Nonconvexity of the Capacity Region of the Multiple-Access Arbitrarily Varying Channel Subject to Constraints

John A. Gubner, *Member, IEEE*, and Brian L. Hughes, *Member, IEEE*

***Abstract*— The random-code capacity region of the multiple-access arbitrarily varying channel subject to both state and input constraints is determined. Consideration of a simple erasure channel shows that the capacity region is not convex in general.**

***Index Terms*—Random code, multiple-access channel, arbitrarily varying channel, state constraint, input constraint, convexity.**

I. Introduction

THE (discrete memoryless) *multiple-access arbitrarily varying channel* (MAVC) subject to constraints models a jammed multiuser channel in which the transmitters and jammers are constrained in average power. Our interest is in characterizing the random-code capacity region of such channels. These results can be applied to determine the performance limits of communications systems, such as spread spectrum, in which the transmitter's code is varied with time in a random (or pseudorandom) manner that is known to the receiver but unknown to the jammers.

Formally, the MAVC with two senders, one jammer, and one receiver is defined to be a transition probability W from $\mathbb{X} \times \mathbb{Y} \times \mathbb{S}$ into $\mathbb{Z}$, where $\mathbb{X}$, $\mathbb{Y}$, $\mathbb{S}$, and $\mathbb{Z}$ are finite sets. We interpret $W(z|x,y,s)$ as the conditional probability that the channel output is $z \in \mathbb{Z}$ given that the channel input symbol from sender 1 is $x \in \mathbb{X}$, the channel input symbol from sender 2 is $y \in \mathbb{Y}$, and the channel *state* (jammer symbol) is $s \in \mathbb{S}$. The channel operation on n-tuples $\boldsymbol{x} = (x_1, \cdots, x_n) \in \mathbb{X}^n$, $\boldsymbol{y} \in \mathbb{Y}^n$, $\boldsymbol{s} \in \mathbb{S}^n$, and $\boldsymbol{z} \in \mathbb{Z}^n$ is given by

$$W^n(\boldsymbol{z}|\boldsymbol{x},\boldsymbol{y},\boldsymbol{s}) := \prod_{k=1}^{n} W(z_k|x_k,y_k,s_k).$$

In general, the state sequence $\boldsymbol{s}$, which is unknown to the senders and the receiver, can be completely arbitrary. However, to model the power limitations in practical communication systems, we require that $\boldsymbol{s}$ satisfy the *state constraint*

$$\ell_n(\boldsymbol{s}) := \frac{1}{n}\sum_{k=1}^{n} \ell(s_k) \;\le\; L \tag{1}$$

where ℓ is a nonnegative function defined on $\mathbb{S}$ and $L \ge 0$. In other words, $\ell(s_k)$ represents the energy in the symbol s_k, and L is a bound on the time-average power of the sequence $\boldsymbol{s}$. In Section II, analogous *input constraints* are imposed on the transmitted codewords.

The main contribution of this paper is the characterization of the random-code capacity region of the MAVC subject to state and input constraints. The analogous results for the single-user arbitrarily varying channel (AVC) were established by Csiszár and Narayan [4]. For the MAVC in the absence of state and input constraints, Jahn [11] established the deterministic-code capacity region, assuming it has a nonempty interior. As an easy corollary, he characterized the *random*-code capacity region, again assuming that the *deterministic*-code capacity region has a nonempty interior. Without this assumption, Jahn did not know whether his characterization of the random-code capacity region still holds (see the paragraph preceding [11, Remark IIA3, p. 214]). By specializing our coding theorem to the unconstrained MAVC, it is clear that Jahn's characterization always holds. Further work on deterministic codes for the MAVC has appeared in [6], [8], [9]. Deterministic codes for the single-user AVC have been studied by several authors, e.g., [1], [2], [5], and the references therein.

For the MAVC, state constraints introduce subtle problems which have no counterpart in the theory of the single-user AVC or the conventional multiple-access channel (MAC). One such problem, near the heart of the present paper, concerns the classic *time-sharing principle* employed in the proofs of many multiuser coding theorems. This principle, which asserts that the capacity region is a convex set, holds for the conventional MAC [3, p. 272] and even for the unconstrained MAVC [11]. However, it fails for the MAVC subject to a state constraint. The difficulty arises as follows [9]. Suppose $n = n_1 + n_2$. Then the inequality in (1) does *not* imply that both

$$\frac{1}{n_1}\sum_{k=1}^{n_1} \ell(s_k) \le L \quad \text{and} \quad \frac{1}{n_2}\sum_{k=1}^{n_2} \ell(s_{n_1+k}) \le L$$

which is a necessary condition to apply the time-sharing principle to an MAVC subject to a state constraint. As a consequence of this observation, we are able to show that the

Manuscript received June 16, 1993; revised June 15, 1994. J. A. Gubner was supported by the Air Force Office of Scientific Research under Grant F49620-92-J-0305. B. L. Hughes was supported by the U.S. Army Research Office under Grant DAAL03-89-K-0130. This work was presented in part at the 1994 IEEE International Symposium on Information Theory, Trondheim, Norway, June 27–July 1, 1994.

J. A. Gubner is with the Department of Electrical and Computer Engineering, University of Wisconsin, Madison, WI 53706 USA.

B. L. Hughes is with the Department of Electrical and Computer Engineering, The Johns Hopkins University, Baltimore, MD 21218 USA.

IEEE Log Number 9406733.

capacity region of this channel is *not* convex in general. We remark here that, whereas the classic time-sharing principle fails, an alternate form of time-sharing based on an auxiliary variable [3, p. 278] will play a central role in this paper.

The remainder of this paper is organized as follows. Our main coding theorem is stated in Section II and proved in Section III. In Section IV, we consider a simple erasure channel and show that its capacity region is not convex.

II. Definitions and Statement of Results

Let N, M, and n be positive integers. A (deterministic) *code* of blocklength n is a triple (f, g, φ) consisting of a codebook $f = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N)$ for sender 1, a codebook $g = (\boldsymbol{y}_1, \cdots, \boldsymbol{y}_M)$ for sender 2, and a decoder φ: $\mathbb{Z}^n \to \{1, \cdots, N\} \times \{1, \cdots, M\}$, where $\boldsymbol{x}_i \in \mathbb{X}^n$, $i = 1, \cdots, N$, and $\boldsymbol{y}_j \in \mathbb{Y}^n$, $j = 1, \cdots, M$. The set of all codes $u = (f, g, \varphi)$ is denoted by $\mathcal{U}_n^{NM}$. A $\mathcal{U}_n^{NM}$-valued random variable is called a *random code*.

The *error probability* of the code $u \in \mathcal{U}_n^{NM}$ for message (i, j) and state sequence $\boldsymbol{s} \in \mathbb{S}^n$ is given by

$$e_{ij}(\boldsymbol{s}, u) := W^n(\{\boldsymbol{z} \in \mathbb{Z}^n : \varphi(\boldsymbol{z}) \neq (i, j)\} | \boldsymbol{x}_i, \boldsymbol{y}_j, \boldsymbol{s}). \quad (2)$$

The corresponding *average error probability* is

$$e(\boldsymbol{s}, u) := \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} e_{ij}(\boldsymbol{s}, u).$$

In Section I, we assumed that ℓ is a nonnegative function defined on $\mathbb{S}$. For convenience, we assume $\min_{s \in \mathbb{S}} \ell(s) = 0$ and set $\ell_{\max} := \max_{s \in \mathbb{S}} \ell(s)$. We also let

$$\mathbb{S}^n(L) := \{\boldsymbol{s} \in \mathbb{S}^n : \ell_n(\boldsymbol{s}) \leq L\}.$$

Note that if $L \geq \ell_{\max}$, then $\mathbb{S}^n(L) = \mathbb{S}^n$ and we say that the state constraint is inactive.

We now impose constraints on the input sequences. Let a and b be nonnegative functions defined on $\mathbb{X}$ and $\mathbb{Y}$, respectively. We assume $\min_x a(x) = \min_y b(y) = 0$. Set $a_{\max} := \max_x a(x)$ and $b_{\max} := \max_y b(y)$. We say that a random code $U = (F, G, \Phi)$ satisfies *input constraint* (A, B) if almost surely we have

$$\frac{1}{n} \sum_{k=1}^{n} a(X_{i,k}) \leq A, \qquad i = 1, \cdots, N, \quad (3a)$$

and

$$\frac{1}{n} \sum_{k=1}^{n} b(Y_{j,k}) \leq B, \qquad j = 1, \cdots, M. \quad (3b)$$

When $A \geq a_{\max}$ and $B \geq b_{\max}$, *all* codes satisfy (3a) and (3b), and the input constraints are said to be inactive.

Definition 1: A pair of nonnegative real numbers, (R_1, R_2), is said to be *achievable* under state constraint L and input constraint (A, B) if, for every $0 < \lambda < 1$, every $\Delta R > 0$, and all sufficiently large n, there exists a $\mathcal{U}_n^{NM}$-valued random code $U = (F, G, \Phi)$ such that $N > \exp[n(R_1 - \Delta R)]$ and $M > \exp[n(R_2 - \Delta R)]$, F and G are independent, (3a) and (3b) are satisfied, and

$$E[e(\boldsymbol{s}, U)] \leq \lambda, \quad \text{for all } \boldsymbol{s} \in \mathbb{S}^n(L). \quad (4)$$

Here and in the sequel, exp, log, and all entropies and mutual informations are to the base 2.

Definition 2: The (random-code, average-error) *capacity region* of the MAVC W, under state constraint L and input constraint (A, B), is defined to be the set of all pairs (R_1, R_2) that are achievable in the sense of Definition 1. We denote this region by $\mathcal{C}(L, A, B)$.

The random coding problem is to characterize $\mathcal{C}(L, A, B)$.

Remark: Following all earlier work on random coding for the MAVC [6]–[8], [11], we have restricted F and G to be independent. If this restriction is dropped, the capacity region can be strictly larger than $\mathcal{C}(L, A, B)$. To keep the present paper to a reasonable size, we defer discussion of this issue to a future paper.

In order to present our result, we require the following notation. If $\mathbb{V}$ is a finite set, $\mathcal{D}(\mathbb{V})$ denotes the set of all probability distributions on $\mathbb{V}$. Given another finite set $\mathbb{S}$, $\mathcal{D}(\mathbb{S}|\mathbb{V})$ denotes the set of all transition probabilities from $\mathbb{V}$ into $\mathbb{S}$.

