List-Decoding for the Arbitrarily Varying Channel Under State Constraints

Anand D. Sarwate, *Member, IEEE*, and Michael Gastpar, *Member, IEEE*

***Abstract*—List-decoding for arbitrarily varying channels (AVCs) under state constraints is investigated. It is shown that rates within ϵ of the randomized coding capacity of AVCs with input-dependent state can be achieved under maximal error with list-decoding using lists of size $O(1/\epsilon)$. Under the average error criterion, an achievable rate and converse bound are given for lists of size L. These bounds are based on two different notions of symmetrizability and do not coincide in general. An example is given which shows that for list size L, the capacity may be positive but strictly smaller than the randomized coding capacity, in contrast to the situation without constraints.**

***Index Terms*—Arbitrarily varying channels (AVCs), list-decoding.**

I. Introduction

The arbitrarily varying channel (AVC) is a model for communication subject to time-varying interference [5]. The time variation is captured by a channel state parameter and coding schemes for these channels are required to have small probability of error for all channel state sequences. In an AVC, the channel state is said to be controlled by a *jammer* who wishes to foil the communication between the encoder and decoder. More details can be found in the survey paper by Lapidoth and Narayan [17].

This paper addresses the problem of list-decoding in an AVC when the state sequence is constrained. The constraint comes by imposing a per-letter cost $l(\cdot)$ on the state sequence and requiring the cost of the state sequence chosen by the jammer for n channel uses to be less than a total budget Λn. The coding schemes in this paper are deterministic; common randomness between the encoder and decoder is not allowed. We consider both the maximal and average error criterion. Under the maximal error criterion, the capacity can be smaller than under the average error criterion. In both cases, we will compare our achievable rates to the capacities for randomized coding.

In list-decoding, the decoder is allowed to output a list of L messages and an error is declared only if the list does not contain the transmitted message. For AVCs without constraints, list-decoding capacities have been investigated under both maximal and average error. For maximal error, Ahlswede [2], [4] found a quantity C_{dep} such that a rate $C_{\mathrm{dep}} - \epsilon$ is achievable with lists of size $O(1/\epsilon)$. We extend this result to the situation with cost constraints and define a quantity $C_{\mathrm{dep}}(\Lambda)$ such that a rate $C_{\mathrm{dep}}(\Lambda) - \epsilon$ is achievable under list-decoding with list size $O(1/\epsilon)$.

The average error list-L capacity $\bar{C}_L$ without constraints was found independently by Blinovsky and colleagues [6], [7] and Hughes [15]. These authors defined the symmetrizability L_{sym} of an AVC and showed that there is a constant list size L_{sym} so that for $L \leq L_{\mathrm{sym}}$, the list-L capacity is 0, and for $L > L_{\mathrm{sym}}$, the list-L capacity is equal to the randomized coding capacity C_r. The number L_{sym} is called the *symmetrizability*. The adversary can cause L_{sym} "degrees" of symmetrizability, so list-decoding requires a list size greater than L_{sym} to guarantee that the correct message is in the list with high probability.

The main result of this paper is that list-decoding under average error is qualitatively different when the state is constrained. The degree to which the jammer can symmetrize the channel depends on the input distribution P and the cost constraint Λ. We define two kinds of symmetrizability, weak and strong, for list-decoding under state constraints. For list sizes L larger than the weak symmetrizability $\tilde{L}_{\mathrm{sym}}(P, \Lambda)$, we show that the coding strategy of Hughes [15], which uses a codebook of fixed type P, yields an achievable rate for the channel. We also prove an outer bound for this channel in terms of a quantity, we call the strong symmetrizability $L_{\mathrm{sym}}(P, \Lambda)$. We construct a jamming strategy that gives a nonvanishing probability of error for codes of type P such that $L \leq L_{\mathrm{sym}}(P, \Lambda)$.

In many cases, $L_{\mathrm{sym}}(P, \Lambda) < \tilde{L}_{\mathrm{sym}}(P, \Lambda)$, which gives a gap between our achievable region and converse. Closing this gap seems nontrivial; we conjecture that the converse can be tightened. However, our results do imply a significant difference between the constrained and unconstrained setting. Without constraints, the list-L capacity $\bar{C}_L$ is either 0 or equal to the randomized coding capacity C_r. We show via a simple example that under cost constraints the list-L capacity $\bar{C}_L(\Lambda)$ may be positive but strictly smaller than the randomized coding capacity $C_r(\Lambda)$. This parallels the result obtained in [12] for list size 1.

II. Definitions

We will use calligraphic type for sets. For an integer M, let $[M] = \{1, 2, \ldots, M\}$. Generally speaking, lower case will

Manuscript received October 06, 2009; revised September 29, 2010; accepted September 04, 2011. Date of current version February 29, 2012. The work of A. D. Sarwate and M. Gastpar was supported in part by the National Science Foundation under Award CCF-0347298. The work of A. D. Sarwate was also supported by the California Institute for Telecommunications and Information Technology (CALIT2) at UC San Diego.

A. D. Sarwate was with the Information Theory and Applications Center, University of California, San Diego, CA 92093-0447 USA. He is now with the Toyota Technological Institute at Chicago, Chicago, IL 60637 USA (e-mail: asarwate@ttic.edu).

M. Gastpar is with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720 USA, and also with the School of Computer and Communication Sciences, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland (e-mail: michael.gastpar@epfl.ch).

Communicated by H. Yamamoto, Associate Editor for Shannon Theory.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2011.2178153

refer to nonrandom quantities and capital letters will refer to random variables. Boldface is used for vectors. Thus, $\mathbf{X}$ is a vector-valued random variable, $\mathbf{x}$ is a fixed vector, and x_i is the ith element of $\mathbf{x}$. For sets $\mathcal{X}$ and $\mathcal{Y}$, the set $\mathcal{P}(\mathcal{X})$ is the set of probability distributions on $\mathcal{X}$. We denote by $\mathcal{P}_n(\mathcal{X})$ the set of all distributions such that $nP(x)$ is an integer for all $x \in \mathcal{X}$, and $\mathcal{P}(\mathcal{Y}|\mathcal{X})$ is the set of all conditional distributions on $\mathcal{Y}$ conditioned on $\mathcal{X}$. For random variables (X, Y) with joint distribution P_{XY}, we will write P_X and P_Y for the marginal distributions and $P_{X|Y}$ for the conditional distribution of X given Y. For a joint distribution $\bar{P} \in \mathcal{P}(\mathcal{X}^m)$, we will denote by P_i the ith marginal of $\bar{P}$. The function $d_{\max}(P, Q)$ will denote the maximum deviation (ℓ_∞ distance) between two probability distributions P and Q.

A. Channel Model and Codes

An AVC is a collection $\mathcal{W} = \{W(\cdot|\cdot, s) : s \in \mathcal{S}\}$ of channels from an input alphabet $\mathcal{X}$ to an output alphabet $\mathcal{Y}$ parameterized by a state s from an alphabet $\mathcal{S}$. In this paper, we assume all alphabets are finite. If $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, $\mathbf{y} = (y_1, y_2, \ldots, y_n)$, and $\mathbf{s} = (s_1, s_2, \ldots, s_n)$ are length n vectors, the probability of $\mathbf{y}$ given $\mathbf{x}$ and $\mathbf{s}$ is given by

$$W(\mathbf{y}|\mathbf{x}, \mathbf{s}) = \prod_{i=1}^{n} W(y_i|x_i, s_i)$$

We are interested in the case where there is a bounded cost function $l : \mathcal{S} \to \mathbb{R}^+$ on the jammer. The cost of an n-tuple is

$$l(\mathbf{s}) = \sum_{k=1}^{n} l(s_k)$$

The state obeys a state constraint Λ if

$$l(\mathbf{s}) \leq n\Lambda \qquad a.s$$

Let $\mathcal{S}^n(\Lambda) = \{\mathbf{s} \in \mathcal{S}^n : l(\mathbf{s}) \leq n\Lambda\}$ be the set of all length-n state sequences satisfying the constraint Λ.

An (n, N, L) deterministic list code $\mathcal{C}$ for the AVC is a pair of maps (ψ, ϕ) where the encoding function is $\psi : [N] \to \mathcal{X}^n$ and the decoding function is $\phi : \mathcal{Y}^n \to \{B : B \subset [N], |B| \leq L\}$. The*rate* of the code is $R = n^{-1} \log(N/L)$. The *codebook* of $\mathcal{C}$ is the set of vectors $\{\mathbf{x}_i : i \in [N]\}$, where $\mathbf{x}_i = \psi(i)$. The decoding region for message i is $D_i = \{\mathbf{y} : i \in \phi(\mathbf{y})\}$. We will often specify a code by the pairs $\{(\mathbf{x}_i, D_i) : i =\in [N]\}$, with the encoder and decoder implicitly defined.

