A Theory of Network Equivalence –
Part II: Multiterminal Channels

Ralf Koetter, Fellow, IEEE, Michelle Effros, Fellow, IEEE, and Muriel Médard, Fellow, IEEE

Abstract—A technique for bounding the capacities of networks of independent channels is introduced. Parts I and II treat point-to-point and multiterminal channels, respectively. Bounds are derived using a new tool called a “bounding model.” Channel 1 is an upper (lower) bounding model for channel 2 if replacing channel 2 by channel 1 in any network yields a new network whose capacity region is a superset (subset) of the capacity region of the original network. This work derives bounding models from noiseless links, with lower bounding models corresponding to points in the channel’s capacity region and upper bounding models corresponding to points in a new channel characterization called an emulation region. Replacing all channels in a network by their noiseless upper (lower) bounding models yields a network of lossless links whose capacity region is a superset (subset) of the capacity region for the original network. This converts a general (often stochastic) network into a network coding instance, enabling the application of tools and results derived in that domain. A channel’s upper and lower bounding models differ when the channel can carry more information in some networks than in others. Bounding the difference between upper and lower bounding models bounds both the accuracy of the technique and the price of separating source-network coding from channel coding.

Keywords: Capacity, network coding, equivalence, bounding channel models

I. INTRODUCTION

Part I of this paper [1] defines upper and lower bounding channel models and derives such models for independent memoryless, point-to-point channels. There, the upper and lower bounds are identical, demonstrating the equivalence, from the perspective of network capacity, between a memoryless, point-to-point channel and an error-free link of the same capacity when each is employed in a larger network.

Part II derives bit-pipe models for independent memoryless multiterminal channels. (See Figure 1.) Each model is parameterized by a vector describing the capacities of its component bit pipes (lossless links). We show that a bit-pipe model is a lower bounding model for a multiterminal channel if and only if its rate vector lies in the channel’s capacity region. We then define the emulation region for a multiterminal channel to be the set of rate vectors describing all upper bounding models for that channel. This makes capacity and emulation regions natural complements: the capacity region characterizes bit pipes that the channel can emulate; the emulation region characterizes bit pipes that can emulate the channel.

The remainder of the paper is organized as follows. Section II defines a parameterized family of bit-pipe channels. Section III characterizes a channel’s capacity and emulation regions. Section IV derives simple emulation region bounds for four canonical channels. Section V investigates the accuracy of capacity bounds derived from bounding models. Examples appear in Section VI. The paper concludes in Section VII.

We apply notation and definitions from [1] throughout.

II. BOUNDING MODELS

As defined in [1, Definition 3], channel $C'$ is a lower bounding model for $C$ (written $C' \subseteq C$) if replacing independent channel $C$ by independent channel $C'$ in a larger network never increases the network capacity $(\mathcal{R}(C_0 \times C') \subseteq \mathcal{R}(C_0 \times C)$ for all channels $C_0$); channel $C'$ is an upper bounding model for $C$ (written $C' \supseteq C$) if...
Consider a multiterminal channel with inputs $x$ from nodes $V_1 = \{1,2\}$ to nodes $V_2 = \{3,4\}$. The channel distribution $p(y|x)$ factors as $p(y|x) = p(y|x_1)p(y|x_{2:4})$, where $p(y|x_{2:4})$ is the distribution on the remaining channel outputs given the remaining channel inputs.

Replacing independent channel $C$ by independent channel $C'$ in a larger network never decreases the network capacity ($\mathcal{R}(C_0 \times C') \geq \mathcal{R}(C_0 \times C)$ for all channels $C_0$); channels $C$ and $C'$ are equivalent (written $C = C'$) if $C'$ is both a lower bounding model and an upper bounding model for $C$. Different upper and lower bounding models are necessary when the rate that a channel can carry varies with the network in which that channel is employed (see, for example, [1, Examples 2 and 3]).

While bounding models can be noisy or noiseless, noiseless models are of particular interest since they enable capacity calculation for stochastic networks using computational tools for noiseless networks (e.g., [2]). The following definition, illustrated in Figure 2, generalizes bit pipes from the point-to-point connections (edges) used in Part 1 to the point-to-multipoint connections (hyperedges) used here. As in [1], a bit pipe of rate $R$ losslessly delivers $\lceil Rn \rceil$ bits in $n$ channel uses.

**Definition 1.** The rate-$R$ broadcast bit pipe $C(i, J, R)$ from node $i$ to (non-empty) receiver set $J$ is a device that noiselessly and simultaneously delivers a single rate-$R$ input from node $i$ to all nodes $j \in J$, giving

$$C(i, J, R) = \left(0, 1^R, \prod_{j \in J} \delta(y^{(j-1)} - x^{(i-1)}), \prod_{j \in J} \{0, 1\}^R \right).$$

Consider a multiterminal channel

$$C = (\mathcal{X}^{V_1}, p(y^{V_2|x^{V_1}}), \mathcal{Y}^{V_2})$$

with inputs $x^{V_1} = (x^{(i-1)} : i \in V_1)$ in alphabet $\mathcal{X}^{V_1} = \prod_{i \in V_1} \mathcal{X}^{(i-1)}$ and outputs $y^{V_2} = (y^{(j-1)} : j \in V_2)$ in alphabet $\mathcal{Y}^{V_2} = \prod_{j \in V_2} \mathcal{Y}^{(j-1)}$. The sets $V_1 \subseteq \mathcal{V}$ of transmitting nodes and $V_2 \subseteq \mathcal{V}$ of receiving nodes are assumed not to intersect ($V_1 \cap V_2 = \emptyset$).

The results that follow bound channel $C$ from below by bit-pipe channel $C(\mathcal{R}_L^{(V_1, V_2)})$ and from above by bit-pipe channel $C(\mathcal{R}_U^{(V_1, V_2)})$; both are defined formally below and illustrated in Figure 3. Each model has input nodes $V_1$, output nodes $V_2$, and a collection of internal nodes. Each internal node is denoted by $\sigma$ for some (non-empty) $A \subseteq V_1$. Input node $a \in V_1$ broadcasts a rate-$\mathcal{R}^{(A \rightarrow B)}$ description of input $X^{(a,1)} \in \mathcal{X}^{(a,1)}$ to all internal nodes $\sigma$ with $A \subseteq \{a\}$. For each (non-empty) $B \subseteq V_2$, internal node $\sigma$ broadcasts a rate-$\mathcal{R}^{(A \rightarrow B)}$ description of $X^{(a,1)} : a \in A$ to receiver set $B$.

The models differ in their internal nodes. The lower bounding model has an internal node for each single-element subset $A$ of $V_1$. The upper bounding model has an internal node for each non-empty subset $A$ of $V_1$.

**Definition 2.** Given a multiterminal channel

$$C = (\mathcal{X}^{V_1}, p(y^{V_2|x^{V_1}}), \mathcal{Y}^{V_2})$$

and a rate vector

$$\mathcal{R}_L^{(V_1, V_2)} = (\mathcal{R}_L^{(\sigma \rightarrow B)} : a \in V_1, B \subseteq V_2),$$

lower bounding bit-pipe channel $C(\mathcal{R}_L^{(V_1, V_2)})$ has transmitters $V_1$, receivers $V_2$, and internal nodes

$$V_{0,L} = \{\sigma : a \in V_1\}.$$

Each transmitter $a \in V_1$ delivers input $X^{(a,1)} \in \mathcal{X}^{(a,1)}$ to internal node $\sigma^{(a)}$ at rate log $|\mathcal{X}^{(a,1)}|$. For each (non-empty) $B \subseteq V_2$, each internal node $\sigma^{(a)} \in V_{0,L}$ broadcasts a rate-$\mathcal{R}_L^{(\sigma \rightarrow B)}$ description of $X^{(a,1)}$ to the nodes in $B$. Thus,

$$C(\mathcal{R}_L^{(V_1, V_2)}) = \prod_{a \in V_1} C(a, \{\sigma^{(a)}\}, \log|\mathcal{X}^{(a,1)}|) \cdot \prod_{a \in V_1} \prod_{B \subseteq V_2} C(\sigma^{(a)}, B, \mathcal{R}_L^{(\sigma \rightarrow B)}).$$

1 The arguments go through unchanged (though with greater notational overhead) when $V_1 \cap V_2 \neq \emptyset$.\[
\]
Definition 3. Given a multiterminal channel
\[ C = (X^{V_1}, p(y^{V_2}|x^{V_1}), \mathcal{Y}^{V_2}) \]
and a rate vector
\[ \mathcal{R}^{(V_1,V_2)} = (R^{(A \rightarrow B)}_U : A \subseteq V_1, B \subseteq V_2), \]
upper bounding bit-pipe channel \( C(\mathcal{R}^{(V_1,V_2)}) \) has transmitters \( V_1 \), receivers \( V_2 \), and internal nodes
\[ V_{0,U} = \{ \sigma^A : A \subseteq V_1 \}. \]
Each transmitter \( a \in V_1 \) broadcasts a rate-log \( |X^{(a,1)}| \) description of input \( X^{(a,1)} \in X^{(a,1)} \) to internal nodes \( \{ \sigma^A \in V_{0,U} : A \supseteq \{ a \} \} \). For each non-empty \( B \subseteq V_2 \), each internal \( \sigma^A \in V_{0,U} \) broadcasts a rate- \( R^{(A \rightarrow B)}_U \) description of \( (X^{(a,1)} : a \in A) \) to the nodes in \( B \). Thus
\[ C(\mathcal{R}^{(V_1,V_2)}) = \prod_{a \in V_1} C(a, \{ \sigma^A \in V_{0,U} : A \supseteq \{ a \} \}, \log |X^{(a,1)}|) \cdot \prod_{A \subseteq V_1} \prod_{B \subseteq V_2} C(\sigma^A, B, R^{(A \rightarrow B)}_U). \]

Lemma 1 Given \( m \)-node network \( N \), for each \( v \in [m] \) let \( (y^{(v,1)}, \ldots, y^{(v,k(v))}) \) denote the channel outputs at node \( v \) and let \( \mathbf{d} = (d(v,k) : v \in [m], k \in [k(v)]) \) be a vector of finite, non-negative integers. If \( N(\mathbf{d}) \) is the network \( N \) modified to delay network output \( y^{(v,k)} \) by \( d(v,k) \) time steps, then
\[ \mathcal{R}(N) = \mathcal{R}(N(\mathbf{d})). \]

Proof. See Appendix I. \( \blacksquare \)

III. LOWER AND UPPER BOUNDING MODELS

Given the bit-pipe channel definitions from Section II, we next characterize the rate vectors \( \mathcal{R}^{(V_1,V_2)}_L \) and \( \mathcal{R}^{(V_1,V_2)}_U \) that yield lower and upper bounding models. We begin by showing that \( C(\mathcal{R}^{(V_1,V_2)}_L) \) is a lower bounding model for \( C \) if and only if \( \mathcal{R}^{(V_1,V_2)}_L \) is in the capacity region of \( C \). The proof, much like that of [1, Lemma 5], involves showing that channel coding on \( C \) can emulate \( C(\mathcal{R}^{(V_1,V_2)}_L) \) so accurately that any sequence of solutions with asymptotically negligible error probability on \( C_0 \times C(\mathcal{R}^{(V_1,V_2)}_L) \) can be run reliably on \( C_0 \times C \).

Lemma 2 Given channel \( C = (X^{V_1}, p(y^{V_2}|x^{V_1}), \mathcal{Y}^{V_2}) \),
\[ C(\mathcal{R}^{(V_1,V_2)}_L) \subseteq C \text{ if and only if } \mathcal{R}^{(V_1,V_2)}_L \subseteq \mathcal{R}(C). \]

Proof. See Appendix II. \( \blacksquare \)

While the capacity region of a channel \( C \) is traditionally defined as the closure of the set of rate vectors achievable on channel \( C \), Lemma 2 suggests an alternative definition: The capacity region of channel \( C \) is the set of rate vectors for all lower bounding models of \( C \); that is,
\[ \mathcal{R}(C) = \{ \mathcal{R}^{(V_1,V_2)}_L : C(\mathcal{R}^{(V_1,V_2)}_L) \subseteq C \}. \]
The corresponding set for upper bounding models is defined as follows.

Definition 4. Given channel
\[ C = (X^{V_1}, p(y^{V_2}|x^{V_1}), \mathcal{Y}^{V_2}), \]
the emulation region of \( C \) is
\[ \mathcal{E}(C) \overset{\text{def}}{=} \{ \mathcal{R}^{(V_1,V_2)}_U : C(\mathcal{R}^{(V_1,V_2)}_U) \supseteq C \}. \]

Since the rate region describing lower bounding models is closed, it is tempting to seek a closure in the rate region for upper bounding models as well. Current techniques for proving closure rely on the continuity of capacity \( \mathcal{R}(C_0 \times C(\mathcal{R}^{(V_1,V_2)}_L)) \) in the components \( \mathcal{R}^{(A \rightarrow B)}_U \).
of vector $R_{U}^{(V_{1},V_{2})}$. To date, $\mathcal{A}(C \times C(R_{U}^{(V_{1},V_{2})}))$ is known to be continuous in $R_{U}^{(A \rightarrow B)}$ only for (strictly) positive $R_{U}^{(A \rightarrow B)}$ [1, Lemma 4]. Since continuity remains an open question when $R_{U}^{(V_{1},V_{2})}$ approaches zero in one or more dimensions [6], [7], Lemma 3 employs the following definition of partial closure, which takes the closure only where continuity is known to apply.

**Definition 5** For any $k \geq 1$, the partial closure of set $A \subseteq \mathbb{R}^{k}$, denoted by $\bar{A}$, is defined as

$$\bar{A} = \{a = (a_{1}, \ldots, a_{k}) \in \mathbb{R}^{k} : \exists \{b_{n} = (b_{n,1}, \ldots, b_{n,k})\}_{n=1}^{\infty} \subseteq A \text{ s.t.} \lim_{n \to \infty} b_{n} = a \text{ and } a_{i} = 0 \Rightarrow b_{n,i} = 0 \text{ for all } n\}.$$

**Lemma 3** Given channel $C = (\mathcal{X}^{V_{1}}, p(y^{V_{2}}|x^{V_{1}}), \mathcal{Y}^{V_{2}})$, if $E \subseteq \mathcal{E}(C)$, then $\bar{E} \subseteq \mathcal{E}(C)$.

**Proof.** Given any $E \subseteq \mathcal{E}(C)$ and $R_{U}^{(V_{1},V_{2})} \in \bar{E} \setminus E$, there exists a sequence $\{R_{U,n}^{(V_{1},V_{2})}\}_{n=1}^{\infty} \subseteq E$ with

$$R_{U,n}^{(V_{1},V_{2})} = (R_{U,n}^{(A \rightarrow B)} : A \subseteq V_{1}, B \subseteq V_{2})$$

$$\lim_{n \to \infty} R_{U,n}^{(V_{1},V_{2})} = R_{U}^{(V_{1},V_{2})}$$

such that if $R_{U,o}^{(A \rightarrow B)} = 0$ then $R_{U,n}^{(A \rightarrow B)} = 0$ for all $n$. Since $\mathcal{A}(C \times C(R_{U}^{(V_{1},V_{2})}))$ is continuous in $R_{U}^{(A \rightarrow B)}$ for all $R_{U}^{(A \rightarrow B)} > 0$ [1, Lemma 4] and closed by definition,

$$\mathcal{A}(C \times C(R_{U}^{(V_{1},V_{2})})) \supseteq \mathcal{A}(C \times C)$$

for all $n$ implies

$$\mathcal{A}(C \times C(R_{U,o}^{(V_{1},V_{2})})) \supseteq \mathcal{A}(C \times C).$$

Just as Lemma 2 derives lower bounding models for $C$ by showing bit-pipe models that channel $C$ can emulate, Theorem 5, below, derives upper bounding models for $C$ by demonstrating bit-pipe models that can emulate channel $C$. The tool for emulating channel $C$ across bit-pipe channel $C(R_{U}^{(V_{1},V_{2})})$ is called a rate-$R_{U}$ “emulation code.” An emulation code is similar to a lossy source code. The encoders map input $x^{V_{1}}$ to a rate-$R_{U}^{(V_{1},V_{2})}$ binary description, and the decoders map the binary description to a reproduction $\hat{y}^{V_{2}}$ of $\tilde{y}^{V_{2}}$. In the construction that follows, the emulation code is designed to approximate the channel behavior $p(y^{V_{2}}|x^{V_{1}})$ in the sense that its end-to-end mapping from $\tilde{y}^{V_{1}}$ to $\tilde{y}^{V_{2}}$ maps typical channel inputs to jointly typical channel outputs. As in [1], the typical set used here is a “restricted typical set” $\hat{A}_{c}^{(N)}$; it is identical to the traditional typical set $A_{c}^{(N)}$ except that elements for which the conditional probability of atypicality is high (greater than $\frac{1}{2}$) are removed. Appendix III defines the set $\hat{A}_{c}^{(N)}$ used here and shows that the probability of $(\hat{A}_{c}^{(N)})^{c}$ decays expo-
nentially in $N$. In that definition and throughout, notation $p \in 2^{-N(H(\pm \varepsilon))}$ is used to mean $2^{-N(H+\varepsilon)} \leq p \leq 2^{-N(H-\varepsilon)}$.