When $\gamma \in \mathcal{D}(\mathbb{V})$, $p \in \mathcal{D}(\mathbb{X}|\mathbb{V})$, $q \in \mathcal{D}(\mathbb{Y}|\mathbb{V})$, and $r \in \mathcal{D}(\mathbb{S}|\mathbb{V})$ are understood, we let V, X, Y, S, and Z be random variables with joint distribution

$$\mathbb{P}_{VXYSZ}(v, x, y, s, z) = \gamma(v)p(x|v)q(y|v)r(s|v)W(z|x, y, s). \quad (5)$$

Observe that

$$E[\ell(S)] = \sum_v \gamma(v) \left(\sum_s \ell(s) r(s|v) \right) \quad (6)$$

and depends only on γ and r. We therefore define

$$\mathcal{D}^L(\mathbb{S}|\gamma) := \{r \in \mathcal{D}(\mathbb{S}|\mathbb{V}) : E[\ell(S)] \leq L\}.$$

Next define

$$I^L(Y \wedge Z|XV) := \inf_{r \in \mathcal{D}^L(\mathbb{S}|\gamma)} I(Y \wedge Z|XV) \quad (7)$$

which depends on γ, p, q, and W, but not r. The quantities $I^L(X \wedge Z|YV)$ and $I^L(XY \wedge Z|V)$ are defined similarly. With these definitions, we set

$$\begin{aligned} \mathcal{R}(L, \gamma, p, q) := \{&(R_1, R_2) : \\ &0 \leq R_1 \leq I^L(X \wedge Z|YV), \\ &0 \leq R_2 \leq I^L(Y \wedge Z|XV), \\ &R_1 + R_2 \leq I^L(XY \wedge Z|V)\}. \end{aligned} \quad (8)$$

Recalling (5) and (6), we similarly observe that $E[a(X)]$ depends only on γ and p and that $E[b(Y)]$ depends only on γ and q. We therefore define

$$\mathcal{D}^A(\mathbb{X}|\gamma) := \{p \in \mathcal{D}(\mathbb{X}|\mathbb{V}) : E[a(X)] \leq A\}$$

and

$$\mathcal{D}^B(\mathbb{Y}|\gamma) := \{q \in \mathcal{D}(\mathbb{Y}|\mathbb{V}) : E[b(Y)] \leq B\}.$$

Let

$$\mathcal{R}(L, A, B) := \bigcup \mathcal{R}(L, \gamma, p, q) \quad (9)$$

where the union is over all sets $|\mathbb{V}| < \infty$, all $\gamma \in \mathcal{D}(\mathbb{V})$, and all $p \in \mathcal{D}^A(\mathbb{X}|\gamma)$ and $q \in \mathcal{D}^B(\mathbb{Y}|\gamma)$.

Our main result is the following coding theorem, which is proved in Section III.

Coding Theorem: $\mathcal{C}(L, A, B)$ is equal to the closure of $\mathcal{R}(L, A, B)$.

As mentioned in Section I, the time-sharing principle does not hold for the MAVC subject to a state constraint. Nevertheless, the use of the auxiliary variable V in (8) can be regarded as a form of time-sharing. In multiuser coding theorems, the use of an auxiliary variable is often equivalent to the operation of taking the convex closure [3, p. 278]; however, this is not generally true of $\mathcal{R}(L, A, B)$ due to the infimum operation in (9). In fact, we show by example in Section IV that $\mathcal{R}(L, A, B)$ is not in general convex. As a consequence, the usual methods for restricting the cardinality of $\mathbb{V}$ in (9)—which depend essentially on convexity—do not apply here.

If the state constraint is inactive, the capacity region is always convex. If $L \geq \ell_{\max}$ then $\mathcal{D}^L(\mathbb{S}|\gamma) = \mathcal{D}^L(\mathbb{S}|\mathbb{V})$ and is independent of γ. Since

$$I(Y \wedge Z|XV) = \sum_v \gamma(v) I(Y \wedge Z|X, V = v)$$

if we enumerate the elements of $\mathbb{V}$ as $v_1, \cdots, v_K$, say, then $I^L(Y \wedge Z|XV)$ can be written as

$$\inf_{r(\cdot|v_1),\ldots,r(\cdot|v_K)} \sum_{k=1}^{K} \gamma(v_k) I(Y \wedge Z|X, V = v_k)$$

which simplifies to

$$\sum_{k=1}^{K} \gamma(v_k) \inf_{r \in \mathcal{D}(\mathbb{S})} I(Y \wedge Z|X, V = v_k) \tag{10}$$

where this last mutual information is computed from the joint distribution on (x, y, s, z)

$$p(x|v_k)q(y|v_k)r(s)W(z|x, y, s).$$

With the foregoing observation, one can prove directly that $\mathcal{R}(L, A, B)$ is convex. If the input constraints are also inactive, the foregoing observation also permits us to conclude that $\mathcal{R}(L, A, B)$ is closed and equals Jahn's random-code capacity region [11]. Let $\mathcal{R}_*(L, A, B)$ denote the right-hand side of (9) with the union restricted to $\mathbb{V}$ such that $|\mathbb{V}| = 1$. Note that $\mathcal{R}_*(L, A, B)$ is closed. It follows from (10) that when the state constraint and input constraints are inactive, $\mathcal{R}(L, A, B)$ is equal to the convex hull of $\mathcal{R}_*(L, A, B)$, which is Jahn's random-code capacity region.

Thus far, we have considered only the average-error probability criterion (4). The definition of the *maximal-error* capacity region, denoted by $\mathcal{C}_m(L, A, B)$, is similar to that of $\mathcal{C}(L, A, B)$ except that (4) is replaced by the *maximal-error criterion*

$$\max_{i,j} E[e_{ij}(\boldsymbol{s}, U)] \leq \lambda, \quad \text{for all } \boldsymbol{s} \in \mathbb{S}^n(L). \tag{11}$$

From the definitions, it is evident that $\mathcal{C}_m(L, A, B) \subseteq \mathcal{C}(L, A, B)$. Implicit in our proof of the coding theorem is a demonstration that $\mathcal{C}(L, A, B) = \mathcal{C}_m(L, A, B)$. This can be seen by observing that the forward part of the coding theorem is proved for the maximal-error criterion (11), while the converse is proved for the average-error criterion (4) (cf. [1, Theorem 2(b)], [11, eq. (5)]).

To conclude this section, we note that it is possible to approximate $\mathcal{R}(L, A, B)$ to any prescribed accuracy as follows. Let $\mathcal{R}_n(L, A, B)$ denote the union of all $\mathcal{R}(L, \tilde{\gamma}, \tilde{p}, \tilde{q})$, where the union is over any fixed set $\widetilde{\mathbb{V}}$ with

$$|\widetilde{\mathbb{V}}| = \binom{n - |\mathbb{X}| - 1}{|\mathbb{X}| - 1} \binom{n - |\mathbb{Y}| - 1}{|\mathbb{Y}| - 1} \tag{12}$$

all $\tilde{\gamma} \in \mathcal{D}(\widetilde{\mathbb{V}})$, all $\tilde{p} \in \mathcal{D}^A(\mathbb{X}|\tilde{\gamma})$, and $\tilde{q} \in \mathcal{D}^B(\mathbb{Y}|\tilde{\gamma})$. Using routine approximation arguments in Appendix II, we have the following result.

Approximation Theorem: For $n \geq 4(|\mathbb{X}| + |\mathbb{Y}|)$

$$\mathcal{R}(L, A, B) \subset \mathcal{R}_n(L, A, B) + \mu_n \mathbf{U} \tag{13}$$

where $\mathbf{U}$ is the unit square $\{(R_1, R_2) : 0 \leq R_1 \leq 1 \text{ and } 0 \leq R_2 \leq 1\}$, and $\mu_n := 2\theta_n \log(\kappa^2/\theta_n)$, $\theta_n := 2(|\mathbb{X}| + |\mathbb{Y}|)/n$, and $\kappa := \max\{|\mathbb{X}|, |\mathbb{Y}|, |\mathbb{Z}|\}$.

III. Proof of the Coding Theorem

Our proof is adapted from [7] and relies heavily on the method of *types* as discussed in [3, pp. 29–33]. Recall that the *type* of an n-tuple $\boldsymbol{x} \in \mathbb{X}^n$ is defined to be the empirical probability distribution $\mathcal{E}_{\boldsymbol{x}}$ given by $\mathcal{E}_{\boldsymbol{x}}(x) = N(x|\boldsymbol{x})/n$ for $x \in \mathbb{X}$, where $N(x|\boldsymbol{x})$ denotes the number of occurrences of x in the n-tuple $\boldsymbol{x}$. The set of types generated by $\mathbb{X}^n$ is denoted by $\mathcal{D}_n(\mathbb{X})$; more precisely, $\mathcal{D}_n(\mathbb{X})$ is the set of $P \in \mathcal{D}(\mathbb{X})$ such that $P = \mathcal{E}_{\boldsymbol{x}}$ for some $\boldsymbol{x} \in \mathbb{X}^n$. In an analogous way, the *joint type* of a pair of n-tuples, $\boldsymbol{x}$ and $\boldsymbol{y}$, is defined by $\mathcal{E}_{\boldsymbol{x},\boldsymbol{y}}(x, y) = N(x, y|\boldsymbol{x}, \boldsymbol{y})/n$ for $x \in \mathbb{X}$ and $y \in \mathbb{Y}$, where $N(x, y|\boldsymbol{x}, \boldsymbol{y})$ denotes the number of occurrences of (x, y) in the n-tuple $((x_1, y_1), \cdots, (x_n, y_n))$. Finally, the conditional type $\mathcal{E}_{\boldsymbol{y}|\boldsymbol{x}}(y|x)$ is given by $\mathcal{E}_{\boldsymbol{x},\boldsymbol{y}}(x, y)/\mathcal{E}_{\boldsymbol{x}}(x)$.

Let $D(\cdot\|\cdot)$ denote the Kullback–Leibler informational divergence [3, p. 20], and recall that the informational divergence dominates the variational distance between two probability distributions [3, p. 58]. We now introduce what we call the ϵ–δ *convention*, by which we mean the following. All of the mutual informations $I(\cdot)$ as well as $I^L(\cdot)$ defined at the beginning of Section II as well as similarly defined entropy functions used below are uniformly continuous functions of the indicated joint distribution $\mathbb{P}_{VXYSZ}$. Thus given $\epsilon > 0$, there exists a $\delta > 0$ such that whenever $D(\mathbb{P}_{VXYSZ}\|\tilde{\mathbb{P}}_{\tilde{V}\tilde{X}\tilde{Y}\tilde{S}\tilde{Z}}) \leq \delta$, any mutual information or entropy evaluated under $\mathbb{P}$ is within $\epsilon/2$ of the same function evaluated under $\tilde{\mathbb{P}}$.