The *maximal* and *average* error probabilities ε_L and $\bar{\varepsilon}_L$ are given by

$$\varepsilon_L = \max_{\mathbf{s} \in \mathcal{S}^n(\Lambda)} \max_i (1 - W(D_i|\mathbf{x}_i, \mathbf{s})) \tag{1}$$

$$\bar{\varepsilon}_L = \max_{\mathbf{s} \in \mathcal{S}^n(\Lambda)} \frac{1}{N} \sum_{i=1}^{N} (1 - W(D_i|\mathbf{x}_i, \mathbf{s})) \tag{2}$$

A rate R is called achievable under maximal (average) list-decoding with list size L if for any $\epsilon > 0$, there exists a sequence of (n, N, L) list codes of rate at least $R - \epsilon$ whose maximal (average) error converges to 0. The list-L capacity is the supremum of achievable rates. We denote the list-L capacities under maximal and average error by $C_L(\Lambda)$ and $\bar{C}_L(\Lambda)$, respectively. We emphasize that in this paper, we consider deterministic codes.

B. Symmetrizability and Information Quantities

A channel $V(y|x_1, x_2, \ldots, x_m) \in \mathcal{P}(\mathcal{Y}|\mathcal{X}^m)$ is *symmetric* if for any permutation π on $[m]$ and for all $(x_1, x_2, \ldots, x_m, y)$

$$V(y|x_1, x_2, \ldots, x_m) = V(y|x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(m)}) \tag{3}$$

A channel $U(s|x_1, x_2, \ldots, x_m) \in \mathcal{P}(\mathcal{S}|\mathcal{X}^m)$*symmetrizes* an AVC $\mathcal{W}$ if the channel

$$V(y|x, x_1, \ldots, x_m) = \sum_{s \in \mathcal{S}} W(y|x, s) U(s|x_1, x_2, \ldots, x_m) \tag{4}$$

is a symmetric channel. Let $\mathcal{U}_{\text{sym}}(m)$ denote the set of channels which symmetrize $\mathcal{W}$:

$$\mathcal{U}_{\text{sym}}(m) = \{U(s|x^m) : V(y|x, x_1, \ldots, x_m) \text{ is symmetric}\} \tag{5}$$

If $U_1, U_2 \in \mathcal{U}_{\text{sym}}(m)$ and generate symmetric channels V_1 and V_2 according to (4), then $\alpha V_1 + (1 - \alpha) V_2$ is the channel generated from $\alpha U_1 + (1 - \alpha) U_2$. Since V_1 and V_2 are symmetric, so is $\alpha V_1 + (1 - \alpha) V_2$ and therefore $\alpha U_1 + (1 - \alpha) U_2 \in \mathcal{U}_{\text{sym}}(m)$. Thus, $\mathcal{U}_{\text{sym}}(m)$ is a closed, convex subset of channels $U(s|x_1, \ldots, x_m)$ defined by equality constraints in (3).

For a distribution $P \in \mathcal{P}(\mathcal{X})$, we define the strong symmetrizing cost $\lambda_m(P)$:

$$\lambda_m(P) = \min_{U \in \mathcal{U}_{\text{sym}}(m)} \max_{\bar{P} \in \mathcal{P}(\mathcal{X}^m) : P_i = P, \forall i \in [m]} \sum_{x^m} \sum_{s} \bar{P}(x^m) U(s|x^m) l(s) \tag{6}$$

This is the smallest expected cost over all symmetrizing channels $U(s|x^m) \in \mathcal{U}_{\text{sym}}(m)$, where the cost is measured over any joint distribution $\bar{P}(x^m)$ with marginals equal to P. The $\max$ and $\min$ are justified because the operations are performed over closed convex sets, and they can be reversed because the expected cost function is linear. We call an AVC *strongly m-symmetrizable* under the constraint Λ if $\lambda_m(P) \leq \Lambda$. We define the strong symmetrizability $L_{\text{sym}}(P, \Lambda)$ of the channel under input P and constraint Λ to be the largest integer m such that $\lambda_m(P) < \Lambda$. That is,

$$L_{\text{sym}}(P, \Lambda) = \max \{m : \lambda_m(P) < \Lambda\} \tag{7}$$

We also define the weak symmetrizing cost $\tilde{\lambda}_m(P)$:

$$\tilde{\lambda}_m(P) = \min_{U \in \mathcal{U}_{\text{sym}}(m)} \sum_{x^m} \sum_{s} P^m(x^m) U(s|x^m) l(s) \tag{8}$$

where P^m is the product distribution $P \times P \times \cdots \times P$. This is the smallest expected cost over all symmetrizing channels $U(s|x^m)$ where the cost is measured over the product distribution P^m. Again, the minimum is attained because $\mathcal{U}_{\text{sym}}(m)$ is closed. We call an AVC *weakly m-symmetrizable* under input P and constraint Λ if $\tilde{\lambda}_m(P) \leq \Lambda$. Similarly, the weak symmetrizability $\tilde{L}_{\text{sym}}(P, \Lambda)$ is the largest integer m such that $\tilde{\lambda}_m(P) < \Lambda$. That is,

$$\tilde{L}_{\text{sym}}(P, \Lambda) = \max \left\{m : \tilde{\lambda}_m(P) < \Lambda\right\} \tag{9}$$

Because the maximization in the definition of the strong symmetrizing cost $\lambda_m(P)$ in (6) is over all joint distributions $\bar{P}$

with marginals equal to P, it includes $\bar{P} = P^m$ in (8), and therefore $\lambda_m(P) \geq \tilde{\lambda}_m(P)$. This, in turn, implies that the strong symmetrizability is smaller than the weak symmetrizability: $L_{\mathrm{sym}}(P, \Lambda) \leq \tilde{L}_{\mathrm{sym}}(P, \Lambda)$.

For a fixed input distribution $P(x)$ on $\mathcal{X}$ and channel $V(y|x)$, let $I(P, V)$ denote the mutual information between the input and output of the channel:

$$I(P, V) = \sum_{x,y} V(y|x)P(x) \log \frac{V(y|x)P(x)}{P(x) \sum_{x'} V(y|x')P(x')} \tag{10}$$

We define the following two sets of distributions:

$$\mathcal{Q}(\Lambda) = \left\{ Q \in \mathcal{P}(\mathcal{S}) : \sum_s l(s)Q(s) \leq \Lambda \right\} \tag{11}$$

$$\mathcal{U}(P, \Lambda) = \left\{ U \in \mathcal{P}(\mathcal{S}|\mathcal{X}) : \sum_{s,x} U(s|x)P(x)l(s) \leq \Lambda \right\} \tag{12}$$

These, in turn, yield two information quantities:

$$C_{\mathrm{std}}(\Lambda) = \max_{P \in \mathcal{P}(\mathcal{X})} \min_{Q \in \mathcal{Q}(\Lambda)} I\left(P, \sum_s W(y|x, s)Q(s) \right) \tag{13}$$

$$C_{\mathrm{dep}}(\Lambda) = \max_{P \in \mathcal{P}(\mathcal{X})} \min_{U \in \mathcal{U}(P, \Lambda)} I\left(P, \sum_s W(y|x, s)U(s|x) \right) \tag{14}$$

The $\max$ and $\min$ exist and can be reversed because the mutual information is continuous, convex in the channel and concave in the input distribution, and the sets of input distributions and channels are closed and bounded. The channels in the second argument of the mutual information correspond to the convex closure and row-convex closure of the AVC $\mathcal{W}$ as defined in [10].

III. Main Results and Context

Capacity results for the AVC depend on the type of codes (randomized, deterministic, or list), error criterion (maximal or average), and the presence of constraints. In general, the maximal error capacity under deterministic coding is not known; a general solution would imply a formula for the zero-error capacity [1], [9]. When randomized coding is allowed, the capacity under maximal error is the same as average error. For an unconstrained AVC (where $\Lambda = \max_s l(s)$), Blackwell *et al.*[5] proved that the capacity under randomized coding is $C_r = C_{\mathrm{std}}(\max_s l(s))$. Ahlswede [3] showed that for unconstrained AVCs, the capacity $\bar{C}_d$ is either 0 or equal to $C_{\mathrm{std}}(\max_s l(s))$.

Under a state constraint Λ, Csiszár and Narayan [11], [12] proved that the randomized coding capacity is $C_r(\Lambda) = C_{\mathrm{std}}(\Lambda)$, and also found the deterministic coding capacity under average error $\bar{C}_d(\Lambda)$. They showed that if the AVC is nonsymmetrizable [14], then $\bar{C}_d(\Lambda) > 0$ and, in fact, $0 < \bar{C}_d(\Lambda) < C_{\mathrm{std}}(\Lambda)$ can hold. The reason this can happen is that the input distribution that maximizes $C_{\mathrm{std}}(\Lambda)$ may permit the jammer to find a channel U that symmetrizes the AVC and satisfies the cost constraint Λ. Therefore, certain input distributions are "disallowed," which lowers the rate.

The results in this paper are for the case of deterministic list codes. Without constraints, Ahlswede [2], [4] showed that a rate $C_{\mathrm{dep}}(\max_s l(s)) - \epsilon$ is achievable under maximal error with lists of size $O(1/\epsilon)$. The same approach works for constrained AVCs under maximal error.

