq \mathbb{P} \left[(Z, V) \notin \mathcal{S} \right] + e^{-\exp(\gamma)} \quad (22)$$

where (22) is (18). Finally, (19) follows applying the union bound to the complement of (20).

In a sense, the covering lemma is dual to the packing lemma in that it gives a lower bound on the probability that there is an “impostor” pair (now defined as one that satisfies a certain property $\mathcal{F}$). If we use the Covering Lemma 5 when $M$ is large enough that $\mathbb{P} \left[ |Z, V| > \log M - \gamma \right]$ is negligible and $\gamma$ is large but negligible with respect to $\log M$, then the probability that at least one of the $M$ independently generated pairs $(Z, V_i) \in \mathcal{F}$ ($V_i$ independent of $Z$) is essentially lower bounded by the probability that $(Z, V) \in \mathcal{F}$ with $(Z, V) \sim P_{ZV}$.

V. ALMOST-LOSSLESS COMPRESSION WITH ENCODED SIDE-INFORMATION

This section considers the setting introduced by Ahlswede-Körner [17] and Wyner [18]. We allow the alphabet $X$ to be finite or countably infinite.

**Definition 4.** A fixed-length source code with compressed side information at the decoder consists of the mappings:

- **Encoder:** $f_1 : \mathcal{X} \rightarrow \{1, \ldots, M_1\}$
- **SL Encoder:** $f_2 : \mathcal{Y} \rightarrow \{1, \ldots, M_2\}$
- **Decoder:** $g : \{1, \ldots, M_1\} \times \{1, \ldots, M_2\} \rightarrow \mathcal{X} \cup \emptyset$

If $X$ and $Y$ are governed by the joint distribution $P_{X,Y}$, the minimal error probability achievable by the class of size-$(M_1, M_2)$-codes defined in Definition 4 is denoted by

$$e_{X\mid Y}(M_1, M_2) = \min_{f_1, f_2, g} \mathbb{P}[g(f_1(X), f_2(Y)) \neq X] \quad (23)$$

Consider the following achievability result.

**Theorem 1.**

$$e_{X\mid Y}(M_1, M_2) \leq \min_{\gamma > 0} \left\{ \mathbb{P}[t_{X\mid U}(X|U) > \log M_1 - \gamma] + \mathbb{P}[t_{Y\mid U}(Y|U) > \log M_2 - \gamma] \right\} + e^{-\exp(\gamma)} + \exp(-\gamma) \quad (24)$$

where $X - \square - Y - \square - U$, which denotes that $X$ and $U$ are conditionally independent given $Y$.

**Proof:** We fix $\gamma > 0$ and $P_{U|Y}$.

*Side information encoder and surrogate codebook:* The side information available at the decompressor takes one of $M_2$ values (which we call the surrogate codebook) drawn from the alphabet $U$. The surrogate codebook (available at both the side information compressor and the decompressor) is:

$$u = (u_1, \ldots, u_{M_2}) \in U^{M_2} \quad (25)$$

Given the surrogate codebook and once it observes $y \in \mathcal{Y}$, the side-information encoder outputs $f_2(y) = \ell \in \{1, \ldots, M_2\}$ such that

$$\pi(u, y) = \min_{j=1,\ldots,M_2} \pi(u_j, y) \quad (26)$$

with arbitrary tie-breaking, where

$$\pi(u, y) = \mathbb{P}[t_{X\mid U}(X|U) > \log M_1 - \gamma| Y = y] \quad (27)$$

Notice that

$$\mathbb{P}[t_{X\mid U}(X|U) > \log M_1 - \gamma] = \mathbb{E} [\pi(U, Y)] \quad (28)$$

in view of $X - \square - Y - \square - U$. For brevity, we have not indicated the dependence of $f_2(y_0)$ on the surrogate codebook $u$.

*Decompressor:* Upon receipt of $(f_1(x), f_2(y)) = (i, \ell)$, the decompressor examines the compression bin

$$G(i, \ell) = \{ x \in f_1^{-1}(i) : \ t_{X\mid U}(x|u_\ell) \leq \log M_1 - \gamma \} \quad (29)$$

and outputs either its unique element or the error warning $g(i, \ell) = \emptyset$ if $|G(i, \ell)| = 1$.