A. Proof of the Forward Result

To prove the forward result, it suffices to prove that $\mathcal{R}(L, \gamma, p, q) \subseteq \mathcal{C}(L, A, B)$ for all $\gamma \in \mathcal{D}(\mathbb{V})$, $p \in \mathcal{D}^A(\mathbb{X}|\gamma)$, and $q \in \mathcal{D}^B(\mathbb{Y}|\gamma)$. Since the capacity region is closed, we may restrict attention to interior points of $\mathcal{R}(L, \gamma, p, q)$. (The case in which $\mathcal{R}(L, \gamma, p, q)$ has no interior points can be handled with minor modifications to the proof to follow.) Suppose (R_1, R_2) is an interior point of $\mathcal{R}(L, \gamma, p, q)$. Then

we can choose $\epsilon > 0$ so small that $R_1 + 2\epsilon < I^L(X \wedge Z|YV)$, $R_2 + 2\epsilon < I^L(Y \wedge Z|XV)$, and $R_1 + R_2 + 2\epsilon < I^L(XY \wedge Z|V)$. Let δ be as in the ϵ–δ convention. Let n be sufficiently large; as will become clear in the course of the proof, how large n needs to be depends only on ϵ, δ, and the cardinalities of the sets $\mathbb{V}$, $\mathbb{X}$, $\mathbb{Y}$, $\mathbb{S}$, and $\mathbb{Z}$. We begin by assuming n is large enough that we can approximate γ by a type $\Gamma \in \mathcal{D}_n(\mathbb{V})$ and we can approximate p and q by conditional types $P \in \mathcal{D}_n(\mathbb{X}|\mathbb{V})$ and $Q \in \mathcal{D}_n(\mathbb{Y}|\mathbb{V})$ such that if

$$\tilde{\mathbb{P}}_{\tilde{V}\tilde{X}\tilde{Y}\tilde{S}\tilde{Z}}(v,x,y,s,z) = \Gamma(v)P(x|v)Q(y|v)\tilde{r}(s|v)W(z|x,y,s) \quad (14)$$

where $\tilde{r} \in \mathcal{D}^L(\mathbb{S}|\Gamma)$, then $I^L(\tilde{X} \wedge \tilde{Z}|\tilde{Y}\tilde{V}) > I^L(X \wedge Z|YV) - \epsilon$, $I^L(\tilde{Y} \wedge \tilde{Z}|\tilde{X}\tilde{V}) > I^L(Y \wedge Z|XV) - \epsilon$, $I^L(\tilde{X}\tilde{Y} \wedge \tilde{Z}|\tilde{V}) > I^L(XY \wedge Z|V) - \epsilon$, and such that

$$E[a(\tilde{X})] \le A \quad \text{and} \quad E[b(\tilde{Y})] \le B. \quad (15)$$

Let $\lfloor t \rfloor$ denote the greatest integer less than or equal to t. For $N = \lfloor \exp(nR_1) \rfloor$ and $M = \lfloor \exp(nR_2) \rfloor$, we will construct a random code $U = (F, G, \Phi)$ with F and G independent, and satisfying (3a) and (3b) almost surely and such that for all i, j and all $\boldsymbol{s} \in \mathbb{S}^n(L)$

$$\begin{aligned} E[e_{ij}(\boldsymbol{s}, U)] &\le 3\exp(-n\delta/8) \\ &\quad + \exp[-n(I^L(X \wedge Z|YV) - R_1 - 2\epsilon)] \\ &\quad + \exp[-n(I^L(Y \wedge Z|XV) - R_2 - 2\epsilon)] \\ &\quad + \exp[-n(I^L(XY \wedge Z|V) - R_1 - R_2 - 2\epsilon)]. \end{aligned} \quad (16)$$

Observe that, given any $\Delta R > 0$ and sufficiently large n, $\exp[n(R_1 - \Delta R)] < N \le \exp(nR_1)$ and $\exp[n(R_2 - \Delta R)] < M \le \exp(nR_2)$.

The Decoder: Fix any $\boldsymbol{v} \in \mathbb{V}^n$ such that $\mathcal{E}_{\boldsymbol{v}} = \Gamma$. Given a set of codewords $f = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N)$ and $g = (\boldsymbol{y}_1, \cdots, \boldsymbol{y}_M)$, we show how to construct a typicality decoder φ. In other words, we shall prescribe a deterministic function typ and take $\varphi = \text{typ}(f, g, \boldsymbol{v})$. If (F, G) is a random set of codewords, we can generate a random code $U = (F, G, \Phi)$ by taking $\Phi = \text{typ}(F, G, \boldsymbol{v})$.

For $\boldsymbol{s} \in \mathbb{S}^n$, let $(\Gamma \times P \times Q \times \mathcal{E}_{\boldsymbol{s}|\boldsymbol{v}} \times W)(v, x, y, s, z)$ denote the distribution

$$\Gamma(v)P(x|v)Q(y|v)\mathcal{E}_{\boldsymbol{s}|\boldsymbol{v}}(s|v)W(z|x,y,s).$$

Then let

$$K(\boldsymbol{s}) := \{(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) : D(\mathcal{E}_{\boldsymbol{v},\boldsymbol{x},\boldsymbol{y},\boldsymbol{s},\boldsymbol{z}} \| \Gamma \times P \times Q \times \mathcal{E}_{\boldsymbol{s}|\boldsymbol{v}} \times W) \le \delta\}$$

and set

$$K := \bigcup_{\boldsymbol{s}' \in \mathbb{S}^n(L)} K(\boldsymbol{s}'). \quad (17)$$

Now, to define a decoder on $\mathbb{Z}^n$, we use the sets

$$K_{ij} := \{\boldsymbol{z} : (\boldsymbol{x}_i, \boldsymbol{y}_j, \boldsymbol{z}) \in K\}$$

as follows. Let

$$E_{ij} := K_{ij} \cap \Big(\bigcup_{i' \ne i} K_{i'j}\Big)^c \cap \Big(\bigcup_{j' \ne j} K_{ij'}\Big)^c \cap \Big(\bigcup_{i' \ne i,\, j' \ne j} K_{i'j'}\Big)^c. \quad (18)$$

Clearly, the $\{E_{ij}\}$ are disjoint. Let φ be any mapping such that $\boldsymbol{z} \in E_{ij}$ implies $\varphi(\boldsymbol{z}) = (i, j)$. This ensures that $\varphi(\boldsymbol{z}) \ne (i, j)$ implies $\boldsymbol{z} \in E_{ij}^c$.

The Codewords: In order for the decoder to work well, we need to restrict the choice of codewords as follows. Recalling that $\mathcal{E}_{\boldsymbol{v}} = \Gamma$, we let (cf. [3, p. 31])

$$\mathbb{T}_P(\boldsymbol{v}) := \{\boldsymbol{x} \in \mathbb{X}^n : \mathcal{E}_{\boldsymbol{v},\boldsymbol{x}} = \Gamma \times P\}$$

where $(\Gamma \times P)(v, x) := \Gamma(v)P(x|v)$. Similarly, we let

$$\mathbb{T}_Q(\boldsymbol{v}) := \{\boldsymbol{y} \in \mathbb{Y}^n : \mathcal{E}_{\boldsymbol{v},\boldsymbol{y}} = \Gamma \times Q\}.$$

We require that

$$\boldsymbol{x}_i \in \mathbb{T}_P(\boldsymbol{v}) \quad \text{and} \quad \boldsymbol{y}_j \in \mathbb{T}_Q(\boldsymbol{v}). \quad (19)$$

It is important to note here that $\boldsymbol{x} \in \mathbb{T}_P(\boldsymbol{v})$ implies

$$\begin{aligned} \frac{1}{n}\sum_{k=1}^n a(x_k) &= \sum_{x \in \mathbb{X}} a(x)\mathcal{E}_{\boldsymbol{x}}(x) \\ &= \sum_x a(x)\Big(\sum_v \mathcal{E}_{\boldsymbol{v},\boldsymbol{x}}(v,x)\Big) \\ &= \sum_v \Gamma(v)\Big(\sum_x a(x)P(x|v)\Big) \\ &= E[a(\tilde{X})] \\ &\le A, \quad \text{by (15)}. \end{aligned} \quad (20)$$

Thus all our codewords satisfy (3a) and (3b).

Remainder of Proof: For any code $u \in \mathcal{U}_n^{NM}$ whose decoder is as just described, and for any $\boldsymbol{s} \in \mathbb{S}^n(L)$, we can always write

$$\begin{aligned} e_{ij}(\boldsymbol{s}, u) &\le W^n(E_{ij}^c|\boldsymbol{x}_i, \boldsymbol{y}_j, \boldsymbol{s}) \\ &\le W^n(K_{ij}^c|\boldsymbol{x}_i, \boldsymbol{y}_j, \boldsymbol{s}) && (21a) \\ &\quad + \sum_{i' \ne i} W^n(K_{i'j}|\boldsymbol{x}_i, \boldsymbol{y}_j, \boldsymbol{s}) && (21b) \\ &\quad + \sum_{j' \ne j} W^n(K_{ij'}|\boldsymbol{x}_i, \boldsymbol{y}_j, \boldsymbol{s}) && (21c) \\ &\quad + \sum_{i' \ne i,\, j' \ne j} W^n(K_{i'j'}|\boldsymbol{x}_i, \boldsymbol{y}_j, \boldsymbol{s}). && (21d) \end{aligned}$$

We first examine the term in (21a). If we set $K_{ij}(\boldsymbol{s}) := \{\boldsymbol{z} : (\boldsymbol{x}_i, \boldsymbol{y}_j, \boldsymbol{z}) \in K(\boldsymbol{s})\}$, then for $\boldsymbol{s} \in \mathbb{S}^n(L)$, $K_{ij}^c \subseteq K_{ij}(\boldsymbol{s})^c$. Using the method of types, along with (19), it is then easy to show that for sufficiently large n (cf., e.g., [6, ineq. (A.19)])

$$\begin{aligned} &W^n(K_{ij}^c|\boldsymbol{x}_i, \boldsymbol{y}_j, \boldsymbol{s}) \\ &\quad \le W^n(K_{ij}(\boldsymbol{s})^c|\boldsymbol{x}_i, \boldsymbol{y}_j, \boldsymbol{s}) \\ &\quad = \sum_{\boldsymbol{z}:(\boldsymbol{x}_i, \boldsymbol{y}_j, \boldsymbol{z}) \in K(\boldsymbol{s})^c} W^n(\boldsymbol{z}|\boldsymbol{x}_i, \boldsymbol{y}_j, \boldsymbol{s}) && (22a) \\ &\quad \le \exp\big[-n\big(3\delta/4 - I(\boldsymbol{x}_i \wedge \boldsymbol{s}|\boldsymbol{v}) \\ &\qquad - I(\boldsymbol{y}_j \wedge \boldsymbol{x}_i\boldsymbol{s}|\boldsymbol{v})\big)\big] \end{aligned}$$

where $I(\boldsymbol{x}_i \wedge \boldsymbol{s}|\boldsymbol{v})$ and $I(\boldsymbol{y}_j \wedge \boldsymbol{x}_i\boldsymbol{s}|\boldsymbol{v})$ are shorthand for the obvious mutual informations computed with the distribution $\mathcal{E}_{\boldsymbol{v},\boldsymbol{x}_i,\boldsymbol{y}_j,\boldsymbol{s}}$.