Theorem 1 (List-Decoding for Maximal Error): Let $\mathcal{W}$ be an AVC with state cost function $l(s)$ and cost constraint Λ. Then, for any $\epsilon > 0$, the rate

$$R = C_{\mathrm{dep}}(\Lambda) - \epsilon$$

is achievable under maximal error using deterministic list codes with list size

$$L = O\left(\frac{1}{\epsilon}\right)$$

Furthermore, the capacity $C_L(\Lambda)$ under maximal error using list codes with list size L is bounded:

$$C_{\mathrm{dep}}(\Lambda) - O(L^{-1}) \leq C_L(\Lambda) \leq C_{\mathrm{dep}}(\Lambda)$$

The proof is given in Appendix A. For the converse, we exhibit a strategy for the jammer that lower bounds the probability of error. The code construction for the lower bound on the capacity proceeds in two steps. First, we show that a codebook containing all codewords of a given type P can be turned into a list code of rate close to

$$I\left(P, \min_{U \in \mathcal{U}(P, \Lambda)} \sum_s W(y|x, s)U(s|x) \right)$$

We can then sample codewords from this code to show that there exists a single codebook with constant list size L whose rate is close to $C_{\mathrm{dep}}(\Lambda) - O(L^{-1})$.

In the absence of state constraints, our definitions of weak and strong symmetrizability are the same, so $L_{\mathrm{sym}}(P, \max_s l(s)) = \tilde{L}_{\mathrm{sym}}(P, \max_s l(s)) = L_{\mathrm{sym}}$. Deterministic list codes for average error without constraints were studied independently by Blinovsky and colleagues [6], [7] and Hughes [15]. They showed a dichotomy similar to [3]: the list-L capacity $\bar{C}_L = 0$ for list sizes $L \leq L_{\mathrm{sym}}$, whereas the list-L capacity $\bar{C}_L = C_r$ for list sizes $L > L_{\mathrm{sym}}$.

Our results for list-decoding under average error are along the lines of [12]. For each list size L, we prove achievable and converse bounds.

Theorem 2 (List-Decoding for Average Error—Converse): Let $\mathcal{W}$ be an AVC with state cost function $l(\cdot)$ and cost constraint Λ. Then, we have the following upper bound on the deterministic list-coding capacity under average error $\bar{C}_L(\Lambda)$:

$$\bar{C}_L(\Lambda) \leq \max_{P \in \mathcal{P}(\mathcal{X}) : L_{\mathrm{sym}}(P, \Lambda) < L} \min_{Q \in \mathcal{Q}(\Lambda)} I\left(P, \sum_s W(y|x, s)Q(s) \right) \tag{15}$$

If for every $P \in \mathcal{P}(\mathcal{X})$, the strong symmetrizability satisfies $L_{\mathrm{sym}}(P, \Lambda) \geq L$, then $\bar{C}_L(\Lambda) = 0$.

The proof of Theorem 2 can be found in Appendix B. To prove the converse, we construct an explicit jamming strategy and give a lower bound on the probability of error for codes whose rate is above that in (15). For a codebook with codewords of type P, the jammer can choose a symmetrizing channel $U \in \mathcal{U}_{\mathrm{sym}}(L)$ such that the expected cost under any joint distribution with marginals equal to P is within the cost constraint.

Operationally, the jammer chooses L codewords from the codebook and uses them as inputs to U to generate a state sequence $\mathbf{s}$ which satisfies the cost constraints.

Theorem 3 (List-Decoding for Average Error—Achievability): Let $\mathcal{W}$ be an AVC with state cost function $l(\cdot)$ and cost constraint Λ. Then, we have the following lower bound on the deterministic list-coding capacity under average error $\bar{C}_L(\Lambda)$:

$$\bar{C}_L(\Lambda) \geq \max_{P \in \mathcal{P}(\mathcal{X}) : \tilde{L}_{\mathrm{sym}}(P,\Lambda) < L} \ \min_{Q \in \mathcal{Q}(\Lambda)} I\left(P, \sum_s W(y|x,s)Q(s)\right)$$

If P^* is the maximizing input distribution for $C_{\mathrm{std}}(\Lambda)$, then for list size $L > \tilde{L}_{\mathrm{sym}}(P^*, \Lambda)$, we have

$$\bar{C}_L(\Lambda) = C_{\mathrm{std}}(\Lambda)$$

The proof of Theorem 2 can be found in Appendix B. The achievability proof uses the codes of Hughes [15]. The existence of a code which is list-decodable is proved in [15] by using measure concentration to show that a random codebook with codewords of fixed type satisfies certain properties with overwhelming probability. However, we use a different decoding rule that extends [15] analogously to [12]. In order to prove that the decoding rule is successful, we require an input distribution P such that the AVC is not weakly L-symmetrizable.

For average error, the achievable rate and converse do not coincide in general, as shown in Section IV.

IV. Example

We will now show via an example that under average error, it is possible that $0 < \bar{C}_L(\Lambda) < C_{\mathrm{std}}(\Lambda)$. In particular, when the jammer must satisfy a constraint, positive rates may be achievable with list sizes that are smaller than the unconstrained symmetrizability, and for a fixed list size, the list-L capacity may be positive but strictly smaller than the randomized coding capacity. The reason for this is that the cost constraint may be such that the distribution P^* that achieves the randomized coding capacity may have a strong symmetrizing cost which is less than the constraint Λ, and therefore, the encoder cannot use that input distribution.

Let the input alphabet $\mathcal{X} = \{0, 1\}$, state alphabet $\mathcal{S} = \{0, 1, \ldots, \sigma\}$ and the channel be defined by

$$Y = X + S \tag{16}$$

with a quadratic cost function $l(s) = s^2$.

Without constraints, Hughes [15] has found that the randomized capacity is

$$C_r(\infty) = -\log \cos \frac{\pi}{\sigma + 3} \tag{17}$$

He also showed that for unconstrained AVCs the list-L capacity obeys a strict threshold:

$$\bar{C}_L(\infty) = \begin{array}{ll} -\log\cos\frac{\pi}{\sigma+3}, & L > \sigma \\ 0, & L \leq \sigma \end{array} \tag{18}$$

We are interested in the case when there is a cost constraint Λ on the jammer. We must calculate the minimum mutual information for different input distributions:

$$I(P, \Lambda) = \min_{Q\mathcal{Q}(\Lambda)} I(X \wedge Y)$$

The randomized-coding capacity under the cost constraint Λ is the max of $I(P, \Lambda)$ over P.

$$C_r(\Lambda) = \sup_{P \in \mathcal{P}(\mathcal{X})} I(P, \Lambda) \tag{19}$$

These calculations can be easily performed numerically.

To calculate the symmetrizability constraints, note that because the channel (16) is deterministic, the symmetry constraints imply that any channel $U \in \mathcal{U}_{\mathrm{sym}}(L)$ must also be symmetric. Therefore, for each $s \in \mathcal{S}$, the probability $U(s|x_1, x_2, \ldots, x_L)$ is only a function of the Hamming weight of $(x_1, x_2, \ldots, x_L)$. By letting t denote this weight, we can consider $\mathcal{U}_{\mathrm{sym}}(L)$ as containing channels of the form $U(s|t)$.

Channels $U \in \mathcal{U}_{\mathrm{sym}}(L)$ are symmetrizing, so for $t > 0$ we have

$$\sum_s W(y|0,s)U(s|t) = \sum_s W(y|1,s)U(s|t-1)$$

from which we can see that for $y = 1, 2, \ldots, \sigma$ and $t = 1, 2, \ldots, L$,

$$U(y|t) = U(y-1|t-1) \tag{20}$$

The only way that $y = 0$ is if $x = 0$ and $s = 0$. Similarly, the only way $y = \sigma + 1$ is if $x = 1$ and $s = \sigma$. Therefore

$$U(0|t) = 0, \qquad t = 1, 2, \ldots, L \tag{21}$$

$$U(\sigma|t) = 0, \qquad t = 0, 1, \ldots, L-1 \tag{22}$$

The conditions (20)–(22) characterize the linear symmetry constraints in $\mathcal{U}_{\mathrm{sym}}(L)$.