Like source and channel coding theorems, emulation coding theorems are proven using random code designs. Showing that the expected performance of a randomly designed emulation code is good proves the existence of a good code. Instead of repeating this argument for each channel, we derive a property of the random code design sufficient to guarantee the existence of good codes.

**Definition 6** A $(2^{NR}(V_1, V_2), N)$ emulation code $(\alpha, \beta)$ for channel $C = (X^{V_1}, p(y^{V_2}|x^{V_1}), Y^{V_2})$ comprises encoders $\alpha = (\alpha^{(A \rightarrow B)}: A \subseteq V_1, B \subseteq V_2)$ and decoders $\beta = (\beta^{(j)} : j \in V_2)$, where

$$\alpha^{(A \rightarrow B)} = \prod_{i \in A} \mathcal{A}^{(i,1)} \rightarrow \mathcal{W}^{(A \rightarrow B)}$$

$$\beta^{(j)} : (A, B) : A \subseteq V_1, (j) \subseteq B \subseteq V_2 \rightarrow \mathcal{Y}^{(j,1)}$$

and $\mathcal{W}^{(A \rightarrow B)} = [2^{NR(U^{A\rightarrow B})}]$ for all non-empty $A \subseteq V_1, B \subseteq V_2$.

When a $(2^{NR}(V_1, V_2), N)$ emulation code $(\alpha, \beta)$ is chosen at random from the family of possible codes, the random code design induces an emulation distribution

$$\hat{p}_N(y^{V_2}|x^{V_1}) = \Pr(\beta(\alpha(x^{V_1})) = y^{V_2}).$$

Given any distribution $p(x^{V_1}) = \prod_{i=1}^{N} p(x_i^{V_1})$ in the set $\mathcal{P}(C)$ of allowed memoryless input distributions for channel $C$ and any constants $\epsilon > 0$ and $c > 1$, the emulation success event $\mathcal{S}^{(N)}(\epsilon, c)$ under input distribution $p(x^{V_1})$ is defined as

$$\mathcal{S}^{(N)}(\epsilon, c) = \left\{(x^{V_1}, y^{V_2}) \in \mathcal{A}_N^{(N)} : p(x^{V_1})\hat{p}_N(y^{V_2}|x^{V_1}) \in 2^{-N(H(X^{V_1}, Y^{V_2}) + \epsilon c)}\right\},$$

and the emulation error probability is defined as

$$Q^{(N)}(\epsilon, c) = \sum_{(x^{V_1}, y^{V_2}) \hat{p}_N(\mathcal{S}^{(N)}(\epsilon, c))} p(x^{V_1})\hat{p}_N(y^{V_2}|x^{V_1}).$$

**Definition 7** Rate vector $\mathcal{R}_U^{(V_1, V_2)}$ suffices to exponentially approximate channel $C$ if for any input distribution $p \in \mathcal{P}(C)$ there exist a constant $c > 1$ and a positive function $\eta(c)$ such that for any $\epsilon > 0$ there is a random $(2^{NR}(V_1, V_2), N)$ emulation code design with expected error probability $Q^{(N)}(\epsilon, c) < 2^{-\eta(c)N}$ for all $N$ sufficiently large. Let

$$\mathcal{S}^{(\alpha)}(C) = \left\{\mathcal{R}_U^{(V_1, V_2)} : \mathcal{R}_U^{(V_1, V_2)} \text{ suffices to exponentially approximate channel } C\right\}.$$

Roughly, Definition 6 declares an emulation code to be successful if $\epsilon$-typical events under the channel distribution remain $\epsilon c$-typical under the emulation distribution, and Definition 4 specifies the set of rates for which the probability of emulation failure can be made to decay exponentially to zero for some constant $c > 1$ and every $\epsilon$. Theorem 5, below, shows that all rates sufficient to exponentially approximate $C$ fall in the emulation region; that is, exponential decay of the failure probability under the given definition of success suffices to guarantee an upper bounding model. Whether this criterion is also necessary remains an open question.

Before turning to Theorem 5, we pause to prove the convexity of set $\mathcal{S}^{(\alpha)}(C)$.

**Lemma 4** Region $\mathcal{S}^{(\alpha)}(C)$ is convex for any channel $C = (X^{V_1}, p(y^{V_2}|x^{V_1}), Y^{V_2})$.

**Proof.** Given any $\mathcal{R}_U^{(V_1, V_2)}, \mathcal{R}_U^{(V_1, V_2)} \in \mathcal{S}^{(\alpha)}(C)$ and $\lambda \in (0, 1/2]$, let

$$\mathcal{R}_U^{(V_1, V_2)} = \lambda \mathcal{R}_U^{(V_1, V_2)} + (1 - \lambda) \mathcal{R}_U^{(V_1, V_2)}.$$

We wish to prove that $\mathcal{R}_U^{(V_1, V_2)} \in \mathcal{S}^{(\alpha)}(C)$.

For $p \in \mathcal{P}(C)$ and $k \in \{1, 2\}, \mathcal{R}_U^{(V_1, V_2)} \in \mathcal{S}^{(\alpha)}(C)$ implies that there exists a constant $c_k \geq 1$ and positive function $\eta_k(c)$ for which random $(2^{NR}(V_1, V_2), N)$ emulation code design achieves error probability $Q^{(N)}(\epsilon, c_k) < 2^{-N\eta_k(c)}$ for any $\epsilon > 0$ and $N$ sufficiently large. Let $(\alpha_k, \beta_k)$ be the randomly drawn code and

$$\hat{p}_{k,N}(y^{V_2}|x^{V_1}) = \Pr(\beta_k(x^{V_1}) = y^{V_2})$$

be the emulation distribution. The success event and error probability for a fixed $p, \epsilon$, and $c_k$ are

$$\mathcal{S}^{(N)}(\epsilon, c) = \left\{(x^{V_1}, y^{V_2}) \in \mathcal{A}_N^{(k)} : p(x^{V_1})\hat{p}_{k,N}(y^{V_2}|x^{V_1}) \in 2^{-N(H(X^{V_1}, Y^{V_2}) + \epsilon c_k)}\right\}$$

$$Q^{(k)}(\epsilon, c) = \sum_{(x^{V_1}, y^{V_2}) \hat{p}_{k,N}(\mathcal{S}^{(N)}(\epsilon, c))} p(x^{V_1})\hat{p}_{k,N}(y^{V_2}|x^{V_1}).$$

Let $N_1 = \lfloor \lambda N \rfloor$ and $N_2 = N - N_1$. Define $(\alpha_{N_1}, \beta_{N_1})$ to be the blocklength-$N_1$ emulation code achieved by independently applying $(\alpha_1, \beta_1, \alpha_{N_1})$ in the first $N_1$ time steps and $(\alpha_{N_2}, \beta_{N_2})$ in the following $N_2$ time steps.
with expected error probability
(n designed in Step 1) with the emulation codes

\[
\mathcal{E}(\lambda N, V) = (\lambda V^1(N_1), \ldots, \lambda V^2(N)) \quad \text{and} \quad \mathcal{Y}(\lambda N, V) = (\lambda V^1(N_1 + 1), \ldots, \lambda V^2(N)).
\]

The code’s emulation distribution is

\[
\hat{p}_N(\mathcal{Y}|\mathcal{X}) = \Pr( (\beta_{1,N_1}(\alpha_{1,N_1}(\mathcal{X})), \beta_{2,N_2}(\alpha_{2,N_2}(\mathcal{X}))) = (\mathcal{Y}_{(1)}, \mathcal{Y}_{(2)})) \quad \text{and} \quad \Pr( (\beta_{1,N_1}(\alpha_{1,N_1}(\mathcal{X}))), \beta_{2,N_2}(\alpha_{2,N_2}(\mathcal{X}))) = (\mathcal{Y}_{(1)}, \mathcal{Y}_{(2)}).
\]

To bound the emulation error probability of \((\alpha_N, \beta_N)\) under input distribution \(p\), note that if \(c = c_1 + c_2\), then an emulation error can occur on \((\alpha_N, \beta_N)\) only if there is an emulation error on at least one of \((\alpha_{N_1}, \beta_{N_1})\) and \((\alpha_{N_2}, \beta_{N_2})\). Therefore, by the union bound,

\[
Q^{(N)}(\epsilon, c_1 + c_2) \leq Q_1^{(N_1)} + Q_2^{(N_2)} \leq 2^{-N_1 \eta_1(\epsilon) + 2^{-N_2 \eta_2(\epsilon)}}.
\]

(a) follows since \(\lambda \leq 1/2\) implies \(\min\{N_1, N_2\} \geq \lambda N - 1\), and (b) holds for all \(N\) sufficiently large. The given exponential error bound proves the desired result.

**Theorem 5** Let \(C = (\lambda V^1, \mathcal{X}|\mathcal{Y})\). Then

\[
\hat{p}_N(C) \subseteq \hat{p}(C).
\]

**Proof.** The proof is similar to [1, Theorem 6]. Fix \(C_0\) and \(\mathcal{R}^{(V_1,V_2)} \in \hat{p}(C)\). Let \(\mathcal{N} = C_0 \times C\) and \(\mathcal{N}_U = C_0 \times \mathcal{C}(\mathcal{R}^{(V_1,V_2)})\). Steps 1-5, below, show that int(\(\mathcal{R}(\mathcal{N})\)) \(\subseteq \mathcal{R}(\mathcal{N}_U)\); this implies \(\mathcal{R}(\mathcal{N}) \subseteq \mathcal{R}(\mathcal{N}_U)\) since \(\mathcal{R}(\mathcal{N}_U)\) is closed by definition. Then \(\mathcal{R}(\mathcal{N}) \subseteq \mathcal{R}(\mathcal{N}_U)\) follows since \(\mathcal{R}(\mathcal{N}_U) = \mathcal{R}(\mathcal{N}_U)\). Finally \(\hat{p}_N(C) \subseteq \hat{p}(C)\) by Lemma 3.

Step 1 - Choose code \(S(\mathcal{N})\); define \(p_t(x^{V_1}, y^{V_2})\). Fix \(R \in \text{int}(\mathcal{R}(\mathcal{N}))\). By [1, Theorem 2], there exists a \(\delta > 0\) and a solution \(S(\mathcal{N})\) of some finite blocklength \(n\) from which we can build, for each \(N\) sufficiently large, a rate-\(R\) stacked solution \(S(\mathcal{N})\) for \(N\)-fold stacked network \(\mathcal{N}\) with expected error probability \(P_e^{(N)} \leq 2^{-N^\delta}\). Stacked solution \(S(\mathcal{N})\) employs inner and outer codes. The outer code applies an independent, randomly designed channel code to each message \(\mathcal{W}^{(u)}(u)\). The inner code independently applies solution \(S(\mathcal{N})\) in each layer of \(\mathcal{N}^\mathcal{U}\). For each \(t \in [n]\), let \(p_t(x^{V_1})\) be the input distribution on channel \(C\) at time \(t\) under solution \(S(\mathcal{N})\), and define \(p_t(x^{V_1}, y^{V_2}) = p_t(x^{V_1})p(y^{V_2}|x^{V_1})\). Then \(p_t(x^{V_1}, y^{V_2}) = \prod_{i=1}^{n} p_t(x^{V_1}(i), y^{V_2}(i))\) is the time-\(t\) distribution across the \(N\) copies of channel \(C\) in \(N\)-fold stacked network \(\mathcal{N}\) under solution \(S(\mathcal{N})\).

Step 2 - Choose \(\epsilon_t\) and channel emulators \((\alpha_{N,t}, \beta_{N,t})\).

By definition of \(\hat{p}(\alpha_{N,t}, \beta_{N,t})\), \(\mathcal{R}^{(V_1,V_2)} \subseteq \hat{p}(\alpha_{N,t}, \beta_{N,t})\) implies that for each input distribution \(p_t(x^{V_1})\) there exists a constant \(c_t > 1\) and positive function \(\eta_t(\epsilon)\) such that for any \(\epsilon > 0\) there is a random \((2^{N_1 \eta_1(\epsilon)} \times N)\) emulation code design with expected error probability \(Q^{(N)}(\epsilon, c_t) = 2^{-\eta_t(\epsilon)n}\) on input distribution \(p_t(x^{V_1})\) for all \(N\) sufficiently large. Let \(c_t = \delta + 1\) \(t\) such that \(\delta\) is the error exponent for \(P_e^{(N)}\) from Step 1. For each \(t\) from \(n-1\) down to 1, set

\[
\epsilon_t = \min\{\delta, \eta_t(\epsilon_t), \ldots, \eta_t(\epsilon_n)\}.
\]

This choice guarantees

\[
\sum_{t=1}^{n} (\epsilon_t + 1)\epsilon_t < \eta_t(\epsilon_t) \quad \forall t \in [n] \quad (1)
\]

\[
\sum_{t=1}^{n} (\epsilon_t + 1)\epsilon_t < \delta \quad (2)
\]

both are useful for Step 5, below.

For each \(t\), randomly design a code \((\alpha_{N,t}, \beta_{N,t})\) to meet the constraint \(Q^{(N)}(\epsilon_t, c_t) < 2^{-\eta_t(\epsilon_t)n}\) for input distribution \(p_t(x^{V_1})\) and \(N\) sufficiently large. Let \(\mathcal{X}_{N,t}(\mathcal{Y}|\mathcal{X})\) be the resulting emulation distribution.

Step 3 - Define solution \(S(\mathcal{N}_U)\).

Solution \(S(\mathcal{N}_U)\) combines the solution \(S(\mathcal{N})\) for network \(\mathcal{N}\) (designed in Step 1) with the emulation codes \(\{(\alpha_{N,t}, \beta_{N,t})\}_{t=1}^{n}\) for channel \(C\) (designed in Step 2).

For each node \(v \in \mathcal{V} \cup \mathcal{V}_0\), let \((\tilde{X}_t(v), \tilde{Y}_t(v))\) be the time-\(t\) node-\(v\) input and output in \(\mathcal{N}_U\). At time \(t\), nodes \(v \in \mathcal{V}_2\) emulation decode network outputs \(\tilde{Y}_t(v)\), giving

\[
\hat{Y}_t(v) = \begin{cases} \tilde{Y}_t(v), & \text{if } v \in \mathcal{V}_2 \text{ and } t \leq \lambda N_1 \varepsilon_t, \\ \hat{Y}_t(v), & \text{otherwise.} \end{cases}
\]

Each node \(v \in \mathcal{V}\) then applies the node encoder \(\hat{X}_t(v)\); while each node \(\sigma^A \in \mathcal{V}_0\) applies the emulation encoders \(\alpha_{N,t}^A = (\alpha_{N,t}^{A-B})(B \subseteq \mathcal{V}_2)\), giving

\[
\hat{X}_t(v) = \begin{cases} \hat{X}_t(v), & \text{if } v \in \mathcal{V}_2 \text{ and } t \leq \lambda N_1 \varepsilon_t, \\ \hat{X}_t(v) \alpha_{N,t}^{y(v) \rightarrow x}, & \text{if } v = \sigma \in \mathcal{V}_0. \end{cases}
\]
as the channel input. After time $n$, each node $v \in V$ applies decoders

$$\hat{W}^{(v)}(\{u\rightarrow V\}, v) : (u, V) \in \mathcal{M}, v \in V$$

from $\mathbf{S}(\mathcal{N})$ to build message reconstruction

$$\hat{W}^{(v)} = \hat{W}^{(v)}(Y^{(v)}_1, n, W^{(v)}_{(v)\rightarrow \epsilon}).$$

Step 4 - Characterize the behavior of $\mathbf{S}(\mathcal{N}(\mathcal{R}_t^{(V_1, V_2)}))$, under the operation of the randomly designed emulation code, independent operation of node decoders.