Now consider a term in the sum in (21b). We can write

$$W^n(K_{i'j}|\boldsymbol{x}_i,\boldsymbol{y}_j,\boldsymbol{s}) = \sum_{\boldsymbol{z}:(\boldsymbol{x}_{i'},\boldsymbol{y}_j,\boldsymbol{z})\in K} W^n(\boldsymbol{z}|\boldsymbol{x}_i,\boldsymbol{y}_j,\boldsymbol{s}). \quad (22b)$$

Similarly, the terms in (21c) and (21d) can be written, respectively, as

$$W^n(K_{ij'}|\boldsymbol{x}_i,\boldsymbol{y}_j,\boldsymbol{s}) = \sum_{\boldsymbol{z}:(\boldsymbol{x}_i,\boldsymbol{y}_{j'},\boldsymbol{z})\in K} W^n(\boldsymbol{z}|\boldsymbol{x}_i,\boldsymbol{y}_j,\boldsymbol{s}) \quad (22c)$$

and

$$W^n(K_{i'j'}|\boldsymbol{x}_i,\boldsymbol{y}_j,\boldsymbol{s}) = \sum_{\boldsymbol{z}:(\boldsymbol{x}_{i'},\boldsymbol{y}_{j'},\boldsymbol{z})\in K} W^n(\boldsymbol{z}|\boldsymbol{x}_i,\boldsymbol{y}_j,\boldsymbol{s}). \quad (22d)$$

To make further progress in bounding the probability of error, we use the following random code. Recalling (19), we take

$$F = (\boldsymbol{X}_1,\ldots,\boldsymbol{X}_N) \qquad \text{and} \qquad G = (\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_M)$$

where F and G are independent and where the $\{\boldsymbol{X}_i\}$ are independent and uniformly distributed on $\mathbb{T}_P(\boldsymbol{v})$ and the $\{\boldsymbol{Y}_j\}$ are independent and uniformly distributed on $\mathbb{T}_Q(\boldsymbol{v})$. On account of (20), all our random codewords satisfy (3a) and (3b).

Remark: In general outline, our proof is very similar to Jahn's proof of the forward result for the unconstrained two-user AVC [11, sec. III-C]. However, a major difference is in the randomization of the codewords. Jahn required that the letters of each codeword be independent; however, this will not guarantee that (3a) and (3b) will hold almost surely even if $E[a(\tilde{X}_{i,k})] \le A$ and $E[b(\tilde{Y}_{j,k})] \le B$ hold for all i, j, k.

Let $e_a(\boldsymbol{x}_i,\boldsymbol{y}_j,\boldsymbol{s})$ denote the right-hand side of (22a). It is clearly upper bounded by 1. Hence

$$\begin{aligned} E[\,e_a(\boldsymbol{X}_i,\boldsymbol{Y}_j,\boldsymbol{s})\,] &\le \exp(-n\delta/4) \\ &\quad + \Pr\Big(\boldsymbol{X}_i \in \{\boldsymbol{x} : I(\boldsymbol{x}\wedge\boldsymbol{s}|\boldsymbol{v}) > \delta/4\}\Big) \\ &\quad + \Pr\Big((\boldsymbol{X}_i,\boldsymbol{Y}_j) \in \{(\boldsymbol{x},\boldsymbol{y}) : I(\boldsymbol{y}\wedge\boldsymbol{x}\boldsymbol{s}|\boldsymbol{v}) > \delta/4\}\Big). \end{aligned} \quad (23)$$

An easy calculation using the method of types shows that for sufficiently large n, these last two probabilities are each upper-bounded uniformly in $\boldsymbol{s}$ and $\boldsymbol{v}$ by $\exp(-n\delta/8)$.

Let $e_b(\boldsymbol{x}_i,\boldsymbol{y}_j,\boldsymbol{x}_{i'})$, $e_c(\boldsymbol{x}_i,\boldsymbol{y}_j,\boldsymbol{y}_{j'})$, and $e_d(\boldsymbol{x}_i,\boldsymbol{y}_j,\boldsymbol{x}_{i'},\boldsymbol{y}_{j'})$ denote the right-hand sides of (22b)–(22d), respectively. Each is clearly upper bounded by 1. We treat only e_d as the others can be treated similarly. For $i' \ne i$ and $j' \ne j$, we can use independence to compute

$$\begin{aligned} &E[\,e_d(\boldsymbol{X}_i,\boldsymbol{Y}_j,\boldsymbol{X}_{i'},\boldsymbol{Y}_{j'}) \mid \boldsymbol{X}_i = \boldsymbol{x}_i,\, \boldsymbol{Y}_j = \boldsymbol{y}_j\,] \quad (24)\\ &\quad = E[\,e_d(\boldsymbol{x}_i,\boldsymbol{y}_j,\boldsymbol{X}_{i'},\boldsymbol{Y}_{j'})\,] \\ &\quad = \sum_{\boldsymbol{z}\in\mathbb{Z}^n} W^n(\boldsymbol{z}|\boldsymbol{x}_i,\boldsymbol{y}_j,\boldsymbol{s}) \\ &\qquad \cdot \Big(\sum_{(\boldsymbol{x},\boldsymbol{y}):(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})\in K} \mathbb{P}(\boldsymbol{X}_{i'} = \boldsymbol{x}, \boldsymbol{Y}_{j'} = \boldsymbol{y})\Big). \quad (25) \end{aligned}$$

Since $\Pr(\boldsymbol{X}_{i'} = \boldsymbol{x}, \boldsymbol{Y}_{j'} = \boldsymbol{y}) = 1/(|\mathbb{T}_P(\boldsymbol{v})|\,|\mathbb{T}_Q(\boldsymbol{v})|)$, and since the cardinality of $\{(\boldsymbol{x},\boldsymbol{y}) : (\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}) \in K\}$ is upper bounded by (recall (17), (14), and the ϵ–δ convention)

$$(n+1)^{|\mathcal{V}||\mathcal{X}||\mathcal{Y}||\mathcal{Z}|}\exp[n(H^L(\tilde{X}\tilde{Y}|\tilde{Z}\tilde{V}) + \epsilon/2)]$$

where

$$H^L(\tilde{X}\tilde{Y}|\tilde{Z}\tilde{V}) := \sup_{\tilde{r}\in\mathcal{D}^L(\mathbb{S}|\Gamma)} H(\tilde{X}\tilde{Y}|\tilde{Z}\tilde{V})$$

the inner sum in (25) is then bounded above by

$$\exp[-n(I^L(\tilde{X}\tilde{Y}\wedge\tilde{Z}|\tilde{V}) - \epsilon)]$$

when n is large enough. Since this bound does not depend on $\boldsymbol{z}$ (recall (14)), the conditional expectation in (24) is upper-bounded by the same quantity. Since the bound does not depend on $\boldsymbol{x}_i$ or $\boldsymbol{y}_j$, we have

$$E[\,e_d(\boldsymbol{X}_i,\boldsymbol{Y}_j,\boldsymbol{X}_{i'},\boldsymbol{Y}_{j'})\,] \le \exp[-n(I^L(\tilde{X}\tilde{Y}\wedge\tilde{Z}|\tilde{V}) - \epsilon)].$$

By a completely analogous procedure, one can show that

$$E[\,e_b(\boldsymbol{X}_i,\boldsymbol{Y}_j,\boldsymbol{X}_{i'})\,] \le \exp[-n(I^L(\tilde{X}\wedge\tilde{Z}|\tilde{Y}\tilde{V}) - \epsilon)]$$

and

$$E[\,e_c(\boldsymbol{X}_i,\boldsymbol{Y}_j,\boldsymbol{Y}_{j'})\,] \le \exp[-n(I^L(\tilde{Y}\wedge\tilde{Z}|\tilde{X}\tilde{V}) - \epsilon)].$$

We now see that

$$\begin{aligned} E[\,e_{ij}(\boldsymbol{s},U)\,] &\le 3\exp(-n\delta/8) \\ &\quad + N\exp[-n(I^L(\tilde{X}\wedge\tilde{Z}|\tilde{Y}\tilde{V}) - \epsilon)] \\ &\quad + M\exp[-n(I^L(\tilde{Y}\wedge\tilde{Z}|\tilde{X}\tilde{V}) - \epsilon)] \\ &\quad + NM\exp[-n(I^L(\tilde{X}\tilde{Y}\wedge\tilde{Z}|\tilde{V}) - \epsilon)] \end{aligned}$$

from which (16) follows. This establishes the forward result.

B. Proof of the Weak Converse

Suppose (R_1, R_2) is an achievable rate pair. We must show it belongs to the closure of $\mathcal{R}(L,A,B)$. It suffices to prove that, for every $0 < \delta < L$, (R_1, R_2) belongs to the closure of $\mathcal{R}(L-\delta,A,B)$ (cf. [8, Lemma 3.1]). Fix $0 < \delta < L$. Let $0 < \lambda < 1$ and $\Delta R > 0$ be arbitrary. Then by Definition 1, for all sufficiently large n, there exist positive integers N and M such that

$$\frac{\log N}{n} > R_1 - \Delta R \quad \text{and} \quad \frac{\log M}{n} > R_2 - \Delta R \quad (26)$$

and such that there exists a random code $U = (F, G, \Phi)$ with F and G independent and satisfying (3a) and (3b) almost surely and

$$\max_{\boldsymbol{s}\in\mathbb{S}^n(L)} E[\,e(\boldsymbol{s},U)\,] \;\le\; \lambda/2. \quad (27)$$

Now, let $r_1,\cdots,r_n$ be any elements of $\mathcal{D}(\mathbb{S})$ such that

$$\frac{1}{n}\sum_{k=1}^{n}\Big(\sum_s \ell(s)r_k(s)\Big) \le L - \delta \quad (28)$$

and set

$$r(\boldsymbol{s}) := \prod_{k=1}^{n} r_k(s_k), \quad \boldsymbol{s} = (s_1, \cdots, s_n) \in \mathbb{S}^n. \tag{29}$$

(Cf. [8, eq. (3.3)] in which the stronger requirement $r_k \in \mathcal{D}^{L-\delta}(\mathbb{S})$ *for each* k was imposed.)