Thus, for each input distribution P, we can find

$$\min_{U \in \mathcal{U}_{\mathrm{sym}}(L)} \sum_{s,t} l(s)U(s|t)\binom{L}{t}P(0)^{L-t}P(1)^t$$

This is a simple linear program. To calculate the strong L-symmetrizing cost, note that the set of all joint distributions $\bar{P}(x_1^L)$ with marginals equal to P is also a convex set defined by linear equality constraints. Likewise, it is simple to numerically evaluate the strong symmetrizing cost

$$\sup_{\bar{P}} \min_{U \in \mathcal{U}_{\mathrm{sym}}(L)} \sum_{s,t} l(s)U(s|t) \sum_{x_1^L : T_{\mathbf{x}} = t/L} \bar{P}(x_1^L)$$

We calculated the achievable rates and converse bounds for $\sigma = 8$, and the results are shown for list sizes $L = 2$ and $L = 4$ in Figs. 1 and 2. For state constraint Λ, the randomized coding capacity $C_r(\Lambda)$ in (19) is given by the dotted line. The achievable rate of Theorem 3 is shown by the solid line, and the converse bound of Theorem 2 by the dashed line. These two

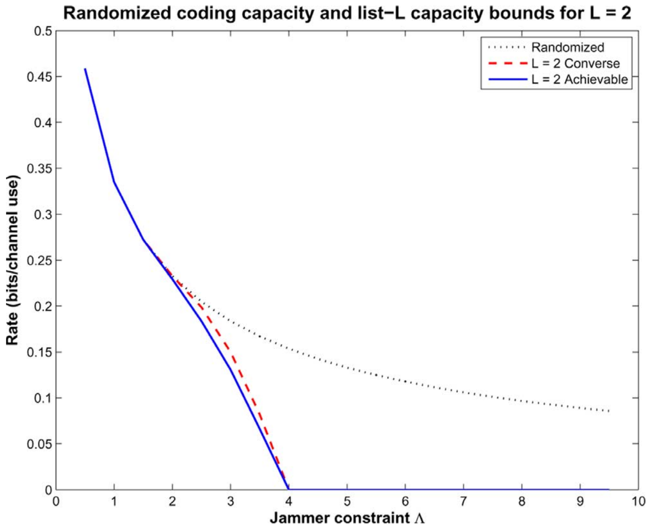

Fig. 1. Randomized coding capacity $C_r(\Lambda)$ and bounds on list-L capacity $\bar{C}_L(\Lambda)$ versus the state constraint Λ for $L = 2$.

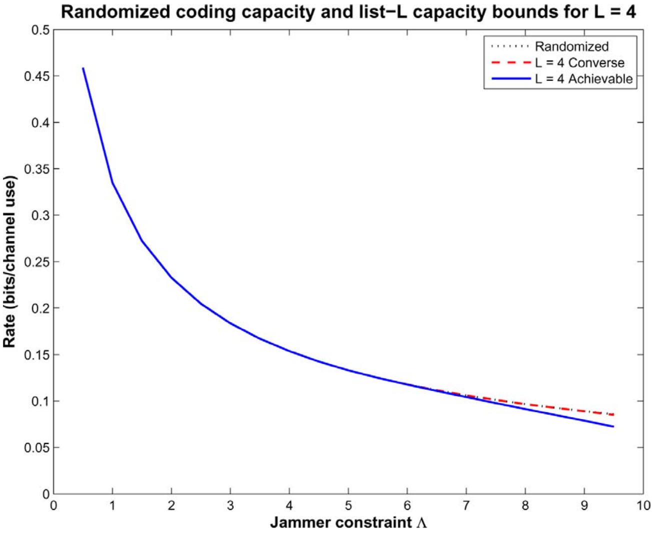

Fig. 2. Randomized coding capacity $C_r(\Lambda)$ and bounds on list-L capacity $\bar{C}_L(\Lambda)$ versus the state constraint Λ for $L = 4$.

curves are given by restricting the optimization over P in the right side of (19).

Figs. 1 and 2 show that when $\Lambda < \sigma^2$, positive rates are achievable for several different list sizes. For a range of Λ, the randomized coding capacity is achievable using lists of size 2 or 4. Fig. 1 also illustrates the fundamental difference between list-decoding with state constraints and list-decoding without constraints: for a range around $\Lambda = 3$, the list-2 capacity $\bar{C}_2(\Lambda)$ is positive but strictly smaller than the randomized coding capacity $C_r(\Lambda)$, in contrast with (18).

V. Discussion

This paper provides several new results on list-decoding for AVCs with state constraints. For maximal error, we showed that rates within $O(1/L)$ of $C_{\text{dep}}(\Lambda)$ are achievable with list codes of list size L. This result can be used together with a construction from [16] to show that $C_{\text{dep}}(\Lambda)$ is the randomized coding capacity of AVCs with input-dependent state [20]. For average error, we provided an achievable rate and converse which do not coincide in general. We conjecture that the converse region of Theorem 2 is not tight and that a stronger converse could be shown. The strong symmetrizing cost in (6) allows optimization over all joint distributions with the same marginals. The converse proof uses a jamming strategy corresponding to taking a random set of L codewords from the codebook as inputs to a symmetrizing channel $U(s|x^L)$ to generate the state sequence. The strong symmetrizing cost is a conservative bound on the cost of such a strategy. It may be that techniques such as in [21] could improve this bound; we leave this for future work. Our results here establish that the behavior of list-decoding for constrained AVCs is fundamentally different than the unconstrained case, analogous to the situation for list size 1.

It may be possible to extend the results in this paper to other situations. Input constraints can be introduced by restricting the maximization over the input distribution to the set of P which satisfy the input constraint. Extensions of the average error results to multiuser scenarios such as [19] may also be possible, but the symmetrizability conditions may become quite baroque. Finally, using the approach here in the Gaussian setting would involve developing measure concentration results which could be interesting in their own right.

Appendix A
Maximal Error

Using standard typicality arguments, we can show the existence of list-decodable codes for maximal error with exponential list size. The codebook is the entire set of typical sequences $\mathcal{T}_P$ and the list is the union of ϵ-shells under the different state sequences. The decoder observes an output sequence $\mathbf{y}$ and outputs a list of all sequences $\mathbf{x} \in \mathcal{T}_P$ such that $\mathbf{x}$ and $\mathbf{y}$ are jointly typical with respect to a joint distribution induced by a channel $U \in \mathcal{U}(P, \Lambda)$. Let

$$\mathcal{W}_{dep}(P,\Lambda) = \Big\{ V : V(y|x) = \sum_s W(y|x,s)U(s|x), \quad U \in \mathcal{U}(P,\Lambda) \Big\} \quad (23)$$

Proof of Theorem 1: Because we are using the maximal error criterion, it is sufficient for the jammer to inflict a large error probability on a single codeword. To prove the converse, we construct a randomized strategy for the jammer for each codeword $\mathbf{x}$ in the code. The behavior of the AVC under this strategy can be bounded via the behavior of an appropriately constructed discrete memoryless channel (DMC). The converse then follows from the strong converse for list-decoding for the DMC [2], [18], [22]. The achievable strategy uses random coding by sampling codewords from the set of sequences of fixed composition to show the existence of a deterministic list code.

Converse: Suppose that for some $\delta > 0$ and $0 < \epsilon < 1$ there exists an unbounded, increasing sequence of blocklengths

$n_1, n_2, \ldots$ and where for each n_t there exists a (n_t, N_t, L) deterministic list code $\mathcal{C}_t$ where

$$\frac{1}{n_t} \log \frac{N_t}{L} > C_{\mathrm{dep}}(\Lambda) + \delta$$

and the maximal error of each code is less than ϵ. Let $N_t(P)$ be the number of codewords of type P in $\mathcal{C}_t$. Since the number of types is at most $(n_t+1)^{|\mathcal{X}|}$, there exists a type P_t such that

$$(n_t+1)^{|\mathcal{X}|} N_t(P_t) > L \exp\left(n_t(C_{\mathrm{dep}}(\Lambda) + \delta)\right)$$

and therefore, for sufficiently large t, there exists a P_t such that the subcode $\mathcal{C}_t(P_t)$ of $\mathcal{C}_t$ consisting of codewords of type P satisfies

$$\frac{1}{n_t} \log \frac{N_t(P_t)}{L} > C_{\mathrm{dep}}(\Lambda) + \delta/2$$

Then, $\mathcal{C}_t(P_t)$ is an $(n_t, N_t(P_t), L)$ deterministic list-code with maximal error less than ϵ.

We will now show that there exists a DMC over which the sequence of codes $\{\mathcal{C}_t(P_t)\}$ cannot achieve arbitrarily small probability of error. This DMC can be approximated by the jammer using a randomized strategy for selecting the state sequence $\mathbf{s}$ based on the transmitted codeword $\mathbf{x}$. Because we are considering maximal error, it is sufficient for the jammer to inflict a large error probability on a single message.

For $\epsilon' > 0$, define the channel from $\mathcal{X}$ to $\mathcal{S}$ as

$$U_{P_t,\epsilon'} = \operatorname*{argmin}_{U \in \mathcal{U}(P_t, \Lambda - \epsilon')} I\left(P_t, \sum_s W(y|x,s) U(s|x)\right)$$

The minimizer is unique by the convexity of the mutual information. For any $\mathbf{x} \in \mathcal{C}_t(P_t)$, let $\mathbf{S}(\mathbf{x}, \epsilon')$ have distribution $U_{P_t,\epsilon'}(\mathbf{s}|\mathbf{x})$. For any $\delta' > 0$, there exists a t sufficiently large such that

$$\mathbb{P}\left(l(\mathbf{S}(\mathbf{x}, \epsilon')) < \Lambda\right) > 1 - \delta' \qquad (24)$$

Consider the DMC formed from the channel $W(y|x,s)$ by choosing s according to the channel $U_{P_t,\epsilon'}(s|x)$:

$$V_{P_t,\epsilon'}(y|x) = \sum_s W(y|x,s) U_{P_t,\epsilon'}(s|x)$$

Because the mutual information is continuous, for any $\delta > 0$, there exists an $\epsilon' > 0$ such that

$$I\left(P, V_{P_t,\epsilon'}\right) \leq \min_{U \in \mathcal{U}(P_t,\Lambda)} I\left(P_t, \sum_s W(y|x,s) U(s|x)\right) + \delta/4$$

For n_t sufficiently large, we have

$$\frac{1}{n_t} \log \frac{N_t(P_t)}{L} > I\left(P, V_{P_t,\epsilon'}\right) + \delta/4 \qquad (25)$$

Let $\hat{\varepsilon}(\mathbf{x}, \mathcal{C}_t(P_t), V_{P_t,\epsilon'})$ be the error for codeword $\mathbf{x}$ in the code $\mathcal{C}_t(P_t)$ on the DMC $V_{P_t,\epsilon'}$. The proof of the strong converse for list coding [18] over the DMC $V_{P_t,\epsilon'}$ shows that for a code with codewords of type P_t, if (25) holds then there exist positive constants c_1 and c_2 such that

$$\frac{1}{N_t} \sum_{i=1}^{N_t} \hat{\varepsilon}(\mathbf{x}(i), \mathcal{C}_t(P_t), V_{P_t,\epsilon'}) \geq 1 - c_1 e^{-c_2 n_t} \qquad (26)$$

where $\mathbf{x}(i)$ is the corresponds to the ith message of the code $\mathcal{C}_t(P_t)$.