$p$ is a product distribution describing the independent operation of the network, respectively, at time $t$. Here $p$ is the product distribution describing the independent operation of node decoders.

Only the channel distribution changes when we run $\mathbf{S}(\mathcal{N}(\mathcal{U}))$ on $\mathcal{N}(\mathcal{U})$. In $\mathcal{N}(\mathcal{U})$, channel $C$ is replaced by the operation of the randomly designed emulation code, giving

$$\hat{p}_N(w, x^n, y^n, \hat{w}) = p(w) \prod_{t=1}^n p(x_t | y_t^{t-1}, w) \prod_{t=1}^n \hat{p}_N(t, y_t^{V_1} | y_t^{V_2} | \hat{y}_t^{V_1} | p(y_t^{V_2} | y_t^{V_1}) | p(\hat{w} | y^n, w).$$

Here $p(w)$ is the distribution on messages, $p(x_t | y_t^{t-1}, w)$ is a product distribution describing the independent operation of node encoders at time $t$, $p(y_t^{V_1} | y_t^{V_2} | \hat{y}_t^{V_1}$ and $p(y_t^{V_2} | \hat{y}_t^{V_1}$ are the distributions for channel $C$ and the rest of the network, respectively, at time $t$, and $p(\hat{w} | y^n, w)$ is the product distribution describing the independent operation of node decoders.

Let $\hat{A}_t^{(N)}, \hat{S}_t^{(N)}$, and $Q_t^{(N)}$ be the restricted typical set, success event, and error probability, respectively, under input distribution $p_t(\hat{z}_t^{V_1})$, parameters $\epsilon_t$ and $\epsilon_t$, and emulation code $(\alpha_t^{N,t}, \beta_t^{N,t})$. Then

$$\hat{S}_t^{(N)} = \{(\hat{z}_t^{V_1}, \hat{z}_t^{V_2}) \in \hat{A}_t^{(N)} : p_t(\hat{z}_t^{V_1}) \hat{p}_N(t, \hat{y}_t^{V_2} | \hat{y}_t^{V_1}) \leq 2^{-N(H(\hat{X}_t^{V_1}, \hat{Y}_t^{V_1} + \epsilon_t + \epsilon_t)} \right\}.$$
\[E \left[ \sum_{t_0=1}^{n} \sum_{(w, x^n, y^n) \in \mathcal{A}_{t_0}} \hat{p}_N(w, x^n, y^n) + \sum_{(w, x^n, y^n, \bar{w}) \in \mathcal{B}} \hat{p}_N(w, x^n, y^n, \bar{w}) \right] \]

\[= E \left[ \sum_{t_0=1}^{n} \sum_{(w, x^n, y^n) \in \mathcal{A}_{t_0}} \hat{p}_N(w, x^n, y^n) + \sum_{(w, x^n, y^n, \bar{w}) \in \mathcal{B}} \hat{p}_N(w, x^n, y^n, \bar{w}) \right] \]

\[\leq \sum_{t_0=1}^{n} 2^{-N \epsilon_0} + 2^{-N \epsilon_1} + 2^{-N \epsilon_2} \sum_{t_0=1}^{n} \sum_{(p, y^n, w) \in \mathcal{B}} \hat{p}_N(w, x^n, y^n, \bar{w}) \]

\[= 2^{-N \epsilon_0} + 2^{-N \epsilon_1} + 2^{-N \epsilon_2} \sum_{t_0=1}^{n} \sum_{(p, y^n, w) \in \mathcal{B}} \hat{p}_N(w, x^n, y^n, \bar{w}) \]

Here \((a)\) breaks the event “emulation fails for one or more times \(t^n\) into the union over \(t_0\) of the events “the first time emulation fails is time \(t_0\),” \((b)\) applies \((3)\), and \((c)\) replaces \(\mathcal{A}_t\) and \(\mathcal{B}\) by supersets \(\mathcal{A}'_t\) and \(\mathcal{B}'\). The bound goes to zero as \(N\) grows by our choice of \(\{\epsilon_t\}_{t=1}^{\infty}\), which guarantees that all exponents are negative as shown in \((1)\) and \((2)\). Since the expected error probability under the random emulation and channel code designs approaches 0, there exists a single sequence of codes that does at least as well. Thus \(R \in \mathcal{B}(N(R_{U,1}^{(V_2,V_2)})]. \]

\[\text{IV. EMULATION REGION BOUNDS}\]

This section describes emulation region bounds for four canonical channels. Notation \(\text{con}(A)\) denotes the partial closure of the convex hull of set \(A\).

\[\text{Lemma 6 Let} \]

\[C = (\mathcal{X}^{(i,1)}), \mathcal{Y}^{(j,1)})\]

\[\text{be a point-to-point channel with capacity } C > 0. \text{ Then} \]

\[E(C) = \{R_{U,1}^{(i,1)}; R_{U,1}^{(i,1)} \geq C\}. \]

\[\text{Proof: The proof of [1, Theorem 6] contains a random design algorithm for emulation codes. For each rate parameter } R_{U,1}^{(i,1)} \geq C, \text{ input distribution } p \in \mathcal{P}(C), \text{ and constant } \epsilon > 0, \text{ the algorithm achieves} \]

\[Q^{(N)}(\epsilon, 10) < 2^{-N \eta(\epsilon)} \]

for positive function \(\eta(\epsilon)\) and \(N\) sufficiently large. Applying the partial closure gives the set described above. Since \(C\) is also in the capacity region \(\mathcal{B}(C)\) of \(C\), the given set describes the full emulation region. 

The next emulation region bound is for the broadcast channel. Unlike the capacity region, which depends only on the conditional marginals at the receivers, the emulation region varies with the conditional distribution on both outputs given the channel input.

\[\text{Theorem 7 Given broadcast channel} \]

\[C = (\mathcal{X}^{(i,1)}, \mathcal{Y}^{(j,1)}, \mathcal{Y}^{(j,2)}, \mathcal{Y}^{(j,1)} \times \mathcal{Y}^{(j,2)}), \]

\[\text{8}\]
Given multiple access channel

\[ \text{Theorem 8} \]

rely on auxiliary random variables.

Source coding problems, emulation region bounds here

interference channels follow. As in many multi-input

let

\[
C_1 = \max_{p(x^{(i,1)})} I(X^{(i,1)}, Y^{(j,1)})
\]

\[
C_2 = \max_{p(x^{(i,1)})} I(X^{(i,1)}, Y^{(j,2)})
\]

\[
C_{12} = \max_{p(x^{(i,1)})} I(X^{(i,1)}, Y^{(j,1)}, Y^{(j,2)})
\]

Proof. See Appendix IV-A. ■

Emulation region bounds for the multiple access and interference channels follow. As in many multi-input source coding problems, emulation region bounds here rely on auxiliary random variables.

**Theorem 8** Given multiple access channel

\[ C = (X^{(i,1)} \times X^{(i,2)}, p(y^{(j,1)}|x^{(i,1)}, x^{(i,2)}), Y^{(j,1)}) \]

let

\[ \mathcal{E}_1 = \left\{ R^{(i,1,2)}_{U_i} \{j, j_2\} : \text{for each } p(x^{(i,1)}, x^{(i,2)}) \right\}
\]

there exists \( p(u]\ | x^{(i,1)} \) s.t. \( |U| \leq |X^{(i,1)}| \),

\[ R^{(i,1,2)}_{U_i} \{j\} > I(X^{(i,1)}, U), \]

\[ R^{(i,1,2)}_{U_i} \{j, j_2\} > I(X^{(i,1)}, X^{(i,2)}, Y^{(j,1)}|U) \}

\[ \mathcal{E}_2 = \left\{ R^{(i,1,2)}_{U_i} \{j\} : \text{for each } p(x^{(i,1)}, x^{(i,2)}) \right\}
\]

there exists \( p(u]\ | x^{(i,2)} \) s.t. \( |U| \leq |X^{(i,2)}| \),

\[ R^{(i,1,2)}_{U_i} \{j\} > I(X^{(i,2)}, U), \]

\[ R^{(i,1,2)}_{U_i} \{j, j_2\} > I(X^{(i,1)}, X^{(i,2)}, Y^{(j,1)}|U) \}

Then \( \mathcal{E}(C) \supseteq \text{con}(\mathcal{E}_1 \cup \mathcal{E}_2) \).

Proof. See Appendix IV-B. ■

**Theorem 9** For interference channel

\[ C = (X^{(i,1)} \times X^{(i,2)}, p(y^{(j,1)}, y^{(j,2)}|x^{(i,1)}, x^{(i,2)}), Y^{(j,1)} \times Y^{(j,2)}) \]

let

\[ \mathcal{E}_1 = \left\{ R^{(i,1,2)}_{U_i} \{j, j_2\} : \text{for each } p(x^{(i,1)}, x^{(i,2)}) \right\} \]

\[ \exists p(u_2|x^{(i,1)}) \text{ and } p(u_1|x^{(i,1)}, u_2) \text{ s.t.} \]

\[ |U_1 \times U_2| \leq |X^{(i,1)}| \]

\[ R^{(i,1)}_{U_1} \{j\} > I(X^{(i,1)}, U_2) \]

\[ R^{(i,2)}_{U_1} \{j\} > I(X^{(i,1)}, X^{(i,2)}, Y^{(j,1)}|U_2) \]

\[ R^{(i,1)}_{U_1} \{j\} > I(X^{(i,1)}, X^{(i,2)}, Y^{(j,1)}|U_2) \]

Let \( \mathcal{E}_2 \) be the region that results when the roles of \( i_1 \) and \( i_2 \) are reversed in \( \mathcal{E}_1 \). Let \( \mathcal{E}_3 \) and \( \mathcal{E}_4 \) be the regions that result when the roles of \( j_1 \) and \( j_2 \) are reversed in \( \mathcal{E}_1 \) and \( \mathcal{E}_2 \), respectively.

Then \( \mathcal{E}(C) \supseteq \text{con}(\mathcal{E}_1 \cup \mathcal{E}_2 \cup \mathcal{E}_3 \cup \mathcal{E}_4) \).

Proof. See Appendix IV-C. ■

**V. Model Accuracy**

Two types of inaccuracy arise in applying bounding models: inaccuracies in the calculation of \( \mathcal{R}(C) \) and \( \mathcal{E}(C) \) and inaccuracies inherent to the use of bounding models. The first case occurs when capacity and emulation regions are difficult to characterize. Here, achievability results give inner bounds on capacity and emulation regions and thus provide lower and upper bounding models, respectively. Any converse can then be used to bound model accuracy. For example, cut-set bounds give an upper bound on capacity and a lower bound on emulation, enabling bounds on the accuracy of both lower and upper bounding models [8].

The second type of inaccuracy results from the generality of bounding models. Finding a bound that applies under all possible network scenarios sometimes implies that the bounds cannot be simultaneously tight. For example in a multiple access channel, the lower bound may be loose when the same device controls both channel inputs; the upper bound may be loose when no node can send information to both transmitters.

Given a channel \( C = (X^{V_1}, p(y^{V_2}|x^{V_1}), Y^{V_2}) \) and network \( C_0 \times C \), any pair \( (R^{(V_1,V_2)}_{L}, R^{(V_1,V_2)}_{U}) \in \mathcal{R}(C) \times \mathcal{E}(C) \) bounds the capacity from above and below

\[ \mathcal{R}(C_0 \times C)(R^{(V_1,V_2)}_{L}) \subseteq \mathcal{R}(C_0 \times C) \subseteq \mathcal{R}(C_0 \times C(R^{(V_1,V_2)}_{U})). \]

If \( \mathcal{R}(R^{(V_1,V_2)}_{L}) = C(R^{(V_1,V_2)}_{L}) \), then the bounds are identical and tight. Otherwise, combining bounds as

\[ \mathcal{R}(C_0 \times C) \subseteq \bigcup_{R^{(V_1,V_2)}_{L} \in \mathcal{R}(C)} \mathcal{R}(C_0 \times C(R^{(V_1,V_2)}_{L})) \]

\[ \mathcal{R}(C_0 \times C) \subseteq \bigcap_{R^{(V_1,V_2)}_{U} \in \mathcal{E}(C)} \mathcal{R}(C_0 \times C(R^{(V_1,V_2)}_{U})). \]
may improve one or both.

For any constant \( c \) and set \( A \subseteq \mathbb{R}^k \), let
\[
cA = \{(ca_1, \ldots, ca_k) : (a_1, \ldots, a_k) \in A \}.
\]

Lemma 10, also described in [8], bounds the accuracy of capacity bounds derived using the bounding models for a given channel \( C \). The bound is derived by comparing the upper and lower bounds to each other and deriving a constant \( c \) for which \( \mathcal{R}_L(C_o \times C) \subseteq \mathcal{U}(C_o \times C) \subseteq c\mathcal{R}_L(C_o \times C) \) for all \( C_o \). For example, if \( c = 1.05 \), then replacing channel \( C \) by its upper or lower bounding model in any network never changes the capacity by more than 5%. The result applies only when the topologies of upper and lower bounding models are the same and equals the maximal ratio between the capacity of a link in the upper bounding model and the capacity of the same link in the lower bounding model. Set
\[
\delta_L(C) = \{ \frac{R_{(a\rightarrow b)}}{R_{(a\rightarrow b)}} : R_{(a\rightarrow b)} \neq 0 \forall (A, B) \text{ s.t. } |A| > 1 \}
\]
is used to describe all upper bounding models with topologies identical to lower bounding models.