*Error probability analysis:* If $(x_0, y_0) \in X \times Y$ is the realization of $(X, Y)$, then $g(f_1(x_0), f_2(y_0)) \neq x_0$ occurs if and only if either of the following conditions occur:

- $E_1 : t_{X\mid U}(x_0|u_\ell(y_0)) \geq \log M_1 - \gamma$
- $E_2 : \exists x \neq x_0 : f_1(x) = f_1(x_0), \ t_{X\mid U}(x|u_\ell(y_0)) \leq \log M_1 - \gamma$

We average the probability of $E_1 \cup E_2$ with respect to:

- $(x_0, y_0) \sim P_{X,Y}$,
- random binning: $\{f_1(x), x \in X\}$ independent, equiprobable on $\{1, \ldots, M_1\}$,
- random surrogate codebook: $u \sim P_U \times \cdots \times P_U$, with all choices independent of each other.
Lemma 5 (where we have used the fact that $\mathbb{P}[E_2] \leq \exp(-\gamma)$ since
\[ \sum_{x \neq x_0} \mathbb{P}[F_1(x) = F_1(x_0)] 1\{f(x | U(x) | U) \leq \log M - \gamma\} \leq \exp(-\gamma) \] (30)
for all $(x, u) \in X \times U$, where (30) follows from (2).
To bound $\mathbb{P}[E_1]$, denote by $U^* = \tilde{f}_2(Y) \in U$, a random variable that depends on both the side information and the surrogate codebook. Then,
\[
\mathbb{P}[E_1] = \mathbb{P}[\pi(U^*, Y) > \log M - \gamma] \\
= \mathbb{E}[\mathbb{P}[\pi(U^*, Y) > \log M - \gamma] \\
= \mathbb{E}[\pi(U^*, Y)] \\
= \int_0^1 \mathbb{P}[\pi(U^*, Y) > t] \, dt \] (34)
where we have used the fact that $0 \leq \pi(u, y) \leq 1$.
To bound upper bound $\mathbb{P}[\pi(U^*, Y) > t]$ we invoke Lemma 5 ($P_{Y | U}, M_2, \gamma, F$) with
\[ F = \{(u, y) \in Y \times U : \pi(u, y) \leq t\} \] (35)
in order to obtain
\[
\mathbb{P}[\pi(U^*, Y) > t] \leq \mathbb{P}[\pi(U, Y) > t] + e^{-\exp(\gamma)} \\
+ \mathbb{P}[\pi(Y, U) > \log M_2 - \gamma] \] (36)
Inserting bound (36) into (34) we conclude that
\[
\mathbb{P}[E_1] \leq \mathbb{P}[\pi(U^*, Y) > \log M_1 - \gamma] + e^{-\exp(\gamma)} \\
+ \mathbb{P}[\pi(Y, U) > \log M_2 - \gamma] \] (37)
where we have used (28).
Adding the averaged upper bounds (30) and (37), there must exist a surrogate codebook and a binning function $f_1$ that satisfy the claimed bound.

VI. LOSSY COMPRESSION WITH SIDE INFORMATION AT THE DECOMPRESSOR

This section considers the setting introduced by Wyner and Ziv [19], [20].

**Definition 5.** A fixed-length lossy compression code with side information at the decompressor is a pair of mappings:

**Encoder:** $f : X \to \{1, \ldots, M\}$

**Decoder:** $c : \{1, \ldots, M\} \times Y \to \hat{X} \cup \hat{e}$

Given a joint distribution $P_{X,Y}$ and a distortion measure $d : X \times \hat{X} \to [0, \infty]$, the minimal probability of excess distortion with side information is denoted by
\[
\epsilon^*_X(Y)(M, d) = \inf \mathbb{P}[d(X, c(f(X), Y))] > d] \] (38)
with the infimum over all codes of size $M$ in Definition 5.

**Theorem 2.** For any set $U$, random transformations $P_{U | X} : X \to U$, $P_{X | U, Y} : U \times Y \to \hat{X}$, positive integer $L$ and $\gamma > 0$,
\[
\epsilon^*_X(Y)(M, d) \leq \mathbb{P}\left[ d\left(X, \hat{X}(U, Y)\right) > d \right] \\
+ \mathbb{P}[f(X, U) > \log ML - \gamma] \\
+ \mathbb{P}[U_{X,Y}(U, Y) \leq \log L + \gamma] \\
+ e^{-\exp(\gamma)} + \exp(-\gamma) \] (39)
where $Y \to X \to U$.