We now connect the DMC $V_{P_t,\epsilon'}$ to the AVC. Because we are considering maximal error, the jammer arbitrarily selects a codeword $\mathbf{x}$ in the code and chooses a state sequence according to the following strategy. For a codeword $\mathbf{x}$, it generates $\mathbf{S}(\mathbf{x}, \epsilon')$. If $l(\mathbf{S}(\mathbf{x}, \epsilon')) > \Lambda$, then it sets $\mathbf{s}$ equal to some fixed sequence $\mathbf{s}_0$ such that $l(\mathbf{s}_0) < \Lambda$. Otherwise, it sets $\mathbf{s} = \mathbf{S}(\mathbf{x}, \epsilon')$. We will now show that this strategy will result in a large error probability when $\mathbf{x}$ is chosen by the encoder.

Let $\varepsilon(\mathbf{x}, \mathcal{C}_t(P_t), \epsilon')$ be the error for codeword $\mathbf{x}$ in $\mathcal{C}_t(P_t)$ under this strategy. Then, from (24), we have

$$\hat{\varepsilon}(\mathbf{x}, \mathcal{C}_t(P_t), V_{P,\epsilon'}) \leq \varepsilon(\mathbf{x}, \mathcal{C}_t(P_t), \epsilon') + \delta'$$

Therefore, using (26), we have

$$\varepsilon(\mathbf{x}, \mathcal{C}_t(P), \epsilon') \geq 1 - c_1 e^{-c_2 n_t} - \delta'$$

which gives a lower bound on the maximal error for the code $\mathcal{C}_t(P_t)$ over the AVC. For sufficiently large n_t, this lower bound can be made larger than ϵ, which is a contradiction. Therefore, the capacity of the AVC under maximal error and list-decoding is upper bounded by $C_{\mathrm{dep}}(\Lambda)$.

Achievability: Let $P \in \mathcal{P}_n(\mathcal{X})$ and $\mathcal{T}_P$ denote the set of all sequences of length n of type P. For any channel $V(y|x)$, we define a channel $V'(x|y)$ by

$$V'(x|y) = \frac{V(y|x)P(x)}{\sum_{x'} V(y|x')P(x')}$$

For a sequence $\mathbf{y}$ and the channel V', define $V' \times T_{\mathbf{y}}$ to be the distribution such that $[V' \times T_{\mathbf{y}}](x,y) = V'(x|y) T_{\mathbf{y}}(y)$. The (V', ϵ)-shell of typical $\mathbf{x}$ sequences around a $\mathbf{y}$ is

$$T_{V'}^{\epsilon}(\mathbf{y}) = \{\mathbf{x} \in \mathcal{T}_P : d_{\max}\left(T_{\mathbf{xy}}, V' \times T_{\mathbf{y}}\right) < \epsilon\}$$

For n sufficiently large, we have

$$\frac{1}{n} \log |T_{V'}^{\epsilon}(\mathbf{y})| \leq H_{V'T_{\mathbf{y}}}(X|Y) + O(\epsilon \log \epsilon^{-1})$$

where the subscript on H indicates the joint distribution under which to take the conditional entropy.

Now, for a fixed $\mathbf{x} \in \mathcal{T}_P$ and $\mathbf{s}$ with $l(\mathbf{s}) \leq n\Lambda$, define a channel from $\mathcal{X}$ to $\mathcal{Y}$ by

$$V_{\mathbf{xs}}(y|x) = \sum_s W(y|x,s) \frac{N(x,s|\mathbf{x},\mathbf{s})}{N(x|\mathbf{x})}$$

Note that $V_{\mathbf{xs}} \in \mathcal{W}_{\mathrm{dep}}(P, \Lambda)$. Let $\mathbf{Y}$ be generated via $W(y|x,s)$ from $(\mathbf{x}, \mathbf{s})$. For each $(x,y) \in \mathcal{X} \times \mathcal{Y}$, applying a Chernoff–Hoeffding bound [13] yields

$$\mathbb{P}\left(|[P \times V_{\mathbf{xs}}](x,y) T_{\mathbf{xY}}(x,y)| > \epsilon\right) \leq 2\exp(-2\epsilon^2 n) \qquad (27)$$

where $[P \times V_{\mathbf{xs}}](x,y) = P(x)V_{\mathbf{xs}}(y|x)$. Therefore, with probability $1 - 2|\mathcal{X}||\mathcal{Y}|\exp(-2\epsilon^2 n)$, the received sequence $\mathbf{Y}$ is jointly typical with $\mathbf{x}$.

For a fixed received sequence $\mathbf{y}$ and constant $\delta > 0$, define the set $\mathcal{V}_P^\delta(\mathbf{y})$ of channels:

$$\mathcal{V}_P^\delta(\mathbf{y}) = \Bigg\{ V \in \mathcal{W}_{\text{dep}}(P,\Lambda) \cap \mathcal{P}_n(\mathcal{Y}|\mathcal{X}) : \\ d_{\max}\left(\sum_y V(y|x)P(x), T_{\mathbf{y}} \right) < \delta \Bigg\}$$

The intersection with $\mathcal{P}_n(\mathcal{Y}|\mathcal{X})$ ensures that $|\mathcal{V}_P^\delta(\mathbf{y})|$ grows polynomially with n. Now define the following set:

$$\mathcal{A}(\mathbf{y}) = \bigcup_{V \in \mathcal{V}_P^\delta(\mathbf{y})} T_{V'}^{(|\mathcal{X}|+1|)\delta}(\mathbf{y})$$

The size of this set is exponential as a function of n and for sufficiently large n, it can be upper bounded:

$$\frac{1}{n}\log|\mathcal{A}(\mathbf{y})| \le \min_{V \in \mathcal{W}_{\text{dep}}(P,\Lambda)} H_{PV}(X|Y) + O(\delta \log \delta^{-1}) \tag{28}$$

Let $A = \max_{\mathbf{y}} |\mathcal{A}(\mathbf{y})|$.

Consider an $(n, |\mathcal{T}_P|, A)$ list code where the codewords are all sequences in $\mathcal{T}_P$ and the decoder outputs the list $\mathcal{A}(\mathbf{y})$. Note that the size of the output list depends on $\mathbf{y}$. We claim that for any $\delta > 0$ and $\delta' > 0$, there exists an n sufficiently large such that this list code has error probability less than δ'.

To see this, fix $\delta > 0$ and $\delta' > 0$ and suppose that some $\mathbf{x}$ was transmitted and the state sequence was $\mathbf{s}$. From (27), there exists an $\epsilon > 0$ and n sufficiently large such that the received $\mathbf{Y}$ satisfies $d_{\max}(P \times V_{\mathbf{xs}}, T_{\mathbf{xY}}) \le \epsilon$ with probability $1 - \delta'$. By choosing ϵ sufficiently small and n sufficiently large, with probability $1 - \delta'$ over the channel we have $\mathbf{x} \in \mathcal{A}(\mathbf{Y})$.

To arrive at the desired code, fix $\eta > 0$. Let $\mathcal{B} = \{\mathbf{X}(i)\}$ be a set of $2^{n(C_{\text{dep}}(\Lambda) - \eta)}$ codewords from $\mathcal{T}_P$ uniformly at random and set the decoder to output $\mathcal{A}(\mathbf{Y}) \cap \mathcal{B}$. We must show this set produced by the decoder has at most $L = O(1/\eta)$ codewords with high probability. This implies that there exists a deterministic $(n, 2^{n(C_{\text{dep}}(\Lambda) - \eta)}, L)$ deterministic list code with small probability of error.