**Lemma 10** Let \( C = (\mathcal{X}^{V_1}, p(y^{V_2}|x^{V_1}), \mathcal{Y}^{V_2}) \). If \( \delta_L(C) \neq 0 \), then
\[
\mathcal{R}_L(C_o \times C) \subseteq \mathcal{U}(C_o \times C) \subseteq \mu(C)\mathcal{R}_L(C_o \times C)
\]
for all \( C_o \), where
\[
\mu(C) = \inf_{(a, b) \in \mathcal{M}(C)} \frac{\max_{(a, b) \in \mathcal{M}(C)} R_{(a\rightarrow b)}^{(a\rightarrow b)}}{R_{(a\rightarrow b)}^{(a\rightarrow b)}}.
\]

**Proof.** For any \( C_0 \) and any \( (\mathcal{R}_L^{(V_1, V_2)}, \mathcal{U}^{(V_1, V_2)}) \) \( \subseteq \mathcal{R}_L(C) \times \delta_L(C) \),
\[
\mathcal{R}(C_0 \times C(\mathcal{R}_L^{(V_1, V_2)})) \subseteq \mathcal{R}_L(N) \subseteq \mathcal{R}(C_0 \times C(\mathcal{R}_L^{(V_1, V_2)})).
\]
Let
\[
c = \frac{\max_{(a, b) \in \mathcal{M}(C)} R_{(a\rightarrow b)}^{(a\rightarrow b)}}{\max_{(a, b) \in \mathcal{M}(C)} R_{(a\rightarrow b)}^{(a\rightarrow b)}}.
\]
Then \( R_{(a\rightarrow b)}^{(a\rightarrow b)} \leq cR_{(a\rightarrow b)}^{(a\rightarrow b)} \) for all \( (a, b) \), and
\[
\mathcal{R}(C_0 \times C(\mathcal{R}_L^{(V_1, V_2)})) \subseteq \mathcal{R}(C_0 \times C(\mathcal{R}_L^{(V_1, V_2)})) \subseteq \mathcal{R}(C_0 \times C(c\mathcal{R}_L^{(V_1, V_2)})).
\]
We next show that
\[
\mathcal{R}(C_0 \times C(c\mathcal{R}_L^{(V_1, V_2)})) \subseteq c\mathcal{R}(C_0 \times C(\mathcal{R}_L^{(V_1, V_2)})).
\]
Let \( \mathcal{N} \) be the \( N \)-fold stacked network for
\[
\mathcal{N} = C_0 \times C(\mathcal{R}_L^{(V_1, V_2)})
\]
and \( \mathcal{N}' \) be the \( N' \)-fold stacked network for
\[
\mathcal{N}' = C_0 \times C(c\mathcal{R}_L^{(V_1, V_2)}).
\]
Set \( N = [c\mathcal{N}'] \). To operate a \((\lambda, \mathcal{R})\)-solution \( \mathcal{S}(\mathcal{N}') \) for \( \mathcal{N}' \) across \( \mathcal{N} \), send the \( \mathcal{N}'C(\mathcal{R}_L^{(V_1, V_2)}) \) bits transmitted across \( \mathcal{N}' \) copies of \( C(c\mathcal{R}_L^{(V_1, V_2)}) \) in \( \mathcal{N}' \) across the \( N \) copies of \( C(\mathcal{R}_L^{(V_1, V_2)}) \) in \( \mathcal{N} \). The resulting code has error probability \( \lambda \) and rate \( N'\mathcal{R}/[c\mathcal{N}'] \), which approaches \( \mathcal{R}/c \) as \( N' \) grows without bound. Thus \( R/c \in \mathcal{R}(C_0 \times C(\mathcal{R}_L^{(V_1, V_2)})), \) which implies \( R \in c\mathcal{R}(C_0 \times C(\mathcal{R}_L^{(V_1, V_2)})) \). Taking the infimum over \( \mathcal{R}(\mathcal{R}_L^{(V_1, V_2)}, \mathcal{R}_L^{(V_1, V_2)}) \) gives the result. \( \blacksquare \)

**Lemma 11** Let \( \mathcal{N} = \prod_{e \in E} C_e \) with
\[
C_e = (\mathcal{X}^{V_1(e)}, p(y^{V_2(e)}|x^{V_1(e)}), \mathcal{Y}^{V_2(e)})
\]
and \( \delta_L(C_e) \neq 0 \) for all \( e \in E \). Then
\[
\mathcal{R}_L(N) \subseteq \mathcal{R}_U(N) \subseteq [\max_{e \in E} \mu(C_e)]\mathcal{R}_L(N).
\]

**Proof.** Consider the bit-pipe models
\[
\prod_{e \in E} C(\mathcal{R}_L^{(V_1(e), V_2(e))}) \quad \text{and} \quad \prod_{e \in E} C(\mathcal{R}_U^{(V_1(e), V_2(e))})
\]
onobtained by replacing each independent channel by its bit-pipe model. The maximal worst-case ratio of edge capacities for this pair of models equals \( \max_{e \in E} \mu(C_e) \). The result thus follows from Lemma 10. \( \blacksquare \)

**VI. Examples**

Figure 4 shows examples of lower and upper bounding models. Their derivations and accuracy bounds appear in Appendix V. For some, equivalent bounding models derived using Lemma 1 are also shown. Most of the bounding models are asymmetrical even when the channel is symmetrical. Symmetrical models can be derived for any symmetrical \( C \) since \( \mathcal{R}(C) \) and \( \delta^o(C) \) are convex; for example if \( C \) is a symmetrical broadcast channel \( C \), then, by the symmetry of \( C \), for any \( R \in \mathcal{R}(C) \) there exists a \( R' \in \mathcal{R}(C) \) that reverses the roles of the two receivers, and \( C(\frac{1}{2}R + \frac{1}{2}R') \) is a symmetrical lower
Fig. 4. Example bounding models: (a) Binary symmetric broadcast channel with noise distribution \(q(z_1, z_2)\), \(E[Z_1] = p_1, E[Z_2] = p_2\). (b) Real Gaussian broadcast channel with \(E[X^2] \leq P, E[Z_i^2] = N_i, E[Z_1Z_2] = \rho \sqrt{N_1N_2}\). (c) Binary symmetric multiple access channel with \(E[Z] = p\). (d) Gaussian multiple access channel with power constraints \((P_1, P_2)\) and noise variance \(N\). (e) Dependent binary symmetric channels with noise distribution \(q(z_1, z_2)\). In each, \(\alpha \in [0, 1]\).
Example 1 Figure 5 (center) shows a diamond network containing a Gaussian broadcast channel followed by a Gaussian multiple access channel. The channel outputs at nodes 2, 3, and 4 are

\[ Y^{(2)} = aX^{(1)} + Z_1 \]
\[ Y^{(3)} = a^{3/2}X^{(1)} + Z_2 \]
\[ Y^{(4)} = a^{3/2}X^{(2)} + a^{1/2}X^{(3)} + Z_3 \]

The power constraint at each node is \( P = 1 \). The variance of each noise random variable \( Z_1, Z_2, \) and \( Z_3 \) is \( N = 1 \). The noise vector \((Z_1, Z_2)\) for the broadcast channel has correlation coefficient \( \rho \in [-1, 1] \).

The first half of the thesis [9] derives bounds on the unicast capacity from node 1 to node 4 in this network when \( \rho = 0 \); Figure 6 shows the resulting inner (o) and outer (+) bounds for 8 values of \( \alpha \) in \([1, 25]\). Bounds for this network with \( \rho = 0 \) are also derived in [10, Theorem 3.2]; those results apply only for \( \alpha \) asymptotically large, thus cannot be shown here. The equivalence tools transform capacity calculation for all \( a \) and \( \rho \) into a simple exercise. Figure 5 (left and right) shows bounding models. Applying those models with \( \alpha = (a^2 - 1)/a^2 \) in the lower bound and \( \alpha = 0 \) in the upper bound yields the solid (lower) and dashed (upper) bounds shown in Figure 6; the difference between these bounds is nowhere greater than \( 1 \) bit per channel use and sometimes very close to zero. The lower bound is achieved with equality using an independent channel code on each channel followed by routing. 

Example 2 Figure 7 shows a network of binary symmetric point-to-point, broadcast, and multiple access channels and its bounding models. The network is similar to a noiseless network from [11] for which cut-ssets are not tight under a pair of unicast demands. The channel outputs are

\[ (Y^{(3,1)}, Y^{(3,2)}) = (X^{(1,2)} + Z_1, X^{(2)} + Z_2) \]
\[ (Y^{(4,1)}, Y^{(4,2)}) = (X^{(1,2)} + Z_3, X^{(3)} + Z_4) \]
\[ (Y^{(5,1)}, Y^{(5,2)}) = (X^{(1,1)} + X^{(4)} + Z_5, X^{(2)} + Z_6) \]
\[ Y^{(6)} = X^{(5)} + Z_7 \]
\[ (Y^{(7,1)}, Y^{(7,2)}) = (X^{(5)} + Z_8, X^{(6)} + Z_9) \]

where

\[ E[Z_i] = \begin{cases} 
  p_1 & \text{if } i \in \{1, 2, 4, 5, 7, 9\} \\
  p_2 & \text{if } i \in \{3, 6, 8\} 
\end{cases} \]

Figure 8(a) shows the set of achievable rates

\[ (R^{(1)} \rightarrow (4)), R^{(2)} \rightarrow (7)) \]

for a pair of unicast connections (1 to 4 and 2 to 7) on the lower (solid line) and the upper (dashed line) bounding networks when \( (p_1, p_2) = (0.02, 0.2) \) and all broadcast channels are (i) physically degraded, (ii) independent, and (iii) \( C_{12} \)-maximizing. The lower bound is identical in all three cases. The upper bound is smallest for physically degraded noise and largest for \( C_{12} \)-maximizing noise. The bounding regions \( \mathcal{R}_1 \) and \( \mathcal{R}_2 \) shown in Figure 8(a) differ additively by at most 0.14 bits per symbol in each dimension and satisfy
Fig. 7. The network of independent binary symmetric channels described in Example 2 (center) and its bounding models (left and right).

Fig. 8. Two-unicast capacity bounds for the network of independent binary symmetric channels described in Example 2. (a) Upper (dashed) and lower (solid) bounds derived using the bounding models. (b) The lower bound from (a) and the cut-set upper bound. Both upper bounds vary with the noise dependence in the broadcast channel receivers. The three curves, in increasing order, arise when all broadcast channels suffer physically degraded, independent, and $C_{12}$-maximizing noise.

Fig. 9. Upper and lower bounding models for a binary erasure broadcast channel with marginal erasure probabilities $p_1 \leq p_2$ at receivers 1 and 2 and probability $p_{12}$ of simultaneous erasures at both receivers.

The set bounds are ever superior to bounds achieved using cut-set models. It is interesting, then, to consider whether cut-set bounds are ever superior to bounds achieved using cut-set models. Examples 2 and 3 show scenarios where bounding models yield capacity outer bounds at least as good as, and sometimes far better than, those achieved by cut-set bounds. It is interesting, then, to consider whether cut-set bounds are ever superior to bounds achieved using cut-set models. By Lemma 6, for each cut $S$ across a channel $C$, there exists an upper bounding model that is tight in cut $S$. As a result, cut-set bounds are never tighter than the intersection of upper bounds found from all bounding models.
Example 3 Wireless erasure networks are networks of broadcast erasure channels without interference. These networks are used to investigate the role of packet loss in wireless network applications. (See, for example, [12]–[15].) To date, available capacity bounds treat only multicast scenarios. Figure 9 shows the lower and upper bounding models for a broadcast erasure channel with loss probabilities \( p_1 \) and \( p_2 \) at its two receivers; here \( 0 \leq p_1 \leq p_2 \) by assumption. The resulting models give

\[
\mu(C) \leq \max_{\alpha} \min \left( \frac{1 - p_2}{\alpha(1 - p_2)}, \frac{p_2 - p_{12}}{(1 - \alpha)(1 - p_1)} \right) = 1 + \frac{p_2 - p_{12}}{1 - p_1},
\]

which is achieved with \( \alpha = \frac{1 - p_1}{1 - p_1 + p_2 - p_{12}} \). Thus \( \mu(C) = 1 + p_2 \) when the noise at the two receivers is independent and \( \mu(C) = 1 + \frac{p_2}{2} \) in the worst case, where \( p_{12} = 0 \). Thus, by Lemma 11, the capacities under arbitrary demands of network coding instances achieved by replacing every channel by its upper (lower) bounding model differ from the capacity of the true network (and each other) by at most 11% when \( p_2 = 0.1, 1\% \) when \( p_2 = 0.01, 0.1\% \) when \( p_2 = 0.001, \) and so on.

VII. Conclusions

The equivalence tools introduced in this two-part series implement the first step of a new strategy for building computational tools that bound the capacities of large networks. Unlike cut-set techniques, which investigate networks in their entirety, the equivalence tools apply a divide-and-conquer strategy, bounding capacities of networks by characterizing the behaviors of their component channels. Since capacities do not compose [1, Example 2], a more nuanced component characterization is required. The characterization proposed here combines capacity and emulation bounds, which give lower and upper bounds, respectively, on a channel’s behavior across all memoryless networks containing that channel. Capacity lower bounds derived using the bounding model strategy describe the optimal performance of a network when source-network coding is separated from independent channel coding on each channel. Capacity upper bounds derived using this strategy employ the maximal rate that can be delivered by each channel, giving network capacity bounds never worse, and sometimes far better, than those obtained by cut-set bounds.

The emulation region, introduced in this work, is a natural complement to the capacity region: the capacity region describes which bit-pipe channels the channel can emulate; the emulation region describes which bit-pipe channels can emulate the channel. As discussed in Section III, the latter question can be tackled using tools from the source coding literature. Simple emulation region bounds for the point-to-point, broadcast, multiple access, and interference channels are provided as examples. Deriving emulation regions for more channels would increase the library of component models for which bounds are available, thereby increasing the family of networks for which lower and upper capacity bounds can be derived.

Many questions remain open for further research. Understanding the capacity impact of a single link removal would enable accuracy bounds when upper and lower bounding models differ. Comparing upper and lower bounds on network capacities can help identify where joint source-network and channel coding may give significant gains. Systematic strategies for recognizing and realizing joint coding opportunities in networks of independent channels remains a challenging open question.

Appendix I

The Impact of Delay on Network Capacity

Lemma 1, also presented in [5], shows that delay has no impact on network capacity. A similar proof is used in [16, Exercise 17.18], published after the submission of this work, to prove that causal delay on the feedback link has no impact on the capacity of a multiple access channel with feedback.

Given a memoryless network \( \mathcal{N} \) and a fixed, non-negative, integer delay vector \( \mathbf{d} \), network \( \mathcal{N}(\mathbf{d}) \) is the modification of network \( \mathcal{N} \) in which output \( \mathbf{Y}^{(v,k)} \) is delayed by \( d(v,k) \) time steps.\(^3\) To make this precise, let \( (Y_t^{(v)}) : v \in \mathcal{V}) \) be the network outputs that result from the action of network \( \mathcal{N} \) on the inputs \( (X_t^{(v)}) : v \in \mathcal{V}) \) transmitted at time \( t \). In network \( \mathcal{N}, Y_{t}^{(v)} \) is received by node \( v \) at time \( t \). In network \( \mathcal{N}(\mathbf{d}) \), the same inputs yield output \( \hat{Y}_t^{(v)} = (Y_t^{(v,1)}, \ldots, Y_t^{(v,k(v))}) \) at node \( v \) at time \( t \), where

\[
\hat{Y}_t^{(v,k)} = \begin{cases} 
Y_t^{(v,k)} & \text{if } t-d(v,k) \geq 1 \\
\text{otherwise} 
\end{cases}
\]

Here \( \hat{Y}_t^{(v,k)} \in \mathcal{Y}^{(v,k)} \) is a constant channel output received until the result of the first transmission arrives.

Proof of Lemma 1. Let \( \mathcal{N} \) be an arbitrary network and \( \mathcal{N}(\mathbf{d}) \) be the network that delays \( \mathcal{Y}^{(v,k)} \) by one time step. The argument that follows proves that \( \mathcal{R}(\mathcal{N}) = \mathcal{R}(\mathcal{N}(\mathbf{d})) \). Since this paper focuses on memoryless networks, it is useful to note that \( \mathcal{N}(\mathbf{d}) \) can be implemented as a causal, memoryless network in which node \( v \) is separated from the network output \( \hat{Y}_t^{(v,k)} \) by \( d(v,k) \) intermediate nodes.
\(\mathcal{R}(N(d_1))\). Since node and channel indices are arbitrary, this implies that no single delay changes the capacity. For any finite \(d\), \(\mathcal{R}(N) = \mathcal{R}(N(d))\) then follows by \(\sum_{j,k} \delta(j,k)\) applications of the single-delay result.

Proving \(\mathcal{R}(N(d_1)) \subseteq \mathcal{R}(N)\) is immediate since any solution \(S(N(d_1))\) can be implemented precisely on \(S(N)\) with the same blocklength, error probability, and rate.

To prove \(\mathcal{R}(N) \subseteq \mathcal{R}(N(d_1))\), note that for any \(R \in \text{int}(\mathcal{R}(N))\), \(\epsilon > 0\), and \(n\) sufficiently large there exists a blocklength-\(n\) \((\epsilon, R)\)-\(S(N)\) solution for \(N\). Let \(X^{(v)}(\cdot)\) and \(\hat{W}^{(u)}_{(v)}(\cdot)\) denote the encoders and decoders for \(S(N)\). We build a solution \(S(N(d_1))\) for \(N(d_1)\) by interleaving two independent applications of solution \(S(N)\) to form a blocklength-\((2n+1)\) solution for \(N(d_1)\) with rate \(\frac{2n+1}{2n} R\) and error probability at most \(2\epsilon\).

Interleaving two applications of the same code ensures that even the delayed channel output arrives before it is needed for each subsequent encoding operation. Adding time step \((2n+1)\) ensures that all channel outputs arrive before decoding commences. Allowing \(n\) to grow without bound gives the desired result. The code is described formally below.