Let α be a $\{1, \cdots, N\}$-valued random variable and β a $\{1, \cdots, M\}$-valued random variable. Let $\boldsymbol{X} = (X_1, \cdots, X_n)$ be an $\mathbb{X}^n$-valued random variable, and similarly, let $\boldsymbol{Y}$, $\boldsymbol{S}$, and $\boldsymbol{Z}$ be $\mathbb{Y}^n$-, $\mathbb{S}^n$-, and $\mathbb{Z}^n$-valued random variables, respectively, whose joint distribution, conditioned on the random code $U = u$, is given by

$$\begin{aligned}\Pr(\boldsymbol{Z} = \boldsymbol{z}, \boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Y} = \boldsymbol{y}, \boldsymbol{S} = \boldsymbol{s}, \beta = j, \alpha = i | U = u) \\ = W^n(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{s}) r(\boldsymbol{s}) \delta_{\boldsymbol{x}_i}(\boldsymbol{x}) \delta_{\boldsymbol{y}_j}(\boldsymbol{y}) / MN\end{aligned} \tag{30}$$

where $u = (f, g, \varphi)$ and $f = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N)$, $g = (\boldsymbol{y}_1, \cdots, \boldsymbol{y}_M)$. Furthermore, $\boldsymbol{x}_i = (x_{i,1}, \cdots, x_{i,n})$ and $\boldsymbol{y}_j = (y_{j,1}, \cdots, y_{j,n})$ are such that

$$\frac{1}{n} \sum_{k=1}^{n} a(x_{i,k}) \le A, \qquad i = 1, \cdots, N, \tag{31a}$$

and

$$\frac{1}{n} \sum_{k=1}^{n} b(y_{j,k}) \le B, \qquad j = 1, \cdots, M. \tag{31b}$$

Observe that $\boldsymbol{S}$ does not necessarily satisfy the state constraint L. If we can show, however, that (27) implies

$$\Pr\big(\Phi(\boldsymbol{Z}) \ne (\alpha, \beta)\big) \le \lambda \tag{32}$$

then a Fano-type weak converse can be carried out.

Using (28) and a Chebyshev bound as in [8, p. 28], we can easily show that $\Pr(\ell_n(\boldsymbol{S}) > L | U = u) \le \lambda/2$ for sufficiently large n for all $u \in \mathcal{U}_n^{NM}$. It then follows that $\Pr(\ell_n(\boldsymbol{S}) > L) \le \lambda/2$ as well. Note that this bound is uniform in $\{r_k\}_{k=1}^n$ satisfying (28). By (30)

$$\begin{aligned}\Pr(\boldsymbol{Z} = \boldsymbol{z}, \alpha = i, \beta = j | U = u) \\ = \frac{1}{NM} \sum_{\boldsymbol{s} \in \mathbb{S}^n} r(\boldsymbol{s}) W^n(\boldsymbol{z}|\boldsymbol{x}_i, \boldsymbol{y}_j, \boldsymbol{s}).\end{aligned}$$

Thus

$$\begin{aligned}\Pr(\Phi(\boldsymbol{Z}) \ne (\alpha, \beta) | U = u) &= \Pr(\varphi(\boldsymbol{Z}) \ne (\alpha, \beta) | U = u) \\ &= \sum_{\boldsymbol{s} \in \mathbb{S}^n} r(\boldsymbol{s}) e(\boldsymbol{s}, u)\end{aligned}$$

and

$$\begin{aligned}\Pr\big(\Phi(\boldsymbol{Z}) \ne (\alpha, \beta)\big) &= E\big[\Pr\big(\Phi(\boldsymbol{Z}) \ne (\alpha, \beta) \mid U\big)\big] \\ &= \sum_{\boldsymbol{s} \in \mathbb{S}^n} r(\boldsymbol{s}) E[e(\boldsymbol{s}, U)] \\ &\le \max_{\boldsymbol{s} \in \mathbb{S}^n(L)} E[e(\boldsymbol{s}, U)] \\ &\quad + \Pr(\ell_n(\boldsymbol{S}) > L) \\ &\le \lambda.\end{aligned}$$

Writing $\Phi(\boldsymbol{Z}) = \big(\Phi_1(\boldsymbol{Z}), \Phi_2(\boldsymbol{Z})\big)$, we see that

$$\Pr(\Phi_1(\boldsymbol{Z}) \ne \alpha) \le \lambda \quad \text{and} \quad \Pr(\Phi_2(\boldsymbol{Z}) \ne \beta) \le \lambda.$$

Using (30) and Fano's inequality [3, p. 53], we can write

$$\begin{aligned}\log NM &= H(\alpha\beta) \\ &\le \lambda \log NM + 1 + I(\alpha\beta \wedge \Phi(\boldsymbol{Z})).\end{aligned}$$

Rearranging, we have

$$(1 - \lambda) \log NM \le I(\alpha\beta \wedge \Phi(\boldsymbol{Z})) + 1$$

and by (26)

$$R_1 + R_2 \le \frac{I(\alpha\beta \wedge \Phi(\boldsymbol{Z})) + 1}{n(1 - \lambda)} + 2\Delta R. \tag{33}$$

Now write

$$\begin{aligned}I(\alpha\beta \wedge \Phi(\boldsymbol{Z})) &= H(\alpha\beta) - H(\alpha\beta|\Phi(\boldsymbol{Z})) \\ &= H(\alpha\beta|U) - H(\alpha\beta|\Phi(\boldsymbol{Z})) \\ &\le H(\alpha\beta|U) - H(\alpha\beta|\Phi(\boldsymbol{Z}), U) \\ &= I(\alpha\beta \wedge \Phi(\boldsymbol{Z})|U) \\ &= \sum_u \Pr(U = u) I(\alpha\beta \wedge \varphi(\boldsymbol{Z})|U = u) \\ &\le \sum_u \Pr(U = u) I(\boldsymbol{XY} \wedge \boldsymbol{Z}|U = u)\end{aligned}$$

where the last step follows by the Data Processing Lemma [3, p. 55, Lemma 3.11(i)]. Equation (30) allows us to write

$$\begin{aligned}I(\alpha\beta \wedge \Phi(\boldsymbol{Z})) &\le \sum_u \Pr(U = u) \\ &\quad \cdot \Big(\sum_{k=1}^{n} I(X_k Y_k \wedge Z_k | U = u)\Big) \\ &= \sum_{k=1}^{n} I(X_k Y_k \wedge Z_k | U).\end{aligned} \tag{34}$$

Now

$$\begin{aligned}I(X_k Y_k \wedge Z_k | U) &= H(Z_k|U) - H(Z_k|X_k Y_k U) \\ &= H(Z_k|U) - H(Z_k|X_k Y_k) \\ &\le H(Z_k) - H(Z_k|X_k Y_k) \\ &= I(X_k Y_k \wedge Z_k)\end{aligned}$$

and so

$$I(\alpha\beta \wedge \Phi(\boldsymbol{Z})) \le \sum_{k=1}^{n} I(X_k Y_k \wedge Z_k) \tag{35}$$

where $I(X_k Y_k \wedge Z_k)$ is computed from the joint distribution on (x, y, s, z)

$$p_k(x) q_k(y) r_k(s) W(z|x, y, s)$$

and $p_k(x) := \Pr(X_k = x)$ and $q_k(y) := \Pr(Y_k = y)$. Note that the independence of X_k and Y_k follows from their conditional independence given $U = u$ together with the independence of F and G. Now let $\mathbb{V} = \{1, \cdots, n\}$, and let $\Pr(\hat{V} = k) = 1/n$. Then the right-hand side of (35) is equal to

$$n \cdot I(\hat{X}\hat{Y} \wedge \hat{Z}|\hat{V})$$

where

$$\hat{\mathbb{P}}_{\hat{V}\hat{X}\hat{Y}\hat{S}\hat{Z}}(k, x, y, s, z) = \frac{1}{n} \cdot p_k(x) q_k(y) r_k(s) W(z|x, y, s).$$

Furthermore, note that $E[\ell(\hat{S})] \le L - \delta$ by (28). Also

$$\begin{aligned} E[a(\hat{X})] &= \frac{1}{n}\sum_{k=1}^{n}\Big(\sum_x a(x)p_k(x)\Big) \\ &= \frac{1}{n}\sum_{k=1}^{n} E[a(X_k)] \\ &= E\Big[\frac{1}{n}\sum_{k=1}^{n} a(X_k)\Big]. \end{aligned}$$

By (30) and (31a), we have

$$\begin{aligned} E\Big[\frac{1}{n}\sum_{k=1}^{n} a(X_k)\Big|U=u\Big] &= \frac{1}{N}\sum_{i=1}^{N}\Big(\frac{1}{n}\sum_{k=1}^{n} a(x_{i,k})\Big) \\ &\le A. \end{aligned} \tag{36}$$

Thus $E[a(\hat{X})] \le A$, and therefore

$$p_k(x) = \hat{\mathbb{P}}_{\hat{X}|\hat{V}}(x|k) \in \mathcal{D}^A(\mathbb{X}|\hat{\mathbb{P}}_{\hat{V}}).$$

From (33) we can now write

$$R_1 + R_2 \le \frac{I^{L-\delta}(\hat{X}\hat{Y} \wedge \hat{Z}|\hat{V}) + n^{-1}}{1-\lambda} + 2\Delta R.$$

In a similar manner, one can obtain

$$R_1 \le \frac{I^{L-\delta}(\hat{X} \wedge \hat{Z}|\hat{Y}\hat{V}) + n^{-1}}{1-\lambda} + \Delta R$$

and

$$R_2 \le \frac{I^{L-\delta}(\hat{Y} \wedge \hat{Z}|\hat{X}\hat{V}) + n^{-1}}{1-\lambda} + \Delta R.$$

Remark: If we replace (31a) by

$$\frac{1}{N}\sum_{i=1}^{N}\Big(\frac{1}{n}\sum_{k=1}^{n} a(x_{i,k})\Big) \le A$$

(which is called a *message-average constraint*, cf. [4, p. 28]) and similarly for (31b), then (36) still holds, and the converse proof goes through unchanged.