Let $R = C_{\text{dep}}(\Lambda) - \eta$. For any fixed $\mathbf{y}$, the probability that any codeword of $\mathcal{B}$ is in $\mathcal{A}(\mathbf{y})$ is upper bounded by $|\mathcal{A}(\mathbf{y})|/|\mathcal{T}_P|$, so from (28), we see that for any $\delta > 0$, we can choose n sufficiently large such that

$$\mathbb{P}(\mathbf{X}(i) \in \mathcal{A}(\mathbf{y})) \le \exp\left(-n\left(C_{\text{dep}}(\Lambda) - O(\delta \log \delta^{-1})\right)\right)$$

Because the codewords are selected independently, Sanov's [8, Th. 12.4.1], bounds the probability that a fraction $L \cdot 2^{-nR}$ of the 2^{nR} codewords end up in $\mathcal{A}(\mathbf{y})$:

$$\begin{aligned}\mathbb{P}(|\mathcal{A}(\mathbf{y}) \cap \mathcal{B}| > L) &\le \exp\Big(-2^{nR} D\left(L2^{-nR} \,\Big\|\, 2^{-n(C_{\text{dep}}(\Lambda) - O(\delta\log\delta^{-1}))} \right) \\ &\qquad + \log(2^{nR}+1) \Big)\end{aligned} \tag{29}$$

Now, we can bound the first summand in the exponent on the right-hand side of (29) (the term $-2^{nR}D(\cdot\|\cdot))$ by

$$\begin{aligned}&-L\log\frac{L}{2^{n(\eta - O(\delta\log\delta^{-1}))}} \\ &\quad - 2^{nR}(1 - L2^{-nR})\log\frac{1 - L2^{-nR}}{1 - 2^{-n(R+\eta-O(\delta\log\delta^{-1}))}} \\ &\qquad \le -nL\left(\eta - O(\delta\log\delta^{-1})\right) - L\log L + 2L\end{aligned} \tag{30}$$

We can pick δ such that $O(\delta\log\delta^{-1}) < \eta/2$ by choosing n sufficiently large. Then, substituting (30) in (29), upper bounding $R < \log|\mathcal{Y}|$, and taking a union bound over all $\mathbf{y}$, we have

$$\begin{aligned}&\mathbb{P}(\exists \mathbf{y} : |\mathcal{A}(\mathbf{y}) \cap \mathcal{B}| > L) \\ &\qquad \le \exp\left(-n\left(L\epsilon/2 - 2\log|\mathcal{Y}|\right) - L\log L + 2L\right)\end{aligned}$$

For sufficiently large n, choosing $L > \lceil \frac{4\log|\mathcal{Y}|}{\eta} \rceil$ makes the exponent negative, showing that with high probability the random selection will produce an $(n, 2^{nR}, L)$ list code under maximal error whose error can be made as small as we like. Therefore, such a deterministic list code exists. ■

Appendix B
Average Error

1) Converse:

Lemma 1 (Approximating Joint Distributions): Let $\mathcal{X}$ be a finite set with $|\mathcal{X}| \ge 2$. For any $\epsilon > 0$ and probability distribution P on $\mathcal{X}$, there exists a $\delta > 0$ such that for any collection of distributions $\{P_i \in \mathcal{P}(\mathcal{X}) : i \in [L]\}$ satisfying

$$d_{\max}(P_i, P) < \delta \qquad \forall i \tag{31}$$

and any joint distribution $\bar{P}(x_1, x_2, \ldots, x_L)$ with

$$\sum_{x_j : j \ne i} \bar{P}(x_1, x_2, \ldots, x_L) = P_i(x_i) \qquad \forall i,\ x_i \in \mathcal{X} \tag{32}$$

there exists a joint distribution $\hat{P}(x_1, x_2, \ldots, x_L)$ such that

$$\sum_{x_j : j \ne i} \hat{P}(x_1, x_2, \ldots, x_L) = P(x_i) \qquad \forall i,\ x_i \in \mathcal{X} \tag{33}$$

and

$$d_{\max}\left(\bar{P}, \hat{P}\right) < \epsilon \tag{34}$$

2) Proof of Lemma 1: Fix $\epsilon > 0$ and P. We consider two cases depending on whether $\min_{x \in \mathcal{X}} P(x) = 0$ or not.

Case 1: First suppose $\min_{x \in \mathcal{X}} P(x) = \beta > 0$. Consider a set of distributions $\{P_i : i \in [L]\}$ satisfying (31) and let $\bar{P}(x_1^L)$ be a joint distribution satisfying (32). We treat probability distributions as vectors in $\mathbb{R}^{|\mathcal{X}|^L}$. We can construct a distribution $\hat{P}$ satisfying (33) and (34) in two steps: first we project $\bar{P}$ onto the set of all vectors whose entries sum to 1 and satisfy (33), and then, we find a $\hat{P}$ close to this projection which is a proper probability distribution.

Let $\mathcal{B}$ be the subspace of $\mathbb{R}^{|\mathcal{X}|^L}$ of all probability distributions P' satisfying the marginal constraints (33). We can summarize these linear constraints in the matrix form

$$AP' = b'$$

where A and b' contain the coefficients corresponding to the constraints in (33). We can assume A has full row-rank by removing linearly dependent constraints. Similarly, the distribution $\bar{P}$ satisfies

$$A\bar{P} = \bar{b}$$

where A and $\bar{b}$ contain the coefficients corresponding to the constraints in (32).

Let $\tilde{P}$ be the Euclidean projection of $\bar{P}$ onto the subspace $\mathcal{B}$:

$$\tilde{P} = \bar{P} + A^T(AA^T)^{-1}(b' - A\bar{P}) \tag{35}$$

The error in the projection is

$$\begin{aligned}\bar{P} - \tilde{P} &= A^T(AA^T)^{-1}(A\bar{P} - b')\\ &= A^T(AA^T)^{-1}(\bar{b} - b')\end{aligned}$$

From (31), all elements of $(\bar{b} - b')$ are in $(-\delta, \delta)$. Since the rows of A are linearly independent, the singular values of A are strictly positive and a function of $|\mathcal{X}|$ and L only. Therefore, there is a positive function $\mu_1(|\mathcal{X}|, L)$ such that

$$\left\| A^T(AA^T)^{-1}(\bar{b} - b') \right\|_2 < \mu_1(|\mathcal{X}|, L) \cdot \delta$$

Since $|\mathcal{X}|$ is finite, there is a function $\mu_2(|\mathcal{X}|, L)$ such that

$$d_{\max}\left(\tilde{P}(x_1^L), \bar{P}(x_1^L)\right) < \mu_2(|\mathcal{X}|, L) \cdot \delta$$

If the $\tilde{P}$ from this projection has all nonnegative entries, then we set $\hat{P} = \tilde{P}$ and choose δ sufficiently small so that $\mu_2(|\mathcal{X}|, L) \cdot \delta < \epsilon$.

If $\tilde{P}$ has entries that are not in $[0, 1]$, then it is not a valid probability distribution. However, since $\bar{P}$ is a probability distribution, we know that

$$\min_{x_1^L} \tilde{P}(x_1^L) > -\mu_2(|\mathcal{X}|, L) \cdot \delta$$

Let P^L be the joint distribution on $\mathcal{X}^L$ with independent marginals P:

$$P^L(x_1, \ldots, x_L) = P(x_1) \cdots P(x_L) \tag{36}$$

Since $\min_x P(x) > \beta$, we have $P^L(x_1^L) > \beta^L$ for all L. Let

$$\alpha = \frac{\mu_2(|\mathcal{X}|, L) \cdot \delta}{\beta^L} \tag{37}$$

and set

$$\hat{P} = (1 - \alpha)\tilde{P} + \alpha P^L \tag{38}$$

Then, $\hat{P}(x_1^L) > 0$ for all x_1^L and by the triangle inequality:

$$\begin{aligned}d_{\max}\left(\bar{P}, \hat{P}\right) &\le d_{\max}\left(\bar{P}, \tilde{P}\right) + d_{\max}\left(\tilde{P}, \hat{P}\right)\\ &< \mu_2(|\mathcal{X}|, L) \cdot \delta + \alpha d_{\max}\left(\tilde{P}, P^L\right)\\ &< \left(1 + \frac{1}{\beta^L}\right) \mu_2(|\mathcal{X}|, L) \cdot \delta\end{aligned}$$

Therefore, for δ sufficiently small, we can choose a $\hat{P}$ such that $d_{\max}\left(\bar{P}, \hat{P}\right) < \epsilon$ for any $\epsilon > 0$.

Case 2: Suppose that $\min_{x \in \mathcal{X}} P(x) = 0$. Let $\mathcal{X}_0 = \{x \in \mathcal{X} : P(x) = 0\}$ and $\mathcal{Z} = \mathcal{X} \setminus \mathcal{X}_0$. Let $Q \in \mathcal{P}(\mathcal{Z})$ be the restriction of P to $\mathcal{Z}$. Then, Q is a probability distribution on $\mathcal{Z}$. First suppose that $|\mathcal{Z}| = 1$. Then, $P(x) = 1$ for some $x \in \mathcal{X}$. Let

$$\hat{P}(x_1^L) = P(x_1) \cdots P(x_L)$$

Since all the marginal distributions P_i of $\bar{P}$ satisfy $d_{\max}(P, P_i) < \delta$, we know that $d_{\max}\left(\bar{P}, \hat{P}\right) < \delta$.