Solution \(S(N(d))\) delivers a message \(W^{(v)}(v) \rightarrow U \in [2^{2nR^{(v)}}]\) for each \((v, U) \in \mathcal{M}\). To transmit these messages using blocklength-\(n\) solution \(S(N)\), each message is broken into two parts

\[
W^{(v)}(v) \rightarrow U = (W^{(v)}(v) \rightarrow U(1), W^{(v)}(v) \rightarrow U(2))
\]

with \(W^{(v)}(v) \rightarrow U(1), W^{(v)}(v) \rightarrow U(2) \in [2^{2nR^{(v)}}]\) corresponding to the first and second halves of its bit stream. We denote the outgoing messages from node \(v\) by \(W^{(v)}(v) \rightarrow s = (W^{(v)}(v) \rightarrow s(1), W^{(v)}(v) \rightarrow s(2))\) and all messages by \(W = (W(1), W(2))\). Solution \(S(N(d_1))\) applies the node encoders from \(S(N)\) to \(W(1)\) in the odd time steps and to \(W(2)\) in the even steps. Formally,

\[
X^{(v)}(2n-1) = X^{(v)}(\hat{Y}^{(v)}_{1:t-1}(\ell), W^{(v)}(v) \rightarrow s(\ell)) \quad \ell \in [n] \times \{1, 2\},
\]

\(Y^{(v)}_{i-1}(\ell) = (Y^{(v)}_1(\ell), \ldots, Y^{(v)}_{i-1}(\ell))\),

and \(Y^{(v)}_i(\ell)\) is the channel output at node \(v\) resulting from the \(i\)th channel use for message \(W(\ell)\); thus \(Y^{(v)} = (Y^{(v)}_{i=1}, \ldots, Y^{(v)}_{i=v(k)})\), and

\[
Y^{(v,k)}_i(\ell) = \begin{cases} 
Y^{(v,k)}_{i=1} + 1, & \text{if } (v, k) = (1, 1) \\
Y^{(v,k)}_{i=1} + 1, & \text{otherwise.}
\end{cases}
\]

The code is causal: each channel input depends only on prior channel outputs. No transmission is required at time \(2n + 1\); the extra time step is included to allow \(Y^{(v)}_{1:n}(2) = Y^{(v)}_{2n+1}\) to reach node 1 before decoding. By the union bound, applying the decoders from \(S(N)\) as

\[
\hat{W}^{(u)}_{(v)}(v) = \hat{W}^{(u)}_{(v)}(v, Y^{(v)}_{1:n}(\ell), W^{(v)}(v) \rightarrow s(\ell))
\]

for all \((u, v) \in \mathcal{M}, v \in V\), and \(\ell \in \{1, 2\}\) reconstructs \(W\) with error probability no greater than \(2\epsilon\).

**APPENDIX II**

**LOWER BOUNDS**

Given a multiterminal channel

\[
\mathcal{C} = (A^{V_1}, p(y^{V_2} | x^{V_1}), Y^{V_2})
\]

the definitions for a solution \(\mathcal{S}(\mathcal{C})\) and the capacity region \(\mathcal{R}(\mathcal{C})\) simplify from [1, Definitions 1 and 2] to Definitions 8 and 9, below. As in [1], \(A^{N} = \{0, 1\}^{NR}\) for any alphabet \(A\). We use notation \(\{0, 1\}^{NR}\) and \([2^{NR}]\) interchangeably.

**Definition 8** For any integer \(N \geq 1\) and rate vector

\[
\mathcal{R}_L^{V_1, V_2} = (R_L^{(a) \rightarrow B}) : a \in V_1, B \subseteq V_2,
\]

a \((2^{NR}_L^{V_1, V_2}, N)\) channel code \((\alpha, \beta)\) for channel \(\mathcal{C}\) comprises a collection of encoding functions \(\alpha = (\alpha^{(v)} : v \in V_1)\) and decoding functions \(\beta = (\beta^{(u) \rightarrow V} : u \in V_1, V \subseteq V_2, v \in V)\), where

\[
\alpha^{(v)} : \bigoplus_{U \subseteq V_2} \mathcal{W}^{(v)}(v) \rightarrow \mathcal{N}^{(v,1)}
\]

and \(\mathcal{W}^{(v)}(v) \rightarrow U = [2^{NR_L^{(v)}}]\). The code’s average error probability is

\[
P_e^{(N)} \equiv \frac{1}{|W|} \sum_{w \in W} \Pr \left( \beta(Y^{V_2}) \neq w \mid X^{V_1} = \alpha(w) \right),
\]

where \(\beta(Y^{V_2}) \neq w\) is notational shorthand for the event that \(\beta^{(u) \rightarrow V} \neq \beta^{(u) \rightarrow V} \neq \beta^{(u) \rightarrow V} \neq \beta^{(u) \rightarrow V}\) for one or more triples \((u, V, v)\) with \(u \in V_1, V \subseteq V_2, v \in V\).

**Definition 9** The capacity region \(\mathcal{R}(\mathcal{C})\) of channel \(\mathcal{C}\) is the closure of all rate vectors \(\mathcal{R}_L^{V_1, V_2}\) such that for any \(\lambda > 0\) and all \(N\) sufficiently large, there exists a \((2^{NR_L^{V_1, V_2}}, N)\) channel code for channel \(\mathcal{C}\) with average error probability \(P_e^{(N)} < \lambda\).
Proof of Lemma 2:

In the argument that follows, transmissions from the internal nodes of \( C(R(V_1, V_2)) \) incur no delay. The network input from node \( \sigma(a) \in V_{0,L} \) at time \( t \) is a function of past and current network outputs \( Y_{1:t}^{(\sigma(a))} = (Y_1^{(\sigma(a))}, \ldots, Y_t^{(\sigma(a))}) \). This choice has no impact on capacity by Lemma 1 and violates no physical constraints since an internal node is a conceptual tool used to describe the model rather than a real physical device.

By our definition of a bounding model, \( C(R(V_1, V_2)) \subseteq C \) implies \( \mathcal{B}(C_0 \times C(R(V_1, V_2))) \subseteq \mathcal{B}(C_0 \times C) \) for all \( C_0 \). Applying this relationship with the trivial network \( C_0 \) that has no inputs and no outputs gives \( \mathcal{B}(C(R(V_1, V_2))) \subseteq \mathcal{B}(C) \), and therefore \( R(V_1, V_2) \in \mathcal{B}(C) \), which is the desired result.

Given any \( C_0 \) and \( R(V_1, V_2) \in \mathcal{B}(C) \), let \( N = C_0 \times C \) and \( N_L = C_0 \times C(R(V_1, V_2)) \). By [1, Lemma 1], it suffices to prove that \( \mathcal{B}(N_L) \subseteq \mathcal{B}(N) \) for all \( N \) sufficiently large. This is shown by fixing any \( \lambda > 0 \) and \( R \in \mathcal{B}(N_L) \) showing that for \( N \) sufficiently large there exists a \( (\lambda, R) \)-\( S(N) \) solution for \( N \)-fold stacked network \( N \). Solution \( S(N) \) runs an \( N \)-fold stacked solution \( S(N_L) \) across \( N \) with the help of a \( (2^N R(V_1, V_2), N) \) channel code for channel \( C \). For reasons to be explained below, we may vary the indexing of the codewords of \( (\alpha, \beta) \) with time; for each \( t \in [n] \), let \( (\alpha_t, \beta_t) \) be channel code \( (\alpha, \beta) \) as indexed for time \( t \).

For each \( v \in V \) and \( t \in [n] \), encoder \( \tilde{X}_t(v) \) in solution \( \mathcal{S}(N_L) \) first channel decodes any outputs from channel \( C \). Let \( Y_t^{(v)} \) and \( \tilde{Y}_t^{(v)} \) be the network output before and after this operation, respectively; then,

\[
\tilde{Y}_t(v) = \begin{cases} 
\sum_{j=0}^{v-1} \tilde{X}_t(v), & \text{if } v \in V \setminus V_2 \\
\sum_{j=0}^{v} \tilde{X}_t(v), & \text{if } v \in V_2.
\end{cases}
\]

Each node \( v \in V \) then applies encoder \( \tilde{X}_t(v) \) to give \( \tilde{X}_t(v) = \tilde{X}_t^{(v)}(\tilde{Y}_t(v)) \). If \( v \in V_L \), it additionally operates internal node encoder \( \dot{X}_t(v) \) and channel encoder \( \alpha_t(v) \), giving

\[
\dot{X}_t(v) = \left\{ \begin{array}{ll}
\tilde{X}_t(v), & \text{if } v \in V \setminus V_1 \\
\alpha_t(v) \tilde{X}_t(v\mid \lambda), & \text{if } v \in V_1.
\end{array} \right.
\]

Node \( v \) transmits \( \dot{X}_t(v) \) at time \( t \).

Let \( \mathbf{	ilde{W}} \) denote the full collection \( \mathbf{	ilde{W}} = (\tilde{W}^{(u,v)} : (u, v) \in M, v \in V) \) of decoders for solution \( S(N_L) \). The corresponding decoder \( \mathbf{\tilde{W}} \) for solution \( S(N) \) applies decoder \( \mathbf{\tilde{W}} \) from \( S(N_L) \) as

\[
\mathbf{\tilde{W}} = \mathbf{\tilde{W}}(\dot{X}_t(v\mid \lambda), \tilde{W}^{(v\mapsto)})
\]

Solution \( S(N) \) can fail in two ways: One or more channel codes \( (\alpha_t, \beta_t) \) can decode incorrectly, or solution \( S(N_L) \) can fail even though all channel codes succeed. Let \( E_t \) denote the event that channel code \( (\alpha_t, \beta_t) \) fails. We bound the failure probability as

\[
\Pr(\mathbf{\tilde{W}} \neq \mathbf{\tilde{W}}) \leq \sum_{t=1}^{n} \Pr(E_t) + \sum_{w \in W} \Pr(\mathbf{\tilde{W}} \neq \mathbf{\tilde{W}} | \mathbf{\tilde{W}} = w) \cap \bigcap_{t=1}^{n} E_t),
\]

where \( \Pr(\mathbf{\tilde{W}} \neq \mathbf{\tilde{W}} | \mathbf{\tilde{W}} = w) \leq \lambda/(n+1) \) by our choice.
of $N$.\footnote{Proving $\Pr(E_2) \leq \lambda/(n+1)$ is slightly subtle. The capacity definition guarantees that the average error probability for $(\alpha, \beta_1)$ can be made arbitrarily small, but the time-$t$ distribution on messages under solution $S(N)$ need not be uniform. Applying a random index assignment to the channel codewords yields $E[\Pr(E_t)] \leq P_c(N)$, which proves the existence of an index assignment for each $t$ with $\Pr(E_t) \leq P_c(N) < \lambda/(n+1)$ as desired.}

The second term follows since $p(w)\cap_{i=1}^n E_i^c \leq p(w)$ for all $w$ and $\Pr(W \neq W|W = w) \cap_{i=1}^n \geq (n+1)$ is the conditional error probability of $S(N)$ given $W = w$; thus the second term is no greater than the error probability of $S(N)$, which is also less than $\lambda/(n+1)$ by our choice of $N$. ■

APPENDIX III

THE TYPICAL SET: DEFINITIONS AND PROPERTIES

The definition of the traditional jointly typical set $A_c^{(N)}$ for a random vector requires every subvector to be typical. The following notation is useful to make this precise. Given $K = |K|$ for some positive integer $K$, let $U^K$ be the random vector $(U_1, \ldots, U_K)$ drawn according to distribution $p(u_1, \ldots, u_K)$ on $\mathcal{U}^K = \prod_{k=1}^K U_k$. We denote $N$ i.i.d. copies of $U^K$ by $U_N^K = (U_1, \ldots, U_N^K)$ and the corresponding alphabet by $\mathcal{U}_N^K = \prod_{k=1}^K U_k$. Given any $A \subseteq K$, $|A|$ sufficiently large. $U_A^A = (U_k : k \in A)$ and $U_A^A = (U_k : k \in A)$ be the subvectors of $U_N^K$ and $U_K^K$ with components in $A$. The corresponding alphabets are $\mathcal{U}_A^A = \prod_{k \in A} U_k$ and $\mathcal{U}_A^A = \prod_{k \in A} U_k$. For any $w^K \in \mathcal{U}_N^K$, we similarly define $w_A^A = (w_k : k \in A)$.

Definition 10 Let $p(u^K)$ be a distribution on $\mathcal{U}_N^K$ with marginals $p(u_A^A)$ on $\mathcal{U}_A^A$ for each $A \subseteq K$. For any $\epsilon > 0$, the traditional jointly typical set for random vector $U_N^K$ is

$$A_c^{(N)} \overset{\text{def}}{=} \left\{ u \in \mathcal{U}_N^K : p(u_A^A) \leq 2^{-N(H(U_A)+\epsilon)} \forall A \subseteq K \right\}.$$ 

When multiple typical sets are in use, we distinguish between them either by context or by adding arguments. For example, $(X, Y) \in A_c^{(N)}$ and $A_c^{(N)}(X, Y)$ refer to the typical set for the joint distribution on $(X, Y)$.

Like [1], this work uses a restricted form of typical set, often simply called the typical set in this work. Restricted typical set $A_c^{(N)}(X, Y)$ is a subset of $A_c^{(N)}$ for which the conditional probability that the next component is jointly typical with all prior components is high; here “high” is probability at least 1/2 for simplicity. For any vector $u_k = (u_1, \ldots, u_k)$, the conditional probability of interest is

$$p(A_c^{(N)}(U_1, \ldots, U_{k+1})|u_1, \ldots, u_k)$$

Lemma 12 proves that the probability of atypicality decays exponentially to zero. The proof is similar to that of [1, Lemma 8].

Lemma 12 If random vector $U_N^K$ is drawn i.i.d. $p(u^K)$, then for any $\epsilon > 0$

$$p(A_c^{(N)}(U_N^K)) < 2^{-N\zeta(\epsilon)}$$

for some function $\zeta(\epsilon) > 0$ and all $N$ sufficiently large.