IV. Example and Discussion

Consider a noiseless erasure channel in which $\mathbb{X} = \mathbb{Y} = \{0, \cdots, \rho-1\}$ and $\mathbb{S} = \{0,1\}$. For $x \in \mathbb{X}$, $y \in \mathbb{Y}$, and $s \in \mathbb{S}$, the channel output is given by

$$z = \begin{cases} x+y, & \text{if } xy = s = 0 \\ \rho, & \text{otherwise} \end{cases}$$

where $z \in \mathbb{Z} = \{0, \cdots, \rho\}$. In other words, if $xy = s = 0$ and $z = x + y$, then $W(z|x,y,s) = 1$. If $xy \ne 0$ or $s \ne 0$, then $W(\rho|x,y,s) = 1$. In all other cases $W(z|x,y,s) = 0$. The state constraint is $\ell(s) = s$ with $0 \le L \le 1$. No input constraints are imposed (or equivalently the constraints are inactive). To reflect this, in this section, we write $\mathcal{R}(L)$ instead of $\mathcal{R}(L, A, B)$.

It is difficult to determine $\mathcal{R}(L)$ exactly for this channel; instead we give inner and outer bounds that are tight for large ρ. For any random variable X, let $E_X = 0$, if $X = 0$ and $E_X = 1$, otherwise. If V, X, Y, S, and Z have joint distribution (5), then

$$\begin{aligned} I(X \wedge Z|YV) &= I(E_X E_Y X \wedge Z|YV) \\ &= I(E_X E_Y \wedge Z|YV) \\ &\quad + I(X \wedge Z|YVE_XE_Y). \end{aligned}$$

Hence

$$\begin{aligned} I^L(X \wedge Z|YVE_XE_Y) &\le I^L(X \wedge Z|YV) \\ &\le 1 + I^L(X \wedge Z|YVE_XE_Y). \end{aligned}$$

Similarly, it can be shown that

$$\begin{aligned} I^L(Y \wedge Z|XVE_XE_Y) &\le I^L(Y \wedge Z|XV) \\ &\le 1 + I^L(Y \wedge Z|XVE_XE_Y), \\ I^L(XY \wedge Z|VE_XE_Y) &\le I^L(XY \wedge Z|V) \\ &\le 2 + I^L(XY \wedge Z|VE_XE_Y). \end{aligned}$$

Recalling (9), we conclude that

$$\mathcal{R}_I(L) \subset \mathcal{R}(L) \subset \mathcal{R}_O(L) \tag{37}$$

where

$$\mathcal{R}_I(L) := \bigcup \mathcal{R}_I(L, \gamma, p, q), \tag{38}$$

$$\begin{aligned} \mathcal{R}_I(L,\gamma,p,q) := \Big\{&(R_1,R_2): \\ &0 \le R_1 \le I^L(X \wedge Z|YVE_XE_Y), \\ &0 \le R_2 \le I^L(Y \wedge Z|XVE_XE_Y), \\ &R_1 + R_2 \le I^L(XY \wedge Z|VE_XE_Y)\Big\} \end{aligned}$$

$$\begin{aligned} \mathcal{R}_O(L) := \Big\{&(R_1,R_2): R_1 \ge 0, R_2 \ge 0, \\ &(|R_1-1|^+, |R_2-1|^+) \in \mathcal{R}_I(L)\Big\} \end{aligned}$$

and the union is over all sets $\mathbb{V}$, $|\mathbb{V}| < \infty$, all $\gamma \in \mathcal{D}(\mathbb{V})$, $p \in \mathcal{D}(\mathbb{X}|\mathbb{V})$, and $q \in \mathcal{D}(\mathbb{Y}|\mathbb{V})$.

In Appendix I, we show that the inner bound is given by

$$\mathcal{R}_I(L) = \begin{cases} \mathcal{R}_1(L), & \text{if } L \ge 1/2 \\ \mathcal{R}_1(L) \cup \mathcal{R}_2(L) \cup \mathcal{R}_3(L), & \text{if } L < 1/2 \end{cases} \tag{39}$$

where

$$\begin{aligned} \mathcal{R}_1(L) := \Big\{&(R_1,R_2): R_1 \ge 0, R_2 \ge 0, \\ &(\sqrt{R_1} + \sqrt{R_2})^2 \le (1-L)\log(\rho-1)\Big\} \end{aligned}$$

and where, letting $\overline{\lambda} = 1 - \lambda$ and $\overline{L} = 1 - L$

$$\begin{aligned} \mathcal{R}_2(L) := \Big\{&(R_1,R_2): 0 \le R_1 \le (\overline{\lambda} - L)s\log(\rho-1), \\ &0 \le R_2 \le \Big[(\lambda - L) + \overline{\lambda}(1-\sqrt{s})^2\Big]\log(\rho-1), \\ &\text{for some } L \le \lambda \le \overline{L}, 0 \le s \le 1\Big\} \end{aligned}$$

$$\mathcal{R}_3(L) := \Big\{(R_1,R_2): (R_2,R_1) \in \mathcal{R}_2(L)\Big\}.$$

The bound (37) reveals several interesting features of $\mathcal{R}(L)$ which are unique to the MAVC. As illustrated in Fig. 1, the capacity region is not convex in general. To our knowledge, this is the first example of a synchronous multiple-access

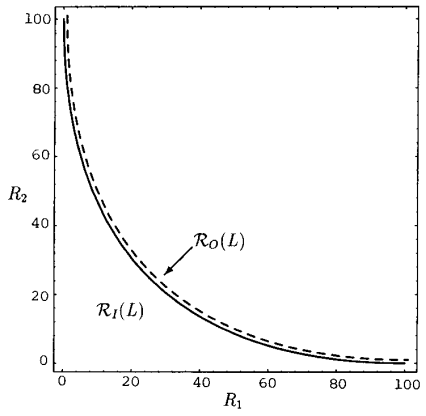

Fig. 1. Inner and outer bounds for $L = 0.5$ and $\log(\rho - 1) = 200$.

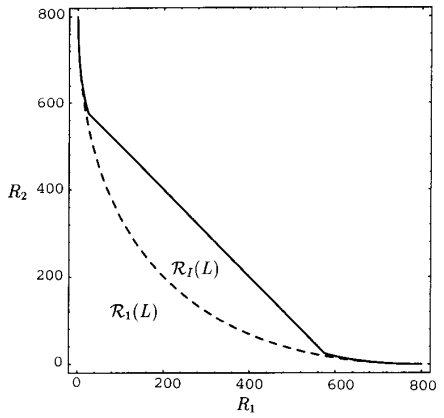

Fig. 2. $\mathcal{R}^1(L)$ and $\mathcal{R}^I(L)$ for $L = 0.2$ and $\log(\rho - 1) = 1000$.

channel with this property. The lack of convexity is due to the state constraint; when this constraint is inactive, $\mathcal{R}(L)$ is always convex.

Perhaps the two best-known prior examples of nonconvexity are the capacity region of the frame-asynchronous multiple-access channel and the capacity region per unit cost of the multiple-access channel subject to input constraints. The first example was obtained independently by Hui and Humblet [10] and Poltyrev [12]. This capacity region is not generally convex because the lack of a common time reference between asynchronous encoders precludes time-sharing. In sharp contrast, for certain MAVC's, time-sharing with the auxiliary variable V enlarges the set of achievable rate pairs, but fails to achieve the full convex hull. The second example, the capacity region *per unit cost* obtained by Verdú [13, eq. (19)], is not the same as the capacity region itself of the conventional MAC subject to input constraints, which is always convex.

To show how time-sharing with the auxiliary variable V can enlarge the set of achievable rate pairs but not achieve the full convex hull, let $\mathcal{R}_*(L)$ consist of those rate pairs that are achievable without time-sharing; i.e., the right side of (9) with the union restricted to $\mathbf{V}$ such that $|\mathbf{V}| = 1$. From Appendix I, it is straightforward to show, for all $0 < L \leq 1$, that $\mathcal{R}_1(L)$ similarly equals the right side of (38) with the union restricted to $\mathbf{V}$ such that $|\mathbf{V}| = 1$. As in (37), it is then easily seen that

$$\mathcal{R}_1(L) \subset \mathcal{R}_*(L) \subset \mathcal{R}_{1,O}(L), \tag{40}$$

where $\mathcal{R}_{1,O}(L)$ is defined similar to $\mathcal{R}_O(L)$ except that $\mathcal{R}_I(L)$ is replaced by $\mathcal{R}_1(L)$. In Fig. 2, $\mathcal{R}_1(L)$ and $\mathcal{R}_I(L)$ are depicted for $\log(\rho - 1) = 1000$. For this value of ρ, the three regions in (37) coincide to within the thickness of the plot lines, as do the three regions in (40). From Fig. 2, we see that

$$\mathcal{R}_*(L) \subset \mathcal{R}(L) \subset \text{convex hull of } \mathcal{R}_*(L)$$

where the inclusions are strict. Thus $\mathcal{R}(L)$, although nonconvex, contains some rate pairs which are achievable only through time-sharing with the auxiliary variable V.

Appendix I
Proof of (39)

Let $\mathbb{V}$, $|\mathbb{V}| < \infty$, $\gamma \in \mathcal{D}(\mathbb{V})$, $p \in \mathcal{D}(\mathbb{X}|\mathbb{V})$, $q \in \mathcal{D}(\mathbb{Y}|\mathbb{V})$ be arbitrary. Our approach is to derive an outer bound to $\mathcal{R}_I(L, \gamma, p, q)$ and then to show this bound is achieved by a particular choice of γ, p, and q.

Fix $r \in \mathcal{D}^L(\mathbb{S}|\gamma)$ and observe for this channel that

$$\begin{aligned} I(XY \wedge Z|VE_XE_Y) &= I(X \wedge Z|YVE_XE_Y) \\ &\quad + I(Y \wedge Z|XVE_XE_Y). \end{aligned}$$

Taking the infimum over $r \in \mathcal{D}^L(\mathbb{S}|\gamma)$, we obtain

$$\begin{aligned} I^L(XY \wedge Z|VE_XE_Y) &\geq I^L(X \wedge Z|YVE_XE_Y) \\ &\quad + I^L(Y \wedge Z|XVE_XE_Y). \end{aligned}$$

The $R_1 + R_2$ bound in $\mathcal{R}_I(L, \gamma, p, q)$ is therefore redundant and can be omitted.