**Proof:** Fix $M, d > 0, U, P_{U | X}, P_{X | U, Y}, L$ and $\gamma > 0$.

**Compressor:** The codebook is the two dimensional array
\[
\mathbf{u} = \begin{bmatrix}
    u_{11} & \cdots & u_{1L} \\
    \vdots & \ddots & \vdots \\
    u_{M1} & \cdots & u_{ML}
\end{bmatrix} \in U^{ML} \] (40)
To compress $x \in X$, we select the coefficient in the matrix $(i(x, u), j(x, u)) \in \{1, \ldots, M\} \times \{1, \ldots, L\}$ that achieves the minimum (arbitrary tie-breaking)
\[
\pi(x, u_{i,j}) = \min_{i,j} \pi(x, u_{i,j}) \] (41)
where $\pi : X \times U \to [0, 1]$ is defined by $\pi(x, u) = \mathbb{P}\left[ \{d(X, \hat{X}(u, Y)) > d\} \cap \{U_{X,Y}(u, Y) \leq \log L + \gamma\} | X = x \right]$ with $Y \sim P_{Y | X = x}$. The compressor outputs
\[
f(x) = i(x, u). \] (42)

**Decompressor:** Upon receipt of $f(x) = i \in \{1, \ldots, M\}$ and $y \in Y$, the decompressor looks for the unique index $j(i, y) = j \in \{1, \ldots, L\}$ such that
\[
u_{U,Y}(u_{i,j}; y) > \log L + \gamma \] (43)
and outputs
\[
c(i, y) = \hat{X}(u_{i,j}, y). \] (44)
If there is no, or more than one, such index, then the decompressor outputs $\hat{e}$.

**Distortion analysis:** Suppose that the realization of $(X, Y)$ is $(x_0, y_0)$, and denote for brevity
\[
i_0 = i(x_0, u) \] (45)
\[j_0 = j(x_0, u) \] (46)
If the decompressor outputs $\hat{e}$ or the resulting distortion exceeds $d$, then at least one of the following events occurs (the reverse need not be true):
\[
\mathcal{E}_1 : \exists j \neq j_0, v_{U,Y}(u_{i_0,j}, y_0) > \log L + \gamma \] (47)
\[
\mathcal{E}_2 : v_{U,Y}(u_{i_0,j_0}, y_0) \leq \log L + \gamma \] (48)
\[
\mathcal{E}_3 : d(x_0, \hat{X}(u_{i_0,j_0}, y_0)) > d \] (49)
We average $\mathbb{P}[\mathcal{E}_1 \cup \mathcal{E}_2 \cup \mathcal{E}_3]$ with respect to:
\[
(x_0, y_0, u) \sim P_{X,Y} \times P_U \] (50)
where the coefficients of the matrix $X$ are independent with identical distribution $P_U$ induced as $P_X = P_{U|X} = P_U$. It is convenient to denote the random indices chosen by the compressor as a function of $X$ and the codebook by

$$I^* = i(X, U) \quad J^* = j(X, U) \quad (U^* = u_{I^*, J^*})$$

$P[E_1]:$ For every $(m, \ell) \in \{1, \ldots, M\} \times \{1, \ldots, L\}$,

$$P[U \neq \{u_{I^*, J^*}Y > \log L + \gamma\}|(I^*, J^*) = (m, \ell)] = \sum_{j=1}^{L} P[U; Y(U_{ij}; Y) > \log L + \gamma] \times (I^*, J^*) = (m, \ell)]$$

$$\leq \sum_{j=1}^{L} P[U; Y(U_{ij}; Y) > \log L + \gamma] \times (I^*, J^*) = (m, \ell)] (49)$$

where (48) holds even if $m \neq 1$ because for any set of $L - 1$ index pairs $K \neq (m, \ell)$, the joint distribution of $Y$ and $\{U_{ij}, (i, j) \in K\}$ conditioned on $(I^*, J^*) = (m, \ell)$, is invariant to $K$. Unconditioning (48)-(49), we obtain

$$P[E_1] \leq \sum_{j=1}^{L} P[U; Y(U_{ij}; Y) > \log L + \gamma]$$

$$= L \sum_{j=1}^{L} P[U; Y(U_{ij}; Y) > \log L + \gamma]$$

$$\leq \exp(-\gamma)$$

where (51) reflects the fact that the codebook entries are identically distributed and independent of $Y$: $P_{U|Y} = P_{U|Y}$, and (52) follows from Lemma 2 ($P_{U|Y}, 1, \log L + \gamma$).

$$P[E_2 \cup E_3] = \int_0^1 P[\pi(X, U^*) > t] \ dt$$

To upper bound the integrand in (53), we invoke Lemma 5 ($P_{X|U}, M, \gamma, F$) with

$$F = \{(x, u) \in \mathcal{X} \times \mathcal{U}: \pi(x, u) \leq t\}$$

(54)

obtain

$$P[\pi(X, U^*) > t] = P[\bigcap_{j=1}^{M} \bigcap_{j=1}^{L} \{(X, U_{ij}) \notin \mathcal{F}\}]$$

$$\leq P[\pi(X, U) > t] + e^{-\exp(\gamma)}$$

(55)

Since (53) holds also with $U$ in lieu of $U^*$, we conclude that

$$P[E_2 \cup E_3] \leq P[\{(d(X, \tilde{X}(U, Y)) > d) \cup \{U_{ij}(U, Y) \leq \log L + \gamma\}]$$

$$+ P[I_{X; U} > \log ML - \gamma] + e^{-\exp(\gamma)}$$

(57)

Together with the union bound, the bounds in (52) and (57) yield the right side in (39). Therefore, a realization of the codebook (40) must exist that also satisfies that bound.