Now suppose $|\mathcal{Z}| \ge 2$. We can construct $\hat{P}$ by first finding a joint distribution $\bar{Q}$ that is close to $\bar{P}$ and then invoking the first case of this proof on $\bar{Q}$ using (35)–(38). From (31), we know that for some $0 < c < |\mathcal{X}^L|$, we have

$$\sum_{x_1^L \notin \mathcal{Z}^L} \bar{P}(x_1, x_2, \ldots, x_L) \stackrel{\triangle}{=} c\delta$$

Define $\bar{Q}$ by

$$\bar{Q}(x_1^L) = \begin{cases} \bar{P}(x_1^L) + |\mathcal{Z}|^{-L} c\delta, & x_1^L \in \mathcal{Z}^L \\ 0, & x_1^L \notin \mathcal{Z}^L \end{cases}$$

Since $\bar{Q}$ has support only on $\mathcal{Z}^L$, we can think of it either as a distribution on $\mathcal{X}^L$ or on $\mathcal{Z}^L$. Note that

$$d_{\max}\left(\bar{P}, \bar{Q}\right) < c\delta$$

Let $\{Q_i : i \in [L]\}$ be the ith marginal distributions of $\bar{Q}$, so that

$$Q_i(x_i) = \sum_{x_j : j \ne i} \bar{Q}(x_1, x_2, \ldots, x_L) = Q_i(x_i)$$

for all $i \in [L]$ and $x_i \in \mathcal{Z}$. Then, we have for some $c' > 0$ that $d_{\max}(Q, Q_i) < c'\delta$.

Now, we can apply Case 1 of this proof [see (35)–(38)] using the set $\mathcal{Z}$ and distributions Q, $\{Q_i : i \in [L]\}$, and $\bar{Q}$. For any $\epsilon_1 > 0$, we can find a $\delta_1 > 0$ such that if $\{Q_i\}$ satisfy $d_{\max}(Q, Q_i) < \delta_1$, then there exists a $\hat{Q}$ with marginals equal to Q such that $d_{\max}\left(\bar{Q}, \hat{Q}\right) < \epsilon_1$. Let $\hat{P}$ be the extension of $\hat{Q}$ to a distribution on $\mathcal{X}^L$ by setting $\hat{P}(x_1^L) = \hat{Q}(x_1^L)$ for $x_1^L \in \mathcal{Z}^L$ and 0 elsewhere. By the triangle inequality

$$\begin{aligned}d_{\max}\left(\bar{P}, \hat{Q}\right) &\le d_{\max}\left(\bar{P}, \bar{Q}\right) + d_{\max}\left(\bar{Q}, \hat{Q}\right)\\ &< c\delta + \epsilon_1\end{aligned}$$

We can choose δ sufficiently small so that δ_1 and ϵ_1 are sufficiently small to guarantee that this distance is less than ϵ. ■

Lemma 2: Let $\mathcal{W}$ be an AVC with state cost function $l(\cdot)$ and constraint Λ and let L be a positive integer. Let $\epsilon > 0$ be arbitrary and suppose P is a distribution with $\lambda_L(P) < \Lambda - \epsilon$. Then, there exists a $\delta > 0$ and n_0 such that for any (n, N, L) list code with $n \geq n_0$ and $N \geq L+1$ whose codewords $\{\mathbf{x}(i) : i \in [N]\}$ satisfy

$$\begin{aligned} d_{\max}\left(T_{\mathbf{x}(i)}, P\right) &< \delta \qquad \forall i \in [N] \\ \lambda_L(T_{\mathbf{x}(i)}) &< \Lambda - \epsilon \qquad \forall i \in [N] \end{aligned}$$

the average error for the code is lower bounded:

$$\max_{\mathbf{s}\in\mathcal{S}^n(\Lambda)} \bar{\varepsilon}_L(\mathbf{s}) > 2\left(\frac{1}{L+2} - \frac{L}{N(L+1)}\right)$$

Proof: From Lemma 1, we can see that for any $\epsilon_1 > 0$, there exists a $\delta_1 > 0$ such that for any set $J \subset [N]$ of codewords with $|J| = L$ and $d_{\max}\left(T_{\mathbf{x}(j)}, P\right) < \delta_1$, we can find a joint type $\bar{P} \in \mathcal{P}(\mathcal{X}^L)$ with marginals equal to P such that the joint type $T_{\mathbf{x}(J)}$ satisfies

$$d_{\max}\left(T_{\mathbf{x}(J)}, \bar{P}\right) < \epsilon_1$$

Now let U^* achieve the minimum in the definition of $\lambda_L(P)$. Since $\lambda_L(P) < \Lambda - \epsilon$, we have

$$\begin{aligned} \sum_{s,x_1^L} l(s)U^*(s|x_1^L)T_{\mathbf{x}(J)}(x_1^L) &\leq \sum_{s,x_1^L} l(s)U^*(s|x_1^L)\bar{P}(x_1^L) \\ &\quad + \epsilon_1\lambda^*|\mathcal{X}|^L \\ &< \Lambda - \epsilon + \epsilon_1\lambda^*|\mathcal{X}|^L \end{aligned}$$

where $\lambda^* = \max_{s\in\mathcal{S}} l(s)$. Now, choose $\epsilon_1 = \epsilon/(2\lambda^*|\mathcal{X}|^L)$ so that

$$\sum_{s,x_1^L} l(s)U^*(s|x_1^L)T_{\mathbf{x}(J)}(x_1^L) < \Lambda - \epsilon/2$$

and choose $\delta = \delta_1$ according to Lemma 1.

Let $\mathcal{J}$ be the set of all subsets of $[N]$ of size L, and let $\mathbf{J}$ be a random variable uniformly distributed on $\mathcal{J}$. Consider the following jamming strategy. The jammer draws a subset $\mathbf{J}$ and for $\mathbf{J} = J$ selects the state sequence according to the random variable $\mathbf{S}(J)$ with distribution

$$Q^n(\mathbf{s}) = \prod_{t=1}^{n} U^*(s_t|\{x_t(j) : j \in J\})$$

The expected cost of $\mathbf{S}(J)$ is

$$\begin{aligned} &\frac{1}{n}\mathbb{E}[l(\mathbf{S}(J))] \\ &\quad = \frac{1}{n}\sum_{t=1}^{n}\sum_{\mathbf{s}} l(s_t)U^*(s_t|\{x_t(j) : j \in J\}) \\ &\quad = \sum_{s,\tilde{x}^L} l(s)U^*(s|\tilde{x}_1,\ldots,\tilde{x}_L)\frac{|\{t : x_t(j) = \tilde{x}_j \ \forall j\}|}{n} \\ &\quad = \sum_{s,\tilde{x}^L} l(s)U^*(s|\tilde{x}_1^L)T_{\mathbf{x}(J)} \\ &\quad < \Lambda - \epsilon/2 \end{aligned}$$

We can also bound the variance of $l(\mathbf{S}(J))$:

$$\mathrm{Var}\left(l(\mathbf{S}(J))\right) \leq \frac{(\lambda^*)^2}{n}$$

Chebyshev's inequality gives the bound:

$$\begin{aligned} \mathbb{P}(l(\mathbf{S}(J)) > \Lambda) &\leq \frac{(\lambda^*)^2}{n(\Lambda - (\Lambda - \epsilon/2))^2} &(39) \\ &\leq \frac{4(\lambda^*)^2}{n\epsilon^2} &(40) \end{aligned}$$

Before continuing, we need some properties of symmetrizing channels used to generate the random variables $\mathbf{S}(J)$. First, we have for any $j \in J$:

$$\begin{aligned} &\mathbb{E}\left[W^n(\mathbf{y}|\mathbf{x}(i), \mathbf{S}(J))\right] \\ &\quad = \sum_{\mathbf{s}} W^n(\mathbf{y}|\mathbf{x}(i), \mathbf{s})U^{*n}(\mathbf{s}|\{x(j') : j' \in J\}) \\ &\quad = \mathbb{E}\left[W^n(\mathbf{y}|\mathbf{x}(j), \mathbf{S}(J \setminus \{j\} \cup \{i\}))\right] \qquad (41) \end{aligned}$$

Using (41), we can see that for some subset $G \subset [N]$ with $|G| = L+1$:

$$\begin{aligned} &\sum_{i\in G}\mathbb{E}\left[\bar{\varepsilon}_L(i, \mathbf{S}(G \setminus \{i\}))\right] \\ &\quad = \sum_{i\in G}\left(1 - \sum_{\mathbf{y}:i\in\psi(\mathbf{y})} \mathbb{E}\left[W^n(\mathbf{y}|\mathbf{x}_i, \mathbf{S}(G \setminus \{i\}))\right]\right) \\ &\quad = L + 1 - \sum_{i\in G}\sum_{\mathbf{y}:i\in\psi(\mathbf{y})} \mathbb{E}\left[W^n(\mathbf{y}|\mathbf{x}_{i_0}, \mathbf{S}(G \setminus \{i_0\}))\right] \end{aligned}$$

Second, because each $\mathbf{y}$ can be decoded to a list of size at most L,

$$\begin{aligned} &\sum_{i\in G}\mathbb{E}\left[\bar{\varepsilon}_L(i, \mathbf{S}(G \setminus \{i\}))\right] \\ &\qquad \geq L + 1 - L\sum_{\mathbf{y}\in\mathcal{Y}^n} \mathbb{E}\left[W^n(\mathbf{y}|\mathbf{x}_{i_0}, \mathbf{S}(G \setminus \{i_0\}))\right] \\ &\qquad = 1 \qquad (42) \end{aligned}$$