Proof. First, divide the atypical set as

$$(A_c^{(N)}(U_N^K))^c = (A_c^{(N)}(U_N^K))^c \cup \left( A_c^{(N)}(U_N^K) \cap (A_c^{(N)}(U_N^K))^c \right). \quad (4)$$

To bound the probability of the first event, divide the traditional atypical set as

$$(A_c^{(N)}(U_N^K))^c = \bigcup_{A \subseteq K} \left\{ u_K^K : \left| -\frac{1}{N} \log p(u_A^A) - H(U_A^A) \right| > \epsilon \right\}.$$ 

Chernoff’s bound (e.g., [17, pp.482-484]), shows that for any i.i.d. random variables $F(1), \ldots, F(N)$,

$$\Pr \left( \sum_{\ell=1}^N F(\ell) > E[F] + \epsilon \right) \leq e^{-N b(\epsilon)},$$

where $b(\epsilon) = -\min_{s>0} [\ln E[e^{sF}] - s(E[F] + \epsilon)]$. Note that $b(\epsilon) \geq 0$ for all $\epsilon > 0$, with equality if and only if $\epsilon = 0$; further, $b(\epsilon)$ grows without bound as $\epsilon > 0$ increases. Given any $A \subseteq K$ and any $\epsilon > 0$, applying Chernoff’s bound first with $F(\ell) = -\log p(U_K^K(\ell))$ and then with $F(\ell) = \log p(U_N^K(\ell))$ and then combining with the union bound gives

$$p \left( \left| -\frac{1}{N} \log p(U_A^A) - H(U_A^A) \right| > \epsilon \right) \leq 2^{-N b_A(\epsilon)}$$
for some \( b_A(\epsilon) > 0 \) independent of \( N \) and all \( N \) sufficiently large. Applying the union bound again gives
\[
p((A^N_ε(U^K))^c) \leq \sum_{A \subseteq K} 2^{-N b_A(\epsilon)} \leq 2^{-N \zeta_K(\epsilon)}
\]
for \( \zeta_K(\epsilon) = \frac{1}{p} \min_{A \subseteq K} b_A(\epsilon) > 0 \) independent of \( N \) and all \( N \) sufficiently large. The given derivation applies not only for \( A^N(U^K) \) but also for \( A^N(U^B) \) for any subset \( B \) of \( K \), giving
\[
p((A^N_ε(U^B))^c) \leq \sum_{A \subseteq B} 2^{-N b_A(\epsilon)} \leq 2^{-N \zeta_K(\epsilon)}.
\]
Next, divide the second event in (4) as
\[
A^N_ε(U^K) \cap (A^N_ε(U^K))^c = \bigcup_{k=1}^{K-1} \left\{ u^K \in A^N_ε \colon p(A^N_ε(U^{k+1})|u^K) < \frac{1}{2} \right\},
\]
giving
\[
p\left( (A^N_ε(U^K) \cap (A^N_ε(U^K))^c) \right) \leq \sum_{k=1}^{K-1} \left\{ \left\{ u^K \in A^N_ε \colon p(A^N_ε(U^{k+1})|u^K) < \frac{1}{2} \right\} \right\}
\]
\[
\leq \sum_{k=1}^{K-1} p\left( \left\{ u^K \in A^N_ε \colon p(A^N_ε(U^{k+1})|u^K) < \frac{1}{2} \right\} \right)
\]
\[
= \sum_{k=1}^{K-1} \sum_{u^K \in A^N_ε(p(A^N(U^{k+1})|u^K) < \frac{1}{2})} p(u^K).
\]
Recall that \( p((A^N_ε(U^B))^c) \leq 2^{-N \zeta_K(\epsilon)} \) for all \( B \subseteq K \) and \( N \) sufficiently large. Therefore, for each \( k \) and \( N \) large enough,
\[
2^{-N \zeta_K(\epsilon)} \geq p\left( (A^N_ε(U^{k+1}))^c \right) \geq \sum_{u^K} p(A^N_ε(U^{k+1})|u^K) p(u^K) \geq \sum_{u^K \in A^N_ε} \left( 1 - \frac{1}{2} p(u^K) \right),
\]
where (a) follows since \( p(A^N_ε(U^{k+1})|u^K) < \frac{1}{2} \) implies \( p((A^N_ε(U^{k+1}))^c|u^K) \geq \frac{1}{2} \). Therefore
\[
\sum_{u^K \in A^N_ε} p(u^K) \leq 2 \cdot 2^{-N \zeta_K(\epsilon)},
\]
giving
\[
p\left( (A^N_ε(U^K) \cap (A^N_ε(U^K))^c) \right) \leq \sum_{k=1}^{K-1} 2^{-N \zeta_K(\epsilon) - \frac{1}{2}} = (K - 1)2^{-N \zeta_K(\epsilon) - \frac{1}{2}}.
\]
Combining the bound on \( p((A^N_ε(U^K))^c) \) with the bound on \( p((A^N_ε(U^K) \cap (A^N_ε(U^K))^c) \) gives
\[
p((A^N_ε(U^K))^c) \leq K 2^{-N \zeta_K(\epsilon) - \frac{1}{2}}.
\]
Since \( |K| = K \) is finite, we have the desired result. ■

**APPENDIX IV**

**EMULATION REGION DERIVATIONS**

Notation and results from Appendix III are here used to derive emulation region bounds. We begin by simplifying the notion of success used in the definition of emulation error from
\[
\mathcal{S}_N(\epsilon,c) = \left\{ (x^{V_1},y^{V_2}) \in \hat{A}_N^c \colon p(x^{V_1}) \hat{p}_N(y^{V_2}|x^{V_1}) \in 2^{-N(H(X^{V_1},Y^{V_2})+\epsilon c)} \right\},
\]
to
\[
\left\{ (x^{V_1},y^{V_2}) \in \hat{A}_N^c \colon p(x^{V_1}) \hat{p}_N(y^{V_2}|x^{V_1}) \in 2^{-N(H(X^{V_1},Y^{V_2})-\epsilon c)} \right\}.
\]
While the original definition is intuitively pleasing due to its similarity with the definition of typicality, the second definition is simpler, sufficient (the lower bound is never used in our derivations), and equivalent, as shown next.

**Lemma 13** Fix channel \( C = (X^{V_1},p(y^{V_2}|x^{V_1}),Y^{V_2}) \), input distribution \( p \in \mathcal{P}(C) \), emulation distribution \( \hat{p}_N(y^{V_2}|x^{V_1}) \), and constant \( c > 1 \). For any \( \epsilon > 0 \), let
\[
Q_N(\epsilon,c) = \sum_{(x^{V_1},y^{V_2}) \in \mathcal{S}_N(\epsilon,c)} p(x^{V_1}) \hat{p}_N(y^{V_2}|x^{V_1})
\]
\[
Q_N'(\epsilon,c) = \sum_{(x^{V_1},y^{V_2}) \in \mathcal{S}_N'(\epsilon,c)} p(x^{V_1}) \hat{p}_N(y^{V_2}|x^{V_1}).
\]
There exists a positive function \( \eta(\epsilon) \) with \( Q_N(\epsilon,c) < 2^{-\eta(\epsilon)N} \) for all \( N \) sufficiently large if and only if there exists a positive function \( \eta'(\epsilon) \) with \( Q_N'(\epsilon,c) < 2^{-\eta'(\epsilon)N} \) for all \( N \) sufficiently large.

**Proof.** Since
\[
(\mathcal{S}_N'(\epsilon,c))^c \subseteq (\mathcal{S}_N(\epsilon,c))^c,
\]

$Q^{(N)}(\epsilon, c) < 2^{-n(c)N}$ implies $Q^{(N)}(\epsilon, c) < 2^{-n(c)N}$.

For the other direction, note that

$$Q^{(N)}(\epsilon, c) = \min_{\beta_N} \{ I(X; N) : \beta_N = \beta_N(\epsilon, c) \}.$$

Since $(x_1, y_2) \in \mathcal{S}^{(N)}(\epsilon, c)$ implies

$$p(x_1)p(y_2|x_1) \leq 2^{-N(H(X_1; Y_{2|X_1}) + \epsilon)},$$

it follows that

$$\sum_{(x_1, y_2) \in \mathcal{S}^{(N)}(\epsilon, c)} p(x_1)p(y_2|x_1) \leq 2^{-N(H(X_1; Y_{2|X_1}) + \epsilon)}.$$

Since $c > 1$ we have the desired result.

Random code designs used to derive emulation regions are similar to those used to derive the rate-distortion theorem: Codewords are drawn at random, and inputs are mapped to jointly typical reproductions where available. In some cases, multi-stage designs similar to multi-stage source code designs (e.g., [18]) are employed: First-stage codes are designed as described above; for each first-stage codeword, a collection of second-stage codewords is drawn according to the conditional distribution on the second reproduction given the first; inputs are mapped to a first-stage codeword and then to a second-stage codeword in the codebook for the first-stage codeword; where possible, the resulting triple is jointly typical; later stages continue similarly. Figure 10 describes the core argument. The algorithm is presented as a second-stage code design, but it works for all stages: First-stage codes are handled by removing the conditioning on the first stage; later stages are handled by setting the first-stage reproduction random variable to a vector representing all prior reproductions.

Lemmas 14 and 15 bound the resulting emulation distribution and its probability of atypicality. The proofs are similar to those of [1, Lemmas 11 and 12].

**Lemma 14** Under the core random emulation code design algorithm, if $(u_0, u_1, u_2) \in A_{\epsilon}^{(N)}$ and $N$ is sufficiently large, then

$$\hat{p}(u_2 | u_0, u_1) \leq 2^{-\beta_N(u_0, u_1)}.$$

**Proof.** Fix $(u_0, u_1, u_2) \in A_{\epsilon}^{(N)}$. Define

$$q(u_0, u_1) = \sum_{u_2 : (u_0, u_1, u_2) \in A_{\epsilon}^{(N)}} p(u_2 | u_0, u_1).$$

The following bounds are useful for simplifying (5). For any $(u_0', u_1', u_2') \in A_{\epsilon}^{(N)}$, $A_{\epsilon}^{(N)} \subseteq A_{\epsilon}^{(N)}$ implies

$$p(u_2 | u_0, u_1) \leq 2^{-N(H(U_2 | U_1) + \epsilon)},$$

$$p(u_0 | u_1) \leq 2^{-N(H(U_0 | U_1, U_2) + \epsilon)}.$$
$\hat{A}_c^{(N)}$ implies $p(\hat{A}_c^{(N)}(U_0, U_1, U_2)|u_0, u_1) \geq 1/2$. Combining this bound with (7) gives

$$\begin{align*}
\frac{1}{2} & \leq \sum_{u_0', u_1', u_2' \in \hat{A}_c^{(N)}} p(u_0'|u_0, u_1) \\
& = \sum_{u_0', u_1', u_2' \in \hat{A}_c^{(N)}} p(u_0'|u_0, u_1) \\
& \leq \left\{ \left\{ u_0', u_1', u_2' \in \hat{A}_c^{(N)} \right\} \right\} \\
& \geq 2^{-N(H(U_2|U_0, U_1)-2\epsilon - 1/N)}.
\end{align*}$$

Combining this bound with (6), gives

$$\begin{align*}
q(u_0, u_1) & = \sum_{u_0', u_1', u_2' \in \hat{A}_c^{(N)}} p(u_0'|u_0, u_1) \\
& \geq 2^{-N(H(U_2|U_0, U_1)-2\epsilon - 1/N)} - 2^{-N(H(U_2|U_0, U_1)+2\epsilon)} \\
& = 2^{-N(I(U_0;U_2|U_1)+4\epsilon+1/N)}.
\end{align*}$$

Returning to (5), for any $(u_0, u_1, u_2) \in \hat{A}_c^{(N)}$,

$$\begin{align*}
\hat{p}_N(u_2|u_0, u_1) & \leq \frac{p(u_2|u_0, u_1)}{q(u_0, u_1) p(u_2|u_0, u_1)} \\
& \leq \frac{q(u_0, u_1)}{2^{-N(I(U_0;U_2|U_1)-4\epsilon)}} \\
& \leq \frac{2^{-N(I(U_0;U_2|U_1)+4\epsilon+1/N)}}{2^{-N(I(U_0;U_2|U_1)-4\epsilon)}} + e^{-2^{N(I(U_0;U_2|U_1)-4\epsilon)}},
\end{align*}$$

where (a) follows from (6) and (7), (b) follows from (8), and (c) applies for $N$ sufficiently large.}

**Lemma 15** Under the core emulation code design algorithm,

$$\hat{p}_N((\hat{A}_c^{(N)}(U_0, U_1, U_2))|u_0, u_1) \leq e^{-2^{N(R(I(U_0;U_2|U_1), I(U_0;U_1))-4\epsilon)}} + p((\hat{A}_c^{(N)}(U_0, U_1, U_2))|u_0, u_1).$$

**Proof.** The bound is trivially satisfied when $(u_0, u_1) \not\in \hat{A}_c^{(N)}$ and also when $(u_0, u_1) \in \hat{A}_c^{(N)}$ but $p((\hat{A}_c^{(N)}(U_0, U_1, U_2))|u_0, u_1) > 1/2$

since in both of these cases

$$\hat{p}_N((\hat{A}_c^{(N)}(U_0, U_1, U_2))|u_0, u_1)$$

$$\hat{p}_N((\hat{A}_c^{(N)}(U_0, U_1, U_2))|u_0, u_1) = p((\hat{A}_c^{(N)}(U_0, U_1, U_2))|u_0, u_1) = 1.$$
Fix $p(x) \in P(C)$, and let $p(x, y_1, y_2) = p(x)p(y_1, y_2|x)$. Fix $\epsilon > 0$ and
\[
R_0 > C_2 = \max_{p(x)} I(X; Y_2)
\]
\[
R_1 > C_{12} - C_2 = \max_{p(x)} I(X; Y_1, Y_2) - \max_{p(x)} I(X; Y_2)
\]
\[
\geq \max_{p(x)} I(X; Y_1|Y_2).
\]

The emulation code $(\alpha_N, \beta_N)$ for broadcast channel $C$ has encoders
\[
\alpha_N = \left( \alpha_N^{(A\to B)} : A \subseteq V_1, B \subseteq V_2 \right)
\]
\[
= \left( \alpha_N^{(1)\to(j_1, j_2)}, \alpha_N^{(2)\to(j_1, j_2)} \right)
\]
\[
= \left( \alpha_N^{(0)}, \alpha_N^{(1)}, \alpha_N^{(2)} \right).
\]

The abbreviated notation is used for simplicity. Since $R_2 = 0$, $\alpha_N^{(2)}$ is the trivial encoder that maps all inputs $x^N$ to the same rate-0 description. The remaining encoders and decoders take the form
\[
\alpha_N^{(0)} : \mathcal{X} \to \mathcal{W}_0, \quad \alpha_N^{(1)} : \mathcal{X} \to \mathcal{W}_1, \quad \alpha_N^{(2)} : \mathcal{X} \to \mathcal{W}_2,
\]
where $\mathcal{W}_0 = [2^{N R_0}] \cup \{0\}$ and $\mathcal{W}_1 = [2^{N R_1}] \cup \{0\}$.

For the random decoder design, draw $\beta_N^{(2)}(w_0), w_0 \in [2^{N R_0}]$, i.i.d. $p(y_2) = \prod_{l=1}^{N} p(y_2(l))$. For each $w_0 \in [2^{N R_0}]$, let $\hat{y}_0 = \beta_N^{(2)}(w_0)$ and draw $\beta_N^{(1)}(w_0, y_1), y_1 \in [2^{N R_1}]$, i.i.d. $p(y_1|y_2) = \prod_{l=1}^{N} p(y_1(l)|y_2(l))$. Let $\epsilon'$ be a function, to be specified later, of $\epsilon$ and $R_0$ that satisfies $0 < \epsilon' < \epsilon$. Fix any $\hat{y}_2 \in \mathcal{Y}_2 \setminus \hat{A}^{(N)}(Y_2)$ and $\hat{y}_1 \in \mathcal{Y}_1 \setminus \hat{A}^{(N)}(Y_1)$; set $\beta_N^{(2)}(0) = \hat{y}_2$ and $\beta_N^{(1)}(w_0, 0) = \hat{y}_1$ for all $w_0$. For the random encoder design, choose $\alpha_N^{(0)}(\hat{y}_2)$ uniformly at random from $\{w_0 \in \mathcal{W}_0 : (\hat{y}_2, \beta_N^{(2)}(w_0)) \in \hat{A}^{(N)}\}$; if the set is empty, set $\alpha_N^{(0)}(\hat{y}_2) = 0$. Let $w_0 = \alpha_N^{(0)}(\hat{y}_2)$; choose $\alpha_N^{(1)}(\hat{y}_2)$ uniformly at random from
\[
\{w_1 \in \mathcal{W}_1 : (\hat{y}_1, \beta_N^{(1)}(w_0, w_1), \beta_N^{(2)}(w_0)) \in \hat{A}^{(N)}\},
\]
setting $\alpha_N^{(1)}(\hat{y}_2) = 0$ if the set is empty.