VII. DATA TRANSMISSION WITH SIDE INFORMATION AT THE ENCODER

This section considers the setup introduced by Gelfand and Pinsker [21].

Definition 6. A size-M data transmission code with side information at the encoder is a pair of mappings:

Encoder: $c: \{1, \ldots, M\} \times \mathcal{S} \rightarrow \mathcal{X}$

Decoder: $g: \mathcal{Y} \rightarrow \{1, \ldots, M\} \cup \emptyset$

The minimal average probability for a given size is

$$e^e(M) = \inf \frac{1}{M} \sum_{m=1}^{M} E[P[g(Y) \neq m|X = c(m, S)]$$

(58)

with the infimum over all codes of size $M$ in Definition 6.

Theorem 3. Fix $P_{Y|XS}$ and $P_S$. For any positive integer $L$, $\gamma > 0$, $\mathcal{X}: \mathcal{U} \times \mathcal{S} \rightarrow \mathcal{X}$, and $P_{U|S}$, where $U$ takes values on an arbitrarily chosen alphabet $\mathcal{U}$, the minimal average error probability achievable by a size-$M$ data transmission code with side information at the encoder is bounded by

$$e^e(M) \leq P[U; S(U, S) > \log L - \gamma] + \exp(-\gamma) + e^{-\exp(\gamma)}$$

$$+ P[U; Y(U, Y) \leq \log ML + \gamma]$$

(59)

where $P_{USXY} = P_{Y|XS}P_{X|US}P_{U|S}PS$ and $P_{X|US}$ is the deterministic transformation $X(\cdot, \cdot)$

Proof:

Encoder: The codebook, available at both encoder and decoder, is the two-dimensional array

$$u = \begin{bmatrix} u_{11} & \cdots & u_{1L} \\ \vdots & \ddots & \vdots \\ u_{ML} & \cdots & u_{ML} \end{bmatrix} \in \mathcal{U}^{ML}$$

(60)

To transmit message $m \in \{1, \ldots, M\}$, when the side information is $s \in \mathcal{S}$, the output of the encoder is:

$$c(m, s) = \tilde{X}(u_{mt}, s)$$

(61)

where

$$\zeta(s, u_{mt}) = \min_{j=1,\ldots,L} \zeta(s, u_{mj})$$

(62)

with arbitrary tie-breaking, and

$$\zeta(s, u) = P[U; Y(U, Y) \leq \log ML + \gamma|U, S = (u, s)]$$

(63)

with $Y \sim P_{Y|XS}(\cdot|\tilde{X}(u, s), s)$. Note that

$$E[\zeta(S, U)] = P[U; Y(U, Y) \leq \log ML + \gamma]$$

(64)

Decoder: Upon receipt of the output $y \in \mathcal{Y}$, the decoder identifies all the index pairs $i, j \in \{1, \ldots, M\} \times \{1, \ldots, L\}$ such that

$$u_{ij}; y > \log ML + \gamma;$$

(65)

if the rows of all those pairs are identical, then the decoder outputs the common row index

$$g(y) = i.$$
If the row indices are not equal, or if no index pair satisfies (65), then \( g(y) = e \).

**Error probability analysis:** Suppose that the message is \( m \in \{1, \ldots, M\} \), the realization of the side information is \( s_0 \in S \), the index chosen by the encoder is \( i_0 \in \{1, \ldots, L\} \), and the decoder receives \( y_0 \in Y \). An error (detected or undetected) occurs only if one or both of the following events occur:

\[
E_1: \quad u_{i_0;Y}(u_{m,i_0}; y_0) \leq \log ML + \gamma
\]
\[
E_2: \quad \exists (i, j), i \neq m \quad u_{i;Y}(u_{i,j}; y_0) > \log ML + \gamma
\]

We proceed to average the probabilities of those events with respect to the message, side information, output, and codebook, whose entries are independently drawn with distribution \( P_U \). Once the desired bound is established on the error probability averaged over the codebook choice, a codebook that satisfies the bound is guaranteed to exist.