Next we write

$$\begin{aligned} &I(X \wedge Z|YVE_XE_Y) \\ &\quad = \sum_{v \in \mathbb{V}} \gamma(v) q(0|v)[1 - p(0|v)] \\ &\qquad \cdot I(X \wedge Z|Y, V = v, E_X = 1, E_Y = 0) \\ &\quad \leq \sum_{v \in \mathbb{V}} \gamma(v) q(0|v)[1 - p(0|v)] r(0|v) \log(\rho - 1). \end{aligned} \tag{41}$$

The inequality follows by recognizing that, given $V = v, E_X = 1$, and $E_Y = 0$, the random variable X, taking values in $\{1, \cdots, \rho - 1\}$, is connected to Z by a single-user erasure channel with probability of erasure $r(1|v)$. Hence, the conditional mutual information on the right is at most $r(0|v) \log(\rho - 1)$.

Similarly, it can be shown that

$$\begin{aligned} &I(Y \wedge Z|XVE_XE_Y) \\ &\quad \leq \sum_{v \in \mathbb{V}} \gamma(v) p(0|v)[1 - q(0|v)] r(0|v) \log(\rho - 1). \end{aligned} \tag{42}$$

Observe that if $a, b \in [0, 1]$ and $t := ab$, then $(1-a)(1-b) \leq (1 - \sqrt{t})^2$, with equality if and only if $a = b = \sqrt{t}$. So, if we set

$$t(v) := q(0|v)[1 - p(0|v)], \tag{43}$$

then (41) and (42) imply

$$I(X \wedge Z|YVE_XE_Y) \le \sum_{v\in\mathbf{V}} \gamma(v)t(v) \cdot r(0|v)\log(\rho-1), \quad (44)$$

$$I(Y \wedge Z|XVE_XE_Y) \le \sum_{v\in\mathbf{V}} \gamma(v)\left(1-\sqrt{t(v)}\right)^2 \cdot r(0|v)\log(\rho-1). \quad (45)$$

Let $T_1 < \cdots < T_K$ denote the distinct values that $t(V)$ takes with positive probability. For $L > 0$, we can always choose $s = T_i$ and $t = T_j$ for some i and j and such that

$$\Pr\{t(V) < s\} < \overline{L} \le \Pr\{t(V) \le s\}$$
$$\Pr\{t(V) > t\} < \overline{L} \le \Pr\{t(V) \ge t\}.$$

Here and throughout this Appendix, for any $x \in [0,1]$, we define $\overline{x} := 1 - x$. With this choice of t and s, it follows that $s \le t$ when $L \ge 1/2$, and $s \ge t$ when $0 < L < 1/2$. Moreover, there exist $a, b \in [0,1]$ such that

$$\Pr\{t(V) < s\} + a\Pr\{t(V) = s\} = \overline{L},$$
$$\Pr\{t(V) > t\} + b\Pr\{t(V) = t\} = \overline{L}.$$

If we set

$$r_t(0|v) := \begin{cases} 0, & t(v) < t \\ b, & t(v) = t \\ 1, & t(v) > t \end{cases}$$

and

$$r_s(0|v) := \begin{cases} 0, & t(v) > s \\ a, & t(v) = s \\ 1, & t(v) < s \end{cases}$$

then both yield $E[\ell(S)] = L$. Since any $r \in \mathcal{D}^L(\mathbb{S}|\gamma)$ provides an upper bound on the infimum, we have

$$I^L(X \wedge Z|YVE_XE_Y) \le \sum_{v\in\mathbf{V}} \gamma(v)r_s(0|v)t(v)\log(\rho-1)$$

$$I^L(Y \wedge Z|XVE_XE_Y) \le \sum_{v\in\mathbf{V}} \gamma(v)r_t(0|v)\left(1-\sqrt{t(v)}\right)^2 \cdot \log(\rho-1). \quad (46)$$

We now restrict attention to the case $L \ge 1/2$. In this case, $t \ge s$ and so

$$\sum_{v\in\mathbf{V}} \gamma(v)r_s(0|v)t(v) \le \Big[\Pr\{t(V) < s\} + a\Pr\{t(V) = s\}\Big]s \le (1-L)t.$$

Similarly, we obtain

$$\sum_{v\in\mathbf{V}} \gamma(v)r_t(0|v)\left(1-\sqrt{t(v)}\right)^2 \le (1-L)(1-\sqrt{t})^2.$$

The four preceding inequalities imply $\mathcal{R}_I(L,\gamma,p,q) \subseteq \mathcal{R}_1(L;t)$, where

$$\mathcal{R}_1(L;t) := \Big\{(R_1,R_2) : 0 \le R_1 \le (1-L)t\log(\rho-1), \; 0 \le R_2 \le (1-L)(1-\sqrt{t})^2\log(\rho-1)\Big\}.$$

From (38), it then follows that $\mathcal{R}_I(L) \subseteq \bigcup_{t\in[0,1]} \mathcal{R}_1(L;t)$. Conversely, for each $0 \le t \le 1$, it is easy to show that the choice $\mathbf{V} = \{t\}$, $\tilde{\gamma}(t) = 1$

$$\tilde{q}(0|v) = 1 - \tilde{p}(0|v) = \sqrt{v}$$
$$\tilde{q}(y|v) = [1-\sqrt{v}]/(\rho-1), \quad y \ne 0$$
$$\tilde{p}(x|v) = \sqrt{v}/(\rho-1), \quad x \ne 0 \quad (47)$$

achieves $\mathcal{R}_I(L,\tilde{\gamma},\tilde{p},\tilde{q}) = \mathcal{R}_1(L;t)$; hence,

$$\mathcal{R}_I(L) = \bigcup_{0\le t\le 1} \mathcal{R}_1(L;t) = \mathcal{R}_1(L).$$

This completes the proof of (39) for $L \ge 1/2$.

We now treat the case $0 < L < 1/2$. To this end, define the probability mass functions

$$\gamma_1(v) := \frac{r_t(1|v)}{L}\gamma(v)$$
$$\gamma_2(v) := \frac{r_s(0|v) + r_t(0|v) - 1}{1-2L}\gamma(v)$$
$$\gamma_3(v) := \frac{r_s(1|v)}{L}\gamma(v).$$

Here we have

$$\sum_{v\in\mathbf{V}} \gamma(v)r_s(0|v)t(v) = L\sum_{v\in\mathbf{V}} \gamma_1(v)t(v) + (1-2L)\sum_{v\in\mathbf{V}} \gamma_2(v)t(v) \le Lt + (1-2L)\sum_{v\in\mathbf{V}} \gamma_2(v)t(v).$$

Using γ_3 instead of γ_1, we similarly obtain

$$\sum_{v\in\mathbf{V}} \gamma(v)r_t(0|v)\left(1-\sqrt{t(v)}\right)^2 \le (1-2L)\sum_{v\in\mathbf{V}} \gamma_2(v)\left(1-\sqrt{t(v)}\right)^2 + L(1-\sqrt{s})^2.$$

Since $\gamma_2(v)$ is nonzero only for $t \le t(v) \le s$, we can write

$$t(v) = \alpha(v)t + \overline{\alpha}(v)s$$

for some $0 \le \alpha(v) \le 1$. It follows that

$$\sum_{v\in\mathbf{V}} \gamma_2(v)t(v) = \alpha t + \overline{\alpha}s$$

where

$$\alpha := \sum_v \gamma_2(v)\alpha(v).$$

Since $(1-\sqrt{x})^2$ is convex on $[t,s]$, we have

$$\left(1-\sqrt{t(v)}\right)^2 \le \alpha(v)(1-\sqrt{t})^2 + \overline{\alpha}(v)(1-\sqrt{s})^2$$

and thus

$$\sum_{v\in\mathbf{V}}\gamma_2(v)\left(1-\sqrt{t(v)}\right)^2 \le \alpha(1-\sqrt{t})^2+\overline{\alpha}(1-\sqrt{s})^2.$$

Substituting into (46), we obtain

$$\begin{aligned} I^L(X\wedge Z|YVE_XE_Y) &\le [Lt+(1-2L)(\alpha t+\overline{\alpha}s)]\\ &\quad\cdot\log(\rho-1)\\ I^L(Y\wedge Z|XVE_XE_Y) &\le \{(1-2L)[\alpha(1-\sqrt{t})^2\\ &\quad+\overline{\alpha}(1-\sqrt{s})^2]\\ &\quad+L(1-\sqrt{s})^2\}\cdot\log(\rho-1). \end{aligned}$$

Upon setting $\lambda := (1-2L)\alpha + L$, we conclude that $\mathcal{R}_I(L,\gamma,p,q) \subseteq \mathcal{R}_2(L;\lambda,t,s)$ where

$$\begin{aligned} &\mathcal{R}_2(L;\lambda,t,s)\\ &:= \Big\{(R_1,R_2): 0\le R_1\le[\lambda t+(\overline{\lambda}-L)s]\log(\rho-1),\\ &\quad 0\le R_2\le[(\lambda-L)(1-\sqrt{t})^2+\overline{\lambda}(1-\sqrt{s})^2]\\ &\quad\cdot\log(\rho-1)\Big\} \end{aligned}$$

$L\le\lambda\le\overline{L}$, and $0\le t\le s\le 1$. Conversely, for each choice of $L\le\lambda\le\overline{L}$ and $0\le t\le s\le 1$, it is straightforward to verify that $\mathcal{R}_I(L,\tilde{\gamma},\tilde{p},\tilde{q}) = \mathcal{R}_2(L;\lambda,t,s)$, where $\mathbf{V}=\{t,s\}$, $\tilde{\gamma}(t)=1-\tilde{\gamma}(s)=\lambda$, and $\tilde{q}$ and $\tilde{p}$ are as in (47). From (38), we conclude that

$$\mathcal{R}_I(L) = \bigcup_{L\le\lambda\le\overline{L},\,0\le t\le s\le 1}\mathcal{R}_2(L;\lambda,t,s). \tag{48}$$

To complete the proof, it only remains to show that the union reduces to the right side of (39) when $0<L<1/2$. To this end, observe that for fixed λ, $(\lambda-L)(1-\sqrt{t})^2+\overline{\lambda}(1-\sqrt{s})^2$ is convex in the region $0\le t\le s\le 1$, and that $\lambda t+(\overline{\lambda}-L)s = R_1/\log(\rho-1)$ is a line segment in this region. The maximum on this line is therefore achieved on the boundary, so that either $s=t$, $t=0$, or $s=1$. Hence

$$\mathcal{R}_I(L) = \mathcal{R}_1(L)\cup\mathcal{R}_2(L)\cup\mathcal{R}_3(L)$$

where $\mathcal{R}_1(L)$, $\mathcal{R}_2(L)$, and $\mathcal{R}_3(L)$ are obtained by restricting the right-side of (48) to $s=t$, $t=0$, and $s=1$, respectively. $\mathcal{R}_1(L)$ and $\mathcal{R}_2(L)$ are then as given in (39). After the change of variable $\hat{t}=(1-\sqrt{t})^2$ and $\hat{\lambda}=\overline{\lambda}$, $\mathcal{R}_3(L)$ also reduces to the corresponding region in (39).