The code has two stages, with emulation distributions
\[
\tilde{p}_N(y_2|x) = \Pr(\beta_N^{(2)}(\alpha_N^{(0)}(x)) = y_2)
\]
\[
\tilde{p}_N(y_2|y_2, x) = \Pr(\beta_N^{(1)}(\alpha_N^{(0)}(x), \alpha_N^{(1)}(x)) = y_1 | \beta_N^{(2)}(\alpha_N^{(0)}(x)) = y_2)
\]
resulting from its first and second stages, respectively. The combined emulation distribution is
\[
\tilde{p}_N(y_1, y_2|\tilde{x}) = \tilde{p}_N(y_2|\tilde{x})\tilde{p}_N(y_1|y_2, \tilde{x}).
\]

By Lemma 13, to show $R \in \hat{C}^{\alpha}(C)$, it suffices to show that there exists a constant $c > 1$ and positive function $\eta(\epsilon)$ for which the emulation distribution satisfies
\[
Q_N^{(\epsilon)}(\epsilon, c) = \sum_{(x, y_1, y_2) \notin \mathcal{A}^{(N)}(X, Y_1, Y_2, \epsilon)} p(x)\tilde{p}_N(y_1, y_2|\tilde{x}) < 2^{-N \eta(\epsilon)},
\]
where
\[
\mathcal{A}^{(N)}(\epsilon, c) = \left\{ (x, y_1, y_2) \in \hat{A}^{(N)}(X, Y_1, Y_2) : p(x)\tilde{p}_N(y_1, y_2|\tilde{x}) \leq 2^{-N(H(X, Y_1, Y_2) - c\epsilon)} \right\}
\]

We begin by bounding $\tilde{p}_N(y_1, y_2|\tilde{x})$. For all $N$ sufficiently large, applying Lemma 14 to the first- and second-stage designs gives
\[
\tilde{p}_N(y_1, y_2|\tilde{x}) \leq 2^{N(9\epsilon')} p(y_2|x)
\]
for all $(x, y_2) \in \hat{A}^{(N)}$ and
\[
\tilde{p}_N(y_1, y_2|\tilde{x}) \leq 2^{N(9\epsilon')} p(y_2|x)
\]
for all $(x, y_2) \in \hat{A}^{(N)}$ with $(x, y_2) \notin \hat{A}^{(N)}$. Therefore, for $(x, y_2) \in \hat{A}^{(N)}$ and $(x, y_2) \notin \hat{A}^{(N)}$,
\[
\tilde{p}_N(y_1, y_2|\tilde{x}) p(x) \leq 2^{N(9\epsilon' + c)} p(x, y_2) \leq 2^{-N(H(X, Y_1, Y_2) - 9\epsilon' - 10c)} \leq 2^{-N(H(X, Y_1, Y_2) - 19c)}.
\]

Fix $c = 19$. Then the emulation succeeds if $(x, y_2) \in \hat{A}^{(N)}$ and $(x, y_2) \notin \hat{A}^{(N)},$ giving
\[
Q_N^{(\epsilon)}(\epsilon, c) = \tilde{p}_N((\hat{A}^{(N)}(X, Y_2) \cap \hat{A}^{(N)}(X, Y_1, Y_2))) + \tilde{p}_N((\hat{A}^{(N)}(X, Y_2) \cap (X, Y_1, Y_2) \notin \hat{A}^{(N)})) + \sum_{\tilde{x}} \tilde{p}_N((\hat{A}^{(N)}(X, Y_2))^{(\epsilon)}(\tilde{x}) p(x)
\]

21
for all

\begin{equation}
E(E_{\min}) \leq \{\text{Lemma 15, (b) Lemma 14, and (c) Lemma 12. Since } R_0 > I(X;Y_2) \text{ and } R_1 > I(X;Y_1|Y_2), \text{ } Q^{(N)}(\epsilon, \epsilon') \text{ decays exponentially to zero for all } N \text{ sufficiently large that } \epsilon < (R_1 - I(X;Y_1|Y_2))/4 \text{ and } \epsilon' > 0 \text{ is chosen to satisfy } \epsilon' \leq \min\{\epsilon, (R_0 - I(X;Y_2))/4, \epsilon/9\}\}. \end{equation}

**IV-B. Multiple Access Channel (Proof of Theorem 8)**

To simplify notation, use \(R = (R_0, R_1, R_2)\) in place of \(R = (R^{(i_1,i_2)\rightarrow(j_1)}, R^{(i_1)\rightarrow(j_1)}, R^{(i_2)\rightarrow(j_1)})\)

and \(C = (X_1 \times X_2, p(y|x_1,x_2), Y)\) in place of

\[
C = (X^{(i_1,i_2)} \times X^{(i_1)}, p(y|j_1,|j_1), X^{(i_2,i_1)}, Y^{(j_1)}).
\]

We show \(\delta_1 \subseteq \delta^0(\mathcal{C})\). Reversing the roles of the encoders shows \(\delta_2 \subseteq \delta^0(\mathcal{C})\). The result then follows from Lemma 4 and Theorem 5.

Given \(p(u,x_1,x_2) = p(u|x_1)p(x_1,x_2)\), fix

\[
p(u,x_1,x_2,y) = p(u,x_1,x_2)p(y|x_1,x_2)
\]

and output

\[
\beta_N(x_1,x_2) = p(u|x_1)p(x_1,x_2)p(y|x_1,x_2).
\]

Fix \(\epsilon > 0\), \(R_0 > I(X_1,X_2;Y|U), R_1 > I(X_1;U), \text{ and } R_2 = 0\). The following argument describes a random code design for an emulation code \((\alpha_N, \beta_N)\) with rate vector approaching \((R_0, R_1, R_2)\) and then uses this design to show \((R_0, R_1, R_2) \in \delta^0(\mathcal{C})\).

The emulation code \((\alpha_N, \beta_N)\) for multiple access channel \(\mathcal{C}\) has encoders

\[
\alpha_N = \left(\alpha_N^{(A \rightarrow B)} : A \subseteq V_1, B \subseteq V_2\right)
\]

\[
= \left(\alpha_N^{(i_1,i_2)\rightarrow(j_1)}, \alpha_N^{(i_1)\rightarrow(j_1)}, \alpha_N^{(i_2)\rightarrow(j_1)}\right)
\]

and decoder \(\beta_N = \beta^{(j)}_N\). Since \(R_2 = 0\), \(\alpha_N^{(2)}\) is the trivial encoder that maps all \(x_2\) to the same rate-0 description.

The remaining mappings are summarized as

\[
\alpha_N^{(1)} : X_1 \rightarrow W_1, \quad \beta_N : W_0 \times W_1 \rightarrow Y
\]

\[
\alpha_N^{(0)} : X_1 \times X_2 \rightarrow W_0, \quad \gamma_N : W_1 \rightarrow U.
\]

where \(W_0 = [2^{NR_0}] \cup \{0\}\) and \(W_1 = [2^{NR_1}] \cup \{0\}\).

Encoders \(\alpha_N^{(0)}\) and \(\alpha_N^{(1)}\) operate at nodes \(\phi^{(i_1,i_2)}\) and \(\sigma^{(i_1)}\) using input \((X_1, X_2)\) losslessly received from nodes \(i_1\) and \(i_2\) and input \(X_1\) losslessly received from node \(i_1\). Node \(j\) operates the decoder.

Mapping \(\gamma_N\) is used by both encoders.

For the random codebook designs, draw \(\gamma_N(w_1), w_1 \in [2^{NR_1}], \text{ i.i.d. from } p(u)\). For each \(w_1 \in [2^{NR_1}]\), set \(\tilde{w} = \gamma_N(w_1)\) and draw \(\beta_N(w_0, w_1), w_0 \in [2^{NR_0}], \text{ i.i.d. from } p(y|u) = \prod_{i=1}^N p(y|u(i))\). Let \(\epsilon'\) be a function, to be described later, of \(\epsilon\) and \(\mathcal{R}\) for which \(0 < \epsilon' < \epsilon\). Fix \(\tilde{u} \in U \setminus A^{(N)}(U)\) and \(\tilde{y} \in Y \setminus A^{(N)}(Y)\); set \(\gamma_N(0) = \tilde{w}\) and \(\beta_N(0, w_1) = \beta_N(w_0, 0) = \tilde{u}\) for all \(w_0, w_1\).

For the random encoder design, choose \(\alpha_N^{(1)}(x_1)\) uniformly at random from

\[
\{w_1 : (\gamma_N(w_1), x_1) \in A^{(N)}_u\};
\]

if the set is empty, set \(\alpha_N^{(1)}(x_1)\) to 0. For each \((x_1, x_2)\), let \(w_1 = \alpha_N^{(1)}(x_1)\) and \(u = \gamma_N(w_1)\), and choose \(\alpha_N^{(0)}(x_1, x_2)\) uniformly at random from

\[
\{w_0 : (x_1, x_2, u, \beta_N(w_0, w_1)) \in A^{(N)}(X_1, X_2, U, Y)\};
\]

if the set is empty, set \(\alpha_N^{(0)}(x_1, x_2)\) to 0. Note that employing index 0 to represent atypical events ensures, at an asymptotically small cost in rate, that the input \((x_1, x_2)\) and output \(\beta_N(\alpha_N(x_1, x_2))\) are jointly typical if and only if all encoders succeed.

For \(N\) sufficiently large, Lemma 14 gives

\[
\beta_N(u|x_1) \leq 2^{N\epsilon'c} p(u|x_1)
\]
for all $(u, x_1) \in \hat{\mathcal{A}}'_c(N)$ and

$$\hat{p}_N(y|u, x_1, x_2) \leq 2^{N\delta_1} p(y|u, x_1, x_2),$$

for all $(u, x_1, x_2, y) \in \hat{\mathcal{A}}_c(N)$ with $(u, x_1) \in \hat{\mathcal{A}}'_c(N)$. The combined emulation distribution is

$$\hat{p}_N(y|x_1, x_2) = \sum_u \hat{p}_N(u|x_1)\hat{p}_N(y|u, x_1, x_2).$$

For any $(x_1, x_2, y) \in \hat{\mathcal{A}}_c(N)$, the summation can be restricted to $u$ with $(u, x_1) \in \hat{\mathcal{A}}'_c(N)$ and $(u, x_1, x_2, y) \in \hat{\mathcal{A}}_c(N)$ since the code never yields a jointly typical reproduction $y$ unless both encoders succeed. Therefore,

$$\hat{p}_N(y|x_1, x_2) = \sum_{w(x_1) \in \hat{\mathcal{A}}'_c(N)} \hat{p}_N(u|x_1)\hat{p}_N(y|u, x_1, x_2) \leq 2^{N(\epsilon + \epsilon')} \sum_{w(x_1) \in \hat{\mathcal{A}}'_c(N)} p(u|x_1)p(y|u, x_1, x_2) = 2^{N(\epsilon + \epsilon')} \sum_{w(x_1) \in \hat{\mathcal{A}}'_c(N)} p(y|u, x_1, x_2) \leq 2^{N(\epsilon + \epsilon')} \sum_{w(x_1) \in \hat{\mathcal{A}}'_c(N)} p(y|x_1, x_2).$$

Therefore, for any $(x_1, x_2, y) \in \hat{\mathcal{A}}_c(N)$,

$$\hat{p}_N(y|x_1, x_2)p(x_1, x_2) \leq 2^{N(\epsilon + \epsilon')} 2^{-N(H(X_1, X_2, Y) - \epsilon)},$$

Set $c = 19$. Then, for $N$ sufficiently large,

\[
Q^{(N)}(\epsilon, c) = \hat{p}_N((U, X_1) \notin \hat{\mathcal{A}}'_c(N)) + \hat{p}_N((U, X_1) \in \hat{\mathcal{A}}'_c(N) \land (U, X_1, X_2, Y) \notin \hat{\mathcal{A}}_c(N)) \leq \sum_{x_1} \hat{p}_N((\hat{\mathcal{A}}'_c(N)(U, X_1))^c|x_1)p(x_1) + \sum_{x_1, x_2} \hat{p}_N((\hat{\mathcal{A}}'_c(N)(U, X_1, X_2, Y))^c|u, x_1, x_2) \leq \delta_1 + \sum_{x_1} p((\hat{\mathcal{A}}'_c(N)(U, X_1))^c|x_1)p(x_1) + \delta_2 + 2^{N(\epsilon + \epsilon')} \sum_{x_1} p((\hat{\mathcal{A}}'_c(N)(U, X_1, X_2, Y))^c|u, x_1, x_2) \leq \delta_1 + p((\hat{\mathcal{A}}'_c(N)(U, X_1))^c) + \delta_2 + 2^{N(\epsilon + \epsilon')} \sum_{x_1} p((\hat{\mathcal{A}}'_c(N)(U, X_1, X_2, Y))^c|u, x_1, x_2),
\]

(b) $\delta_1 + p((\hat{\mathcal{A}}'_c(N)(U, X_1))^c) + \delta_2 + 2^{N(\epsilon + \epsilon')} \sum_{x_1} p((\hat{\mathcal{A}}'_c(N)(U, X_1, X_2, Y))^c|u, x_1, x_2)$

and the marked inequalities follow from (a) Lemma 15, (b) Lemma 14, (c) the definition of $p(u, x_1, x_2, y)$, and (d) Lemma 12. For all $c < (R_2 - I(X_1, X_2; Y|U))/4$, it gives the desired exponential error decay.

**IV-C. Interference Channel (Proof of Theorem 9)**

Since the proof is similar to earlier examples, the following sketch focuses on the random code design. For simplicity, use $(R_{00}, R_{01}, R_{10}, R_{11})$ to denote

\[
(P^{(i_1, i_2)}_{i_1, i_2} \rightarrow (j_1, j_2)), P^{(i_1, i_2)}_{i_1, i_2} \rightarrow (j_1, j_2), R^{(i_1)}_{i_1} \rightarrow (j_1, j_2)), R^{(i_1)}_{i_1} \rightarrow (j_1, j_2))
\]

by $\mathcal{C} = (X_1 \times X_2, p(y|x_1, x_2), Y_1 \times Y_2)$. Fix

$$p(u_1, u_2, x_1, x_2, y_1, y_2) = p(u_2|x_1)p(u_1|x_1, u_2)p(x_1, x_2)p(y_1, y_2|x_1, x_2),$$

and let $\mathcal{W}_i = [2^{N R_i}] \cup \{0\}$. The random codebook designs proceed as follows. Draw

\[
\gamma^{(2)}_N(u_{10}), \quad u_{10} \in [2^{N R_{10}}],
\]

i.i.d. $\prod_{i=1}^N p(u_2(\ell))$. For each $u_2 = \gamma^{(2)}(u_{10})$, draw

\[
\gamma^{(1)}_N(u_{10}, w_{11}), \quad w_{11} \in [2^{N R_{11}}],
\]

i.i.d. $\prod_{i=1}^N p(u_1(\ell)|u_2(\ell))$ and draw

\[
\beta^{(2)}_N(w_{00}, u_{10}), \quad w_{00} \in [2^{N R_{00}}],
\]

The other rates, denoted by $(R_{02}, R_{20}, R_{12}, R_{21}, R_{22})$, are set to zero. Represent

\[
\mathcal{C} = (X^{(i_1, i_2)} \times X^{(i_2, i_1)}, p(y^{(j_1, j_2)}|x^{(i_1, i_2)}, x^{(i_2, i_1)}), Y^{(j_1, j_2)} \times Y^{(j_2, j_1)})
\]

by $\mathcal{C} = (X_1 \times X_2, p(y|x_1, x_2), Y_1 \times Y_2)$. Fix

$$p(u_1, u_2, x_1, x_2, y_1, y_2) = p(u_2|x_1)p(u_1|x_1, u_2)p(x_1, x_2)p(y_1, y_2|x_1, x_2),$$

and let $\mathcal{W}_i = [2^{N R_i}] \cup \{0\}$. The random codebook designs proceed as follows. Draw

\[
\gamma^{(2)}_N(u_{10}), \quad u_{10} \in [2^{N R_{10}}],
\]

i.i.d. $\prod_{i=1}^N p(u_2(\ell))$. For each $u_2 = \gamma^{(2)}(u_{10})$, draw

\[
\gamma^{(1)}_N(u_{10}, w_{11}), \quad w_{11} \in [2^{N R_{11}}],
\]

i.i.d. $\prod_{i=1}^N p(u_1(\ell)|u_2(\ell))$ and draw

\[
\beta^{(2)}_N(w_{00}, u_{10}), \quad w_{00} \in [2^{N R_{00}}],
\]
i.i.d. \( \prod_{\ell=1}^{N} p(y_\ell(\ell)|u_\ell(\ell)) \). For each 
\[
(z_1, z_2) = (\gamma_N^{(1)}(w_{10}, w_{11}), \gamma_N^{(2)}(w_{10}), \beta_N^{(2)}(w_{00}, w_{10})),
\]
draw 
\[
\beta_N^{(1)}(w_{00}, w_{01}, w_{10}, w_{11}), \quad w_{01} \in [2^{NR_0}],
\]
i.i.d. \( \prod_{\ell=1}^{N} p(y_\ell(\ell)|u_\ell(\ell), y_\ell(\ell), y_{-\ell}(\ell)) \). The encoders are designed as follows.