Denoting the chosen entry of the codebook matrix (60) by \( U^* \) and the corresponding output (response of the random transformation \( P_X|\{S=U\} \)) by \( Y^* \), we have

\[
P[E_1] = P[u_{i;Y}(U^*; Y^*) \leq \log ML + \gamma]
\]
\[
= E[\zeta(S, U^*)]
\]
\[
= \int_0^1 P[\zeta(S, U^*) > t] \, dt.
\]

Letting

\[ F = \{(s, u) \in X \times U: \zeta(s, u) \leq t\}, \]

the probability in the right side of (69), conditioned on the message being \( m \) is

\[
P[\zeta(S, U^*) > t|m] = P[\bigcup_{j=1}^{L} \{(S, U_{m,j}) \notin F\}]
\]
\[
\leq P[\zeta(S, U) > t] + P[u_{i;S}(U; S) > \log L - \gamma] + e^{-\exp(\gamma)}
\]

where (72) follows from Lemma 5 (\( P_{SU}, L, \gamma, F \)).

Substituting (72) in (69), and recalling (64) we obtain:

\[
P[E_1] \leq P[u_{i;Y}(U; Y) \leq \log ML + \gamma] + e^{-\exp(\gamma)}
\]
\[
+ P[u_{i;S}(U; S) > \log L - \gamma]
\]

Lemma 3 (\( P_{UY}, (M - 1)L, \log ML + \gamma \)) yields

\[
P[E_2] \leq \frac{(M - 1)L}{ML} \exp(-\gamma)
\]

where the union in (74) is over \( (i, j) \in \{1, \ldots, m - 1, m + 1, \ldots, M\} \times \{1, \ldots, L\} \); therefore, the random variables \( \{U_{ij}\} \) are indeed independent of \( Y^* \) and are distributed according to \( P_U \).

The desired bound (59) is the sum of (72) and (76).

**VIII. Multicase**

**Definition 7.** A size-(\( M_1, M_2 \)) data transmission code for two-user multiple-access consists of the mappings:

- **Encoder 1:** \( c_1: \{1, \ldots, M_1\} \rightarrow X_1 \)
- **Encoder 2:** \( c_2: \{1, \ldots, M_2\} \rightarrow X_2 \)
- **Decoder:** \( g: Y \rightarrow \{1, \ldots, M_1\} \times \{1, \ldots, M_2\} \cup \{\varepsilon\} \)

For a given \( P_{Y|X_1 X_2} \) the minimal average (over independent equiprobable messages) error probability achieved by a size-(\( M_1, M_2 \)) data transmission code for two-user multiple-access is denoted by \( e^*_{Y|X_1 X_2}(M_1, M_2) \).

**Theorem 4.** For all \( \gamma > 0 \) and \( P_{X_1|U}, P_{X_2|U}, P_U \), where \( U \in U \), an arbitrary auxiliary set, the minimal average error probability, \( e^*_{Y|X_1 X_2}(M_1, M_2) \), is upper bounded by

\[
\begin{align*}
&\leq P[X_1,Y|X_2,U(X_1; Y|X_2,U) \leq \log M_1 + \gamma] \\
&+ P[X_2,Y|X_1,U(X_2; Y|X_1,U) \leq \log M_2 + \gamma] \\
&+ P[X_1,X_2,Y|U(X_1, X_2; Y|U) \leq \log M_1 M_2 + \gamma] \\
&+ 3 \exp(-\gamma)
\end{align*}
\]

with \( X_1 \rightarrow U \rightarrow X_2 \) and \( U \rightarrow X_1 \rightarrow X_2 \rightarrow Y \).

**Proof:** Fix \( \gamma > 0 \), \( P_{X_1|U}, P_{X_2|U} \) and \( P_U \). The decoder and the encoders agree beforehand on a common auxiliary value \( u \in U \).

**Encoders:** Known at the decoder and the corresponding encoders, the codebooks are the arrays:

\[
c_1(u) = [c_1(u) \cdots c_{M_1}(u)] \in X_1^{M_1}
\]
\[
c_2(u) = [d_1(u) \cdots d_{M_2}(u)] \in X_2^{M_2}
\]

The encoders are simply:

\[
c_1(i) = c_i(u) \quad i = 1, \ldots, M_1
\]
\[
c_2(i) = d_i(u) \quad i = 1, \ldots, M_2
\]

**Decoder:** Upon receipt of the output \( y \in Y \), the decoder identifies the index pairs \( (i, j) \in \{1, \ldots, M_1\} \times \{1, \ldots, M_2\} \) such that

\[
i_{X_1;Y|X_2,U}(c_1(u); y|d_j(u), u) > \log M_1 + \gamma
\]
\[
i_{X_2;Y|X_1,U}(d_j(u); y|c_i(u), u) > \log M_2 + \gamma
\]
\[
i_{X_1,X_2;Y|U}(c_i(u), d_j(u); y|u) > \log M_1 M_2 + \gamma
\]

If there is one and only one such pair, then the decoder outputs \( g(y) = (i, j) \). If there are none or more than one, the decoder outputs \( g(y) = e \).