Appendix II
Proof of the Approximation Theorem

The proof is an easy consequence the following two lemmas.

Recall our assumption that $\min_x a(x) = \min_y b(y) = 0$. Then Lemma 1 is straightforward. In the lemma, d is the variational distance [3, p. 58].

Lemma 1: Suppose $|\mathbf{V}|<\infty$ and that $p\in\mathcal{D}(\mathbb{X}|\mathbf{V})$ and $q\in\mathcal{D}(\mathbb{Y}|\mathbf{V})$ are given. Fix any positive integer n. Then for each $v\in\mathbf{V}$, there exist types $p'(\cdot|v)\in\mathcal{D}_n(\mathbb{X})$ and $q'(\cdot|v)\in\mathcal{D}_n(\mathbb{Y})$ such that

$$\begin{aligned} d\big(p(\cdot|v),p'(\cdot|v)\big) &\le 2|\mathbb{X}|/n\\ d\big(q(\cdot|v),q'(\cdot|v)\big) &\le 2|\mathbb{Y}|/n \end{aligned} \tag{49}$$

and such that both

$$\sum_x a(x)p'(x|v) \le \sum_x a(x)p(x|v)$$

and

$$\sum_y b(y)q'(y|v) \le \sum_y b(y)q(y|v).$$

There are two immediate implications of Lemma 1. First, the sum of the two upper bounds in (49) is equal to θ_n, and second, for any $\gamma\in\mathcal{D}(\mathbf{V})$, if $p\in\mathcal{D}^A(\mathbb{X}|\gamma)$ and $q\in\mathcal{D}^B(\mathbb{Y}|\gamma)$, then $p'\in\mathcal{D}^A(\mathbb{X}|\gamma)$ and $q'\in\mathcal{D}^B(\mathbb{Y}|\gamma)$.

Lemma 2: For $|\mathbf{V}|<\infty$, $\gamma\in\mathcal{D}(\mathbf{V})$, $p\in\mathcal{D}^A(\mathbb{X}|\gamma)$, and $q\in\mathcal{D}^B(\mathbb{Y}|\gamma)$, let $\mathbb{P}_{VXYSZ}$ be as in (5). For $n\ge 4(|\mathbb{X}|+|\mathbb{Y}|)$, let p' and q' be as in Lemma 1, and set

$$\begin{aligned} &\mathbb{P}_{VX'Y'SZ'}(v,x,y,s,z)\\ &\quad:=\gamma(v)p'(x|v)q'(y|v)r(s|v)W(z|x,y,s). \end{aligned} \tag{50}$$

Then

$$\begin{aligned} I^L(X\wedge Z|YV) &\le I^L(X'\wedge Z'|Y'V)+\mu_n\\ I^L(Y\wedge Z|XV) &\le I^L(Y'\wedge Z'|X'V)+\mu_n\\ I^L(XY\wedge Z|V) &\le I^L(X'Y'\wedge Z'|V)+\mu_n. \end{aligned}$$

Proof: Observe that $\mathbb{P}_{Z|XYV}=\mathbb{P}_{Z'|X'Y'V}$. Then using (49) and (50)

$$\big|H(Z|XYV)-H(Z'|X'Y'V)\big| \le \theta_n\log|\mathbb{Z}|.$$

Next, note that (49) implies that for each $v\in\mathbf{V}$

$$d(\mathbb{P}_{XYSZ|V}(\cdot|v),\mathbb{P}_{X'Y'SZ'|V}(\cdot|v)) \le \theta_n.$$

Since for $n\ge 4(|\mathbb{X}|+|\mathbb{Y}|)$, $\theta_n\le 1/2$, we can follow [3, p. 33, Lemma 2.7] and show that

$$\begin{aligned} \big|H(Z|V)-H(Z'|V)\big| &\le \theta_n\log(|\mathbb{Z}|/\theta_n)\\ \big|H(Y|V)-H(Y'|V)\big| &\le \theta_n\log(|\mathbb{Y}|/\theta_n)\\ \big|H(X|V)-H(X'|V)\big| &\le \theta_n\log(|\mathbb{X}|/\theta_n)\\ \big|H(YZ|V)-H(Y'Z'|V)\big| &\le \theta_n\log(|\mathbb{Y}||\mathbb{Z}|/\theta_n)\\ \big|H(XZ|V)-H(X'Z'|V)\big| &\le \theta_n\log(|\mathbb{X}||\mathbb{Z}|/\theta_n). \end{aligned}$$

Then writing, for example, $I(X\wedge Z|YV) = H(Z|YV) - H(Z|XYV) = H(YZ|V) - H(Y|V) - H(Z|XYV)$, the lemma follows easily. ∎

When p' and q' are as in Lemma 1, note that as v runs though $\mathbf{V}$, there are only $|\mathcal{D}_n(\mathbb{X})|$ possibilities for $p'(\cdot|v)$ and $|\mathcal{D}_n(\mathbb{Y})|$ possibilities for $q'(\cdot|v)$. If we let $m := |\mathcal{D}_n(\mathbb{X})||\mathcal{D}_n(\mathbb{Y})|$, then by [3, p. 39], m is equal to the right-hand side of (12). Hence, in an obvious way, we can partition $\mathbf{V}$ into m equivalence classes, say $\mathbf{V}_1,\cdots,\mathbf{V}_m$.

Recalling (7), we set $I_m^L(Y' \wedge Z'|X'V) := \inf I(Y' \wedge Z'|X'V)$, where the infimum is restricted to those $r \in \mathcal{D}^L(\mathbb{S}|\gamma)$ such that $r(\cdot|v)$ is a constant of v on each $\mathbb{V}_i$. Note that $I^L(Y' \wedge Z'|X'V) \leq I_m^L(Y' \wedge Z'|X'V)$. If we now return to (50), we see that $\mathbb{P}_{X'Y'SZ'|V}(\cdot|v)$ is constant for v on each equivalence class. Hence, we may identify the $\tilde{\mathbb{V}}$ in the theorem with the collection of equivalence classes of $\mathbb{V}$, and in an obvious way construct

$$\mathbb{P}_{\tilde{V}\tilde{X}\tilde{Y}\tilde{S}\tilde{Z}}(\tilde{v}, x, y, s, z) := \tilde{\gamma}(\tilde{v})\tilde{p}(x|\tilde{v})\tilde{q}(y|\tilde{v})\tilde{r}(s|\tilde{v})W(z|x, y, s)$$

where if $\tilde{v} = \mathbb{V}_i$

$$\begin{aligned}\tilde{\gamma}(\mathbb{V}_i) &= \sum_{v \in \mathbb{V}_i} \gamma(v) \\ \tilde{p}(x|\mathbb{V}_i) &= p'(x|v_i) \\ \tilde{q}(y|\mathbb{V}_i) &= q'(y|v_i) \\ \tilde{r}(s|\mathbb{V}_i) &= r(s|v_i)\end{aligned}$$

and v_i is any element of the equivalence class $\mathbb{V}_i$. Furthermore, it is not hard to see that $I_m^L(Y' \wedge Z'|X'V) = I^L(\tilde{Y} \wedge \tilde{Z}|\tilde{X}\tilde{V})$. Thus for any $\mathcal{R}(L, \gamma, p, q)$ on the right-hand side of (9), we have

$$\begin{aligned}\mathcal{R}(L, \gamma, p, q) &\subset \mathcal{R}(L, \gamma, p', q') + \mu_n \mathbf{U} \\ &\subset \mathcal{R}(L, \tilde{\gamma}, \tilde{p}, \tilde{q}) + \mu_n \mathbf{U}\end{aligned}$$

and then (13) follows.

REFERENCES

[1] R. Ahlswede, "Elimination of correlation in random codes for arbitrarily varying channels," *Z. Wahrscheinlichkeitstheorie verw. Geb.*, vol. 44, pp. 159–175, 1978.

[2] I. Csiszár, "Arbitrarily varying channels with general alphabets and states," *IEEE Trans. Inform. Theory*, vol. 38, pp. 1725–1742, Nov. 1992.

[3] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic Press, 1981.

[4] I. Csiszár and P. Narayan, "Arbitrarily varying channels with constrained inputs and states," *IEEE Trans. Inform. Theory*, vol. 34, pp. 27–34, Jan. 1988.

[5] ——,"The capacity of the arbitrarily varying channel revisited: Positivity, constraints," *IEEE Trans. Inform. Theory*, vol. 34, pp. 181–193, Mar. 1988.

[6] J. A. Gubner, "On the deterministic-code capacity of the multiple-access arbitrarily varying channel," *IEEE Trans. Inform. Theory*, vol. 36, pp. 262–275, Mar. 1990.

[7] ——, "Random coding for the constrained multiple-access arbitrarily varying channel," *Proc. 28th Annual Allerton Conf. Commun. Contr. Comput.* (University of Illinois, Urbana, Oct. 1990), pp. 684–690.

[8] ——, "State constraints for the multiple-access arbitrarily varying channel," *IEEE Trans. Inform. Theory*, vol. 37, pp. 27–35, Jan. 1991.

[9] ——, "On the capacity region of the discrete additive multiple-access arbitrarily varying channel," *IEEE Trans. Inform. Theory*, vol. 38, pp. 1344–1347, July 1992.

[10] J. Y. N. Hui and P. A. Humblet, "The capacity region of the totally asynchronous multiple-access channel," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 207–216, Mar. 1985.

[11] J.-H. Jahn, "Coding of arbitrarily varying multiuser channels," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 212–226, Mar. 1981.

[12] G. S. Poltyrev, "Coding in an asynchronous multiple-access channel," *Prob. Inform. Trans.*, pp. 12–21, July–Sept. 1983.

[13] S. Verdú, "On channel capacity per unit cost," *IEEE Trans. Inform. Theory*, vol. 36, pp. 1019–1030, Sept. 1990.