- Map \( X_1 \) to a jointly typical 
\[
U_2 = \gamma_N^{(2)}(w_{10}), \quad w_{10} \in [2^{NR}] .
\]
- Given \( w_{10} \) and \( U_2 \) from above, map \( (X_1, U_2) \) to a jointly typical 
\[
U_1 = \gamma_N^{(1)}(w_{10}, w_{11}), \quad w_{11} \in [2^{NR_1}],
\]
and map \( (X_1, X_2, U_2) \) to a jointly typical 
\[
Y_2 = \beta_N^{(2)}(w_{00}, w_{10}), \quad w_{00} \in [2^{NR_0}].
\]
- Given \( (w_{10}, w_{11}, w_{00}) \) and \( (U_1, U_2, Y_2) \) from the previous steps, map \( (X_1, X_2, U_1, U_2, Y_2) \) to a jointly typical 
\[
Y_1 = \beta_N^{(1)}(w_{00}, w_{01}, w_{10}, w_{11}), \quad w_{01} \in [2^{NR_0}].
\]

When a jointly typical codeword is not available, the string is encoded to index 0, which is represented by an atypical codeword. The error probability bounds are derived as in the prior examples.

**APPENDIX V**

**EXAMPLE MODELS:**

**DERIVATIONS AND ACCURACY BOUNDS**

The following examples employ notation from Appendix IV.

**V.A. Binary Symmetric Broadcast Channel**

Let \( C = (\{0, 1\}, p(y_1, y_2|x), \{0, 1\}^2) \) be a binary symmetric broadcast channel with 
\[
(Y_1, Y_2) = (X \oplus Z_1, X \oplus Z_2).
\]
Noise vector \( (Z_1, Z_2) \sim q(z_1, z_2) \) is independent of \( X \); components \( Z_1 \) and \( Z_2 \) of that vector may be dependent. Without loss of generality, assume that \( 0 \leq p_1 \leq p_2 \leq 1/2 \), where 
\[
p_1 = \Pr(Z_1 = 1) = q(1, 0) + q(1, 1) \\
p_2 = \Pr(Z_2 = 1) = q(0, 1) + q(1, 1).
\]

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2014.2319075, IEEE Transactions on Information Theory

V-B. Gaussian Broadcast Channel

Let $\mathcal{C} = (\mathbb{R}, p(y_1, y_2|x), \mathbb{R}^2)$ be a two-receiver real Gaussian broadcast channel with

$$(Y_1, Y_2) = (a_1 X + Z_1, a_2 X + Z_2),$$

and power constraint $P$. Here $Z_1$ and $Z_2$ are zero-mean Gaussian random variables with variances $N_1$ and $N_2$ and covariance $\rho \sqrt{N_1 N_2}$, satisfying $N_1/a_1^2 \leq N_2/a_2^2$. Figure 4(b) gives example lower and upper bounding models $C(\mathcal{R}_L(Y_1, Y_2))$ and $C(\mathcal{R}_U(Y_1, Y_2))$. Again, $\mathcal{R}_U(Y_1, Y_2) = (I(U; Y_2), I(X; Y_1|U), 0) \in \mathcal{A}(\mathcal{C})$ and $\mathcal{R}_L(Y_1, Y_2) = (C_2, C_{12} - C_2, 0) \in \mathcal{E}(\mathcal{C})$. The lower bound is evaluated with $X = U + V$ for independent Gaussian random variables $U$ and $V$ of power $\rho P$ and $(1 - \rho P)$. For the upper bound, $C_i = (1/2) \log(1 + (a_i^2 P)/N_i)$ for $i \in \{1, 2\}$, and

$$C_{12} = \frac{1}{2} \log \left( 1 + \frac{(a_1^2 N_2 - 2a_1 a_2 \rho \sqrt{N_1 N_2} + a_2^2 N_1) P}{N_1 N_2 (1 - \rho^2)} \right).$$

The cut set bounds between the transmitter and all receivers are identical in the channel and upper bound.
The accuracy bounds are

\[
\mu(C) \leq \min_{\alpha \in [0,1/2]} \max \left\{ \frac{1}{2} \log \left( 1 + \frac{p}{N_2/a_2^2} \right), \frac{1}{2} \log \left( 1 + \frac{1}{(1-\alpha)p + N_2/a_2^2} \right), \frac{1}{2} \log \left( 1 + \frac{1}{(1-\alpha)p + N_1/a_1^2} \right) \right\}
\]

Figures 12(a) and (b) plot this bound as a function of noise correlation \( \rho \). In (a), each curve corresponds to a different value of \( N_1/a_1^2 = N_2/a_2^2 \). For each \( \rho \), \( \mu(C) \) improves (decreases) as the SNR increases. For each SNR, \( \mu(C) = 1 \) when \( \rho = 1 \): the upper and lower bounding models are identical to each other and equivalent to the noisy channel. In (b), the SNR at receiver 1 is fixed while the SNR at receiver 2 varies from one curve to the next. For some parameters, a receiver that observes both channel outputs can solve from one curve to the next. For some parameters, a receiver that observes both channel outputs can solve.

V-C. Binary Symmetric Multiple Access Channel

Let \( C = \{\{0,1\}^2, p(y|x_1, x_2), \{0,1\}\} \) be a binary symmetric multiple access channel with

\[ Y = X_1 \oplus X_2 \oplus Z. \]

Noise \( Z \) is independent of \( (X_1, X_2) \), and \( \Pr(Z = 1) = p \in [0,1/2] \). Figure 4(c) shows the channel and a pair of bounding models. The lower and upper bounds correspond to the points

\[ R^{(V_1, V_2)}_{\{1\}} = \left( (1-H(p)), (1-\alpha)(1-H(p)) \right) \]

\[ \mathcal{R}^{(V_1, V_2)}_{\{2\}} = \left( 1 - H(p), 0, 0 \right) \]

in \( \mathcal{A}(C) \) and \( \mathcal{E}(C) \), respectively. While the topologies of these bounds differ, a lower bound of the same topology as the upper bound can be derived by applying Lemma 1 to each lower bounding models different to each other and each SNR, \( \mu \) of noise correlation \( \rho \).

V-D. Gaussian Multiple Access Channel

Let \( C = (\mathbb{R}^2, p(y|x_1, x_2), \mathbb{R}) \) be a Gaussian multiple access channel with

\[ Y = X_1 + X_2 + Z, \]

\[ E([X_i]^2) \leq P_i, \quad \text{and} \quad Z \sim N(0, N). \]

Figure 4(d) shows lower bounding models \( \mathcal{C}(R^{(V_1, V_2)}_{\{1\}}) \) and upper bounding model \( \mathcal{C}(R^{(V_1, V_2)}_{\{2\}}) \) for \( \mathcal{C} \). Vector \( R^{(V_1, V_2)}_{\{1\}} \) lies on the dominant face of capacity region

\[ \mathcal{A}(C) = \left\{ (R_1, R_2) : R_1 \leq \frac{1}{2} \log \left( 1 + \frac{P_1}{N} \right), R_1 + R_2 \leq \frac{1}{2} \log \left( 1 + \frac{P_1 + P_2}{N} \right) \right\}. \]

For the upper bound, we evaluate point

\[ R^{(V_1, V_2)}_{\{2\}} = \left( R_0, R_1, R_2 \right) \]

\[ = (I(X_1, X_2; Y|U), I(X_1; U), 0), \]

with \( X_1 = U + V \) and \( X_2 = \beta X_1 + W \); the given \( R^{(V_1, V_2)}_{\{2\}} \) falls in the emulation region by Theorem 8. Here \( \alpha \in [0,1], \beta \in [-\sqrt{P_2/P_1}, \sqrt{P_2/P_1}], \) and \( U, V, \) and \( W \) are independent Gaussian random variables with variances \( (1-\alpha)P_1, \alpha P_1, \) and \( P_2 - \beta^2 P_1 \), giving

\[ R_0 = \frac{1}{2} \log \left( 1 + \frac{(1 + \beta)^2 \alpha - \beta^2}{N} P_1 + P_2 \right) \]

\[ R_1 = \frac{1}{2} \log(1/\alpha). \]

The upper bounding model in the figure is evaluated at \( \beta = \alpha/(1-\alpha) \). The channel can carry up to \( (1/2) \log(1 + (\sqrt{P_2} + \sqrt{P_1})^2/N) \) bits per channel use when both transmitters are controlled (directly or indirectly) by a single transmitter; thus the sum rate of the upper bound is tight when \( \alpha = 1 \) and \( \beta = \sqrt{P_2/P_1} \). Since the bounding model topologies differ, Lemma 10 does not apply.

V-E. Binary Symmetric Channels with Dependent Noise

Let \( C = \{\{0,1\}^2, p(y_1, y_2|x_1, x_2), \{0,1\}^2\} \) be a pair of binary symmetric channels defined by

\[ (Y_1, Y_2) = (X_1 \oplus Z_1, X_1 \oplus Z_2). \]
Suppose that \((Z_1, Z_2)\) is independent of \((X_1, X_2)\) but \(Z_1\) and \(Z_2\) are dependent on each other. Since the channel distribution \(p(y_1, y_2 | x_1, x_2)\) does not factor as \(p(y_1 | x_1)p(y_2 | x_2)\), the channel models for \(p(y_1 | x_1)\) and \(p(y_2 | x_2)\) do not apply; instead, we apply the bounding models for the interference channel. Figure 4(e) shows the union bound. Evaluating Theorem 9 when \(U_1\) and \(U_2\) are deterministic gives

\[
\delta_1 \supseteq \{ R_{\mathcal{U}}^{(1, 2)} \in [0, \infty)^4 : R_{00} \geq 1 - H(Z_2), \quad R_{00} + R_{01} \geq 2 - H(Z_1, Z_2) \}.
\]

The cut-set bound between all inputs and outputs is

\[
\max_{p(x_1, x_2)} I(X_1, X_2; Y_1, Y_2) = 2 - H(Z_1, Z_2),
\]

in both the channel and its upper bounding model.

ACKNOWLEDGMENT

R. Koetter was with the Technical University of Munich when the early drafts of this work were written. He spent many of his last working hours pursuing its completion with the loving support of his wife Nuala, to whom this paper is dedicated. The final manuscript was completed after his passing. His co-authors take full responsibility for any errors or omissions.

REFERENCES


Ralf Koetter (S’92–M’95–SM’06–F’09) was born in Königstein im Taunus, Germany, on October 10, 1963. He received a Diploma in Electrical Engineering from the Technische Universität Darmstadt, Germany, in 1990 and the Ph.D. degree from the Department of Electrical Engineering, Linköping University, Sweden, in 1996. From 1996 to 1997, he was a Visiting Scientist at the IBM Almaden Research Center, San Jose, CA. He was a Visiting Assistant Professor at the University of Illinois, Urbana-Champaign, and a Visiting Scientist at CNRS in Sophia-Antipolis, France, from 1997 to 1998. During 1999–2006 he was member of the faculty at the University of Illinois, Urbana-Champaign. In 2006, Prof. Koetter joined the faculty of the Technische Universität München, Germany, as the Head of the Institute for Communications Engineering (Lehrstuhl für Nachrichtentechnik). He passed away on February 2, 2009.

His research interests were in coding theory and information theory, and in their applications to communication systems. During 1999–2001, Prof. Koetter was an Associate Editor for Coding Theory and Techniques for the IEEE TRANSACTIONS ON COMMUNICATIONS, and during 2000–2003, he served as an Associate Editor for Coding Theory for the IEEE TRANSACTIONS ON INFORMATION THEORY. He was Technical Program Co-Chair for the 2008 International Symposium on Information Theory, and twice Co-Editor-in-Chief for special issues of the IEEE TRANSACTIONS ON INFORMATION THEORY. During 2003–2008, he was a member of the Board of Governors of the IEEE Information Theory Society. He received an IBM Invention Achievement Award in 1997, an NSF CAREER Award in 2000, an IBM Partnership Award in 2001, and a Xerox Award for faculty research in 2006. He received the IEEE Information Theory Society Paper Award in 2004, the Vodafone Innovationspreis in 2008, the Best Paper Award from the IEEE Signal Processing Society in 2008, and the IEEE Communications Society & Information Theory Society Joint Paper Award twice, in 2009 and in 2010.
Michelle Effros (S’93-M’95-SM’02-F’09) received the B.S. degree with distinction in 1989, the M.S. degree in 1990, and the Ph.D. degree in 1994, all in electrical engineering from Stanford University. She joined the faculty at the California Institute of Technology in 1994, where she is currently the George van Osdol Professor of Electrical Engineering. Her research interests include information theory, network coding, data compression, and communications.

Prof. Effros received Stanford’s Frederick Emmons Terman Engineering Scholastic Award (for excellence in engineering) in 1989, the Hughes Masters Full-Study Fellowship in 1989, the National Science Foundation Graduate Fellowship in 1990, the AT&T Ph.D. Scholarship in 1993, the NSF CAREER Award in 1995, the Charles Lee Powell Foundation Award in 1997, the Richard Feynman-Hughes Fellowship in 1997, and an Okawa Research Grant in 2000. She was cited by Technology Review as one of the world’s top 100 young innovators in 2002. She and her co-authors received the Communications Society and Information Theory Society Joint Paper Award in 2009. She became a fellow of the IEEE in 2009.

She is a member of Tau Beta Pi, Phi Beta Kappa, and Sigma Xi. She served as the Editor of the IEEE INFORMATION THEORY SOCIETY NEWSLETTER from 1995 to 1998 and as a Member of the Board of Governors of the IEEE Information Theory Society from 1998 to 2003 and 2008 to the present. She currently serves as First Vice President of the Information Theory Society. She has been a member of the Advisory Committee for the Computer and Information Science and Engineering (CISE) Directorate at the National Science Foundation from 2009 to 2012. She served on the IEEE Signal Processing Society Image and Multi-Dimensional Signal Processing (IMDSP) Technical Committee from 2001 to 2007 and on ISAT from 2006 to 2009. She served as Associate Editor for the joint special issue on Networking and Information Theory in the IEEE TRANSACTIONS ON INFORMATION THEORY and the IEEE/ACM TRANSACTIONS ON NETWORKING and as Associate Editor for Source Coding for the IEEE TRANSACTIONS ON INFORMATION THEORY from 2004 to 2007. She has served on numerous technical program committees and review boards, including serving as general co-chair for the 2009 Network Coding Workshop and technical program committee co-chair for the 2012 IEEE International Symposium on Information Theory.

Muriel Médard (S’90-M’95-SM’02-F’08) received B.S. degrees in electrical engineering and computer science as well as mathematics in 1989, the B.S. degree in humanities in 1990, the M.S. degree in electrical engineering in 1991, and the Sc.D. degree in electrical engineering in 1995, all from the Massachusetts Institute of Technology (MIT), Cambridge.

She is a Professor of Electrical Engineering and Computer Science at MIT. She was previously an Assistant Professor in the Electrical and Computer Engineering Department and a member of the Coordinated Science Laboratory at the University of Illinois Urbana-Champaign. From 1995 to 1998, she was a Staff Member at MIT Lincoln Laboratory in the Optical Communications and the Advanced Networking Groups. Her research interests are in the areas of network coding and reliable communications, particularly for optical and wireless networks.

Prof. Médard has served as an Associate Editor for the Optical Communications and Networking Series of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, as an Associate Editor in Communications for the IEEE TRANSACTIONS ON INFORMATION THEORY, and as an Associate Editor for the OSA JOURNAL OF OPTICAL NETWORKING. She has also served as Guest Editor for the IEEE/OSA JOURNAL OF LIGHTWAVE TECHNOLOGY, for the 2006 Joint Special Issue of the IEEE TRANSACTIONS ON INFORMATION THEORY and the IEEE/ACM TRANSACTIONS ON NETWORKING on “Networking and Information Theory” and for the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY Special Issue on “Statistical Methods for Network Security and Forensics.” She serves as an associate editor for the IEEE/OSA JOURNAL OF LIGHTWAVE TECHNOLOGY. She is a recipient of the William R. Bennett Prize in the Field of Communications Networking, the 2002 IEEE Leon K. Kirchmayer Prize Paper Award, and the Best Paper Award at the Fourth International Workshop on the Design of Reliable Communication Networks (DRCN 2003). She received the NSF CAREER Award in 2001 and was corecipient of the 2004 Harold E. Edgerton Faculty Achievement Award at MIT. She was named a 2007 Gilbreth Lecturer by the National Academy of Engineering. She serves as a member of the Board of Governors of the IEEE Information Theory Society.