**Error Probability Analysis:** If the observed output is \( y_0 \), and the transmitted messages are \( (m_1, m_2) \in \{1, \ldots, M_1\} \times \{1, \ldots, M_2\} \), an error occurs only if any of the following events occur

\[
E_1: i_{X_1,Y|X_2,U}(c_{m_1}(u); y_0|d_{m_2}(u), u) \leq \log M_1 + \gamma
\]
\[
E_2: i_{X_2,Y|X_1,U}(d_{m_2}(u); y_0|c_{m_1}(u), u) \leq \log M_2 + \gamma
\]
\[
E_3: i_{X_1,X_2,Y|U}(c_{m_1}(u), d_{m_2}(u); y_0|u) \leq \log M_1 M_2 + \gamma
\]
\[
E_4: \exists \neq m_1 i_{X_1,Y|X_2,U}(c_{m_1}(u); y_0|d_{m_2}(u), u) > \log M_1 + \gamma
\]
\[
E_5: \exists \neq m_2 i_{X_2,Y|X_1,U}(d_{m_2}(u); y_0|c_{m_1}(u), u) > \log M_2 + \gamma
\]
\( \mathcal{E}_6 : \exists i \neq m_1, \exists j \neq m_2, i_{X_1, X_2, Y_1}^u (c_1(u), d_1(u), \gamma Y_0 | u) > \log M_1 M_2 + \gamma \)

Note that \( \mathcal{E}_4 \) is necessary but not sufficient for the list of pairs that satisfy (82)-(84) to include a pair \((i, m_2)\), \(i \neq m_1\). This as well as similar considerations for \( \mathcal{E}_5 \) and \( \mathcal{E}_6 \) only overestimates the error probability.

We average the probabilities over the choice of \( u, c_1(u), \ldots, c_{M_1}(u), d_1(u), \ldots, d_{M_2}(u) \) drawn according to

\[
P_U P_{X_1 | U} \times \cdots \times P_{X_1 | U} \times P_{X_3 | U} \times \cdots \times P_{X_2 | U}
\]

Consequently, the three probabilities appearing in (77) correspond to \( P_1 \mathcal{E}_1, P_2 \mathcal{E}_2, P_3 \mathcal{E}_3 \), respectively.

To complete the proof we proceed to show that \( P_1 \mathcal{E}_i \), \( i = 4, 5, 6 \) are upper bounded by \( \exp(-\gamma) \). Let us introduce the random variables \( X_1, X_2, Y_1, Y_2 \) which, conditioned on \( U \), are independent of each other and of \( (X_1, X_2, Y_1) \) and are distributed according to \( P_{X_1 | U} \) and \( P_{X_2 | U} \), respectively.

Invoking the union bound, the averaged probability of \( \mathcal{E}_4 \), conditioned on \((m_1, c_1(u), u, \ldots, c_{M_1}(u), d_1(u), \ldots, d_{M_2}(u))\), is upper bounded by

\[
\begin{align*}
(M_1 - 1) \frac{1}{M_1} \mathbb{P} \left[ \{ X_1, X_2, Y_1, X_2 Y_1 | d_{m_2}(u) > \log M_1 + \gamma \} \right] & \\
& \leq \frac{M_1 - 1}{M_1} \exp(-\gamma)
\end{align*}
\]

(85)

where (85) follows from Lemma 2 \((P_{X_1, X_2, Y_1 | U}(d_{m_2}(u), u), M_1 - 1, \log M_1 + \gamma)\). Swapping the user indices, we obtain an analogous bound for \( \mathcal{E}_5 \). Finally, conditioned on \((m_1, m_2, u)\), the averaged probability of \( \mathcal{E}_6 \) is upper bounded by

\[
\begin{align*}
(M_1 - 1)(M_2 - 1) \frac{1}{M_1 M_2} \mathbb{P} \left[ \{ X_1, \tilde{X}_2, Y_1, \tilde{X}_2 Y_2 | u > \log M_1 M_2 + \gamma \} \right] & \\
& \leq \frac{M_1 - 1 M_2 - 1}{M_1 M_2} \exp(-\gamma)
\end{align*}
\]

(86)

where (86) follows from Lemma 2 \((P_{X_1, X_2, Y_1 | U} = u, (M_1 - 1)(M_2 - 1), \log M_1 M_2 + \gamma)\).

The conclusion is that there must exist an auxiliary value \( u \in U \), and codebooks \( (c_1(u), c_2(u)) \in \alpha_{M_1} \times \alpha_{M_2} \) with which the error probability is bounded by the right side of (77).

IX. BROADCAST

We consider the special case in which there is no public message to be transmitted to both receivers.

**Definition 8.** A size-\((M_1, M_2)\) code for two-user broadcast consists of the mappings:

Encoder:

\[
c : \{1, \ldots, M_1\} \times \{1, \ldots, M_2\} \rightarrow X
\]

Decoder 1:

\[
g_1 : Y_1 \rightarrow \{1, \ldots, M_1\} \cup \{e\}
\]

Decoder 2:

\[
g_2 : Y_2 \rightarrow \{1, \ldots, M_2\} \cup \{e\}
\]

The following result gives a non-asymptotic counterpart to Marton’s inner bound to the capacity region [23].

**Theorem 5.** For all positive integers \( M_1, M_2, L_2, \gamma > 0 \), \( f : U^2 \rightarrow X \) and joint distribution \( P_{U_1 U_2} \) on \( U^2 \) there exists a \((M_1, M_2)\) broadcast code such that either decoder makes an error with probability not exceeding

\[
\begin{align*}
P_{U_1, Y_1}(U_1; Y_1) & \leq \log M_1 + \gamma \\
+ P_{U_2, Y_2}(U_2; Y_2) & \leq \log M_2 L_2 + \gamma \\
+ P_{U_2, U_1}(U_2; U_1) & > \log L_2 - \gamma + 2 \exp(-\gamma) + e^{-c \exp(\gamma)}
\end{align*}
\]

where \( P_{Y_1, Y_2 U_1 U_2} = P_{Y_1, Y_2 X = f(U_1, U_2) P_{U_1 U_2}} \).

**Proof:** Fix \( M_1, M_2, L_2, \gamma > 0 \), \( f \), and \( P_{U_1 U_2} \). Encoder:

The encoder uses two codebooks:

\[
u = \begin{bmatrix} u_1 & \cdots & u_{M_1} \end{bmatrix} \in U^{M_1}
\]

\[
v = \begin{bmatrix} v_{11} & \cdots & v_{L_2} \\
\vdots & \ddots & \vdots \\
v_{M_1 1} & \cdots & v_{M_1 L_2} \end{bmatrix} \in U^{L_2 M_2}
\]

To send messages \((m_1, m_2) \in \{1, \ldots, M_1\} \times \{1, \ldots, M_2\} \), the encoder transmits

\[
c(m_1, m_2) = f(u_{m_1}, v_{m_2, l})
\]

where \( l \in \{1, \ldots, L_2\} \) achieves

\[
\min_j \zeta(u_{m_1}, v_{m_2, j})
\]

with

\[
\zeta(u_1, u_2) = P_{Y_1 Y_2 U_1 U_2}((u_1, u_2) | (u_1, u_2))
\]

and

\[
Y_{u_1 u_2} = \{(y_1, y_2) \in Y_1 \times Y_2 : y_{U_{1, Y_1}}(u_1; y_1) \leq \log M_1 + \gamma \text{ or } y_{U_{2, Y_2}}(u_2; y_2) \leq \log M_2 L_2 \}
\]

Note that

\[
\mathbb{E}[\zeta(U_1, U_2)] \leq P_{U_1, Y_1}(U_1; Y_1) \leq \log M_1 + \gamma \\
+ P_{U_2, Y_2}(U_2; Y_2) \leq \log M_2 L_2 + \gamma
\]

**Decoder 1:** Upon receipt of \( y_1 \in Y_1 \), Decoder 1 outputs the unique index \( m_1 \in \{1, \ldots, M_1\} \) that satisfies

\[
u_{1, Y_1}(u_{m_1}; y_1) > \log M_1 + \gamma
\]

If there is no such index or a multiplicity of them, then Decoder 1 outputs \( g_1(y_1) = e \).

**Decoder 2:** Upon receipt of \( y_2 \in Y_2 \), Decoder 2 lists all the pairs \((m, l) \in \{1, \ldots, M_2\} \times \{1, \ldots, L_2\}\) that satisfy

\[
u_{2, Y_2}(u_{m, l}; y_2) > \log L_2 M_2 + \gamma
\]

If all such pairs share the first coordinate, then Decoder 2 outputs it: \( g_2(y_2) = m \). If the list is empty or there are discrepancies in the first coordinate, then it signals error \( g_2(y_2) = e \).

**Error probability analysis:** Let the transmitted messages be \((1, 1)\), the index that achieves (91) be \( f \), and the observed outputs be \((y_1, y_2)\). An error occurs at either decoder only if any number of the following events take place:
Next, we average the probabilities of those events with respect to the arrays \( u \) and \( v \), whose entries are independently drawn with distributions \( P_{U_1} \) and \( P_{U_2} \), respectively. Under such random choices, we denote by \( U_1^* \) and \( U_2^* \) the coefficients selected by the encoder from \( u \) and \( v \), respectively, and the corresponding outputs by \( Y_1^* \) and \( Y_2^* \).

Since \( \{U_{1,m} = 2, \ldots, M\} \) are independent of any other random variable and have distribution \( P_{U_1} \), Lemma 3 \((P_{U_1} Y_1, M_1 - 1, \log M_1 + \gamma)\) yields

\[
\mathbb{P}[\mathcal{E}_1] \leq \mathbb{E} \left[ \mathbb{P} \left[ \cup_{m \neq 1} \{u_i, y_i (U_{1,m} ; Y_1^*) > \log M_1 + \gamma \right\} Y_1^* \right] \\
\leq \frac{M_1 - 1}{M_1} \exp(-\gamma) \\
\leq \exp(-\gamma)
\]

(97)

Analogously, with \( \{U_{2,m,j} = 2, \ldots, M_2, j = 1, \ldots, L_2\} \) distributed according to \( P_{U_2} \) independent of the remaining random variables, Lemma 3 \((P_{U_2} Y_2, (M_2 - 1)L_2, \log M_2 L_2 + \gamma)\) yields

\[
\mathbb{P}[\mathcal{E}_2] \leq \mathbb{P} \left[ \cup \{u_i, y_i (U_{2,m,j} ; Y_2^*) > \log M_2 L_2 + \gamma \right\} \\
\leq \frac{(M_2 - 1)L_2}{M_2 L_2} \exp(-\gamma) \\
\leq \exp(-\gamma)
\]

(98)

where the union is over all \((m, j)\) with \( m \neq 1 \). We proceed to deal with the event \( \mathcal{E}_3 \cup \mathcal{E}_4 \):

\[
\mathbb{P}[\mathcal{E}_3 \cup \mathcal{E}_4] = \mathbb{E}[\mathcal{E}_3^c \cup \mathcal{E}_4^c] \\
= \int_0^1 \mathbb{P}[\mathcal{E}_3^c (U_1^*, U_2^*) > t] \, dt
\]

(101)

(102)

While \( U_1^* \sim P_{U_1} \), it does not hold that \( (U_1^*, U_2^*) \sim P_{U_1 U_2} \), so we cannot simply use the bound in (94) to upper bound the expectation in (101).

Letting

\[
\mathcal{F} = \{(u_1, u_2) \in U^2 : \chi(u_1, u_2) \leq t\}
\]

we deal with the probability in (102) using the covering lemma:

\[
\mathbb{P}[\mathcal{F}] \leq \mathbb{P} \left[ \bigcup_{j=1}^{L_2} \{u_j, y_{j} (U_{1,2,m,j} ; y) \notin \mathcal{F} \} \right] \\
\leq \mathbb{P}[\mathcal{F}] + e^{-\exp(\gamma)} + \mathbb{P}[u_j, y_{j} (U_{1,2}, y) > \log L_2 - \gamma]
\]

(104)

where inequality (105) follows from Lemma 5 \((P_{U_1 U_2}, L_2, \gamma, \mathcal{F})\).

Since \( \mathbb{P}[\{U_1, U_2 \notin \mathcal{F}\}] \leq \mathbb{P}[\{U_1, U_2 \} > t] \), integrating the right side of (105) with respect to \( t \), yields

\[
\mathbb{P}[\mathcal{E}_3 \cup \mathcal{E}_4] \leq \mathbb{E}[\chi(U_1, U_2)] + e^{-\exp(\gamma)} + \mathbb{P}[u_j, y_{j} (U_{1,2}, y) > \log L_2 - \gamma]
\]

(106)

We may now conclude that averaged over the random choice of the codebook arrays (88)-(89), the probability that either decoder selects the wrong codeword is upper bounded by the sum of (98), (100), and (106), which is equal to the right side of (87) in view of (94). Therefore, there must exist a realization of the arrays for which the corresponding code satisfies (87).

\[
\text{ REFERENCES}
\]