We now bound the probability of error for this jamming strategy. The expected error, averaged over the random variable $\mathbf{J}$ and the randomly selected state sequence $\mathbf{S}(\mathbf{J})$, is

$$\mathbb{E}_{\mathbf{J},\mathbf{S}(\mathbf{J})}\left[\bar{\varepsilon}_L(\mathbf{S}(\mathbf{J}))\right] = \frac{1}{\binom{N}{L}}\frac{1}{N}\sum_{J\in\mathcal{J}}\sum_{i=1}^{N}\mathbb{E}\left[\bar{\varepsilon}_L(i, \mathbf{S}(J))\right]$$

Then

$$\begin{aligned} &\mathbb{E}_{\mathbf{J},\mathbf{S}(\mathbf{J})}\left[\bar{\varepsilon}_L(\mathbf{S}(\mathbf{J}))\right] \\ &\quad \geq \frac{1}{\binom{N}{L}}\frac{1}{N}\sum_{G\subset[N]:|G|=L+1}\sum_{i\in G}\mathbb{E}\left[\bar{\varepsilon}_L(i, \mathbf{S}(G \setminus \{i\}))\right] (43) \end{aligned}$$

Now, we can rewrite the inner sum using (42):

$$\begin{aligned} \mathbb{E}_{\mathbf{J},\mathbf{S}(\mathbf{J})}\left[\bar{\varepsilon}_L(\mathbf{S}(\mathbf{J}))\right] &\geq \frac{\binom{N}{L+1}}{\binom{N}{L}\cdot N} \\ &= \frac{1}{L+1} - \frac{L}{N(L+1)} \end{aligned}$$

Finally, we can add in the bound (40) to obtain

$$\begin{aligned}\frac{1}{L+1}-\frac{L}{N(L+1)} &\leq \mathbb{E}_{\mathbf{J},\mathbf{S}(\mathbf{J})}\left[\bar{\varepsilon}_L(\mathbf{S}(\mathbf{J}))\right] \\ &\leq \max_{\mathbf{s}\in\mathcal{S}^n(\Lambda)} \bar{\varepsilon}_L(\mathbf{s}) + \mathbb{P}\left(l(\mathbf{S}(\mathbf{J}))>\Lambda\right) \\ &\leq \max_{\mathbf{s}\in\mathcal{S}^n(\Lambda)} + \bar{\varepsilon}_L(\mathbf{s})\frac{4(\lambda^*)^2}{n\epsilon^2}\end{aligned}$$

Now, we can choose n_0 large enough such that

$$\max_{\mathbf{s}\in\mathcal{S}^n(\Lambda)} \bar{\varepsilon}_L(\mathbf{s}) > \frac{1}{L+2}-\frac{L}{N(L+1)}$$

■

Lemma 3: Let $\mathcal{W}$ be an AVC with state cost function $l(\cdot)$ and constraint Λ and let L be a positive integer. For any $\epsilon > 0$, there exists a $\nu(L,\mathcal{W},\epsilon) > 0$ and n_0 such that for any (n,N,L) list code (ϕ,ψ) with $n \geq n_0$ and $N > L+1$ whose codewords $\{\mathbf{x}(i) : i \in [N]\}$ satisfy

$$\lambda_L(T_{\mathbf{x}(i)}) < \Lambda - \epsilon \qquad \forall i \in [N] \tag{44}$$

the error must satisfy

$$\max_{\mathbf{s}\in\mathcal{S}^n(\Lambda)} \bar{\varepsilon}_L(\mathbf{s}) > \nu(L,\mathcal{W},\epsilon)$$

Proof: Fix $\epsilon > 0$. For each $P \in \mathcal{P}(\mathcal{X})$ from Lemma 1, we know there is a $\delta(P) > 0$ such that any joint distribution $\bar{P}$ with marginals within $\delta(P)$ of P can be approximated by a $\hat{P}$ with marginals equal to P such that $d_{\max}\left(\bar{P},\hat{P}\right) < \epsilon$.

Let

$$\mathcal{B}(P) = \{P' \in \mathcal{P}(\mathcal{X}) : d_{\max}(P,P') < \delta(P)\}$$

Then, $\{\mathcal{B}(P) : P \in \mathcal{P}(\mathcal{X})\}$ is an open cover of $\mathcal{P}(\mathcal{X})$. Since $\mathcal{P}(\mathcal{X})$ is compact, there is a constant r and finite subcover $\{\mathcal{B}(P_j) : j \in [r]\}$. From this finite cover, we can create a partition $\{A_j : j \in [r]\}$ of $\mathcal{P}$ such that $A_j \subseteq \mathcal{B}(P_j)$ for all j.

Now consider an (n,N,L) code whose codewords $\mathcal{C}$ satisfy (44). Let $F_j = \{i \in [N] : T_{\mathbf{x}(i)} \in A_j\}$. We can bound the error

$$\bar{\varepsilon}_L(\mathbf{s}) = \frac{1}{Nr}\sum_{j=1}^{r}\sum_{i\in F_j}\bar{\varepsilon}_L(i,\mathbf{s}) \geq \frac{|F_j|}{Nr}\left(\frac{1}{|F_j|}\sum_{i\in F_j}\bar{\varepsilon}_L(i,\mathbf{s})\right)$$

Since the collection $\{F_j\}$ partitions the codebook, for some j, we have $|F_j| \geq N/r$. From Lemma 2, the jammer can force the error to be lower bounded by

$$\max_{\mathbf{s}\in\mathcal{S}^n(\Lambda)} \bar{\varepsilon}_L(\mathbf{s}) \geq \frac{1}{r^2}\left(\frac{1}{L+1}-\frac{L}{N(L+1)}\right)$$

Note that the constant r is a function of ϵ, $\mathcal{W}$, and L, so we can set $\nu(L,\mathcal{W},\epsilon)$ to be this lower bound. For any $\epsilon > 0$, we have exhibited a jamming strategy such that the error is bounded away from 0. ■

Theorem 2 follows from the preceding lemma. Suppose that there exists a sequence of (n,N,L) codes $\{\mathcal{C}_n\}$ of rate

$$\sup_{P\in\mathcal{P}(\mathcal{X}):L_{\mathrm{sym}}(P,\Lambda)<L}\ \min_{Q\in\mathcal{Q}(\Lambda)} I\left(P,\sum_s W(y|x,s)Q(s)\right)+\delta$$

For each n, let $\mathcal{B}^- = \{\mathbf{x}(i) \in \mathcal{C}_n : L_{\mathrm{sym}}(T_{\mathbf{x}},\Lambda) < L\}$ and $\mathcal{B}^+ = \{\mathbf{x}(i) \in \mathcal{C}_n : L_{\mathrm{sym}}(T_{\mathbf{x}},\Lambda) \geq L\}$. Clearly, $|\mathcal{B}^-| \geq N/2$ or $|\mathcal{B}^+| \geq N/2$ or both. In the first case, the adversary can choose the state according to Q so that the channel is a DMC with transition probabilities $V_Q(y|x) = \sum_s W(y|x,s)Q(s)$. The rate of the subcode containing codewords $\mathbf{x}$ with $L_{\mathrm{sym}}(T_{\mathbf{x}},\Lambda) < L$ is greater than the mutual information $I(T_{\mathbf{x}},V_Q)$ for each $\mathbf{x}$, and therefore, the average error cannot converge to 0. In the second case, Lemma 3 shows that the average error is at least $\nu(L,\mathcal{W},\epsilon)/2$.

3) Achievability Under Average Error: Given a P such that the weak symmetrizing cost satisfies $\tilde{L}_{\mathrm{sym}}(P) > \Lambda$, we can use the coding scheme of Hughes [15] modified in the natural way suggested by Csiszár and Narayan [12] for list size 1. The technical issue is to prove that the decoding rule is unambiguous; that is, it should always produce a list of L or fewer codewords. The codebook consists of N constant-type codewords drawn uniformly from the codewords of type P. In order to describe the decoding rule we will use, we define the set

$$\begin{aligned}\mathcal{G}_\eta(\Lambda) = \{P_{XSY} \in \mathcal{P}(\mathcal{X}\times\mathcal{S}\times\mathcal{Y}) :&\\ D\left(P_{XSY}\ \|\ P_X \times P_S \times W\right) &\leq \eta,\\ \mathbb{E}[l(S)] &\leq \Lambda\}\end{aligned} \tag{45}$$

where

$$(P_X \times P_S \times W)(x,s,y) = P_X(x)P_S(s)W(y|x,s)$$

The set $\mathcal{G}_\eta(\Lambda)$ contains joint distributions which are close to those generated from the AVC $\mathcal{W}$ via independent inputs with distribution P_X and P_S.

Definition 1 (Decoding Rule): Let $\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_N$ be a given codebook and suppose $\mathbf{y}$ was received. Let $\psi(\mathbf{y})$ denote the list decoded from $\mathbf{y}$. Then, put $i \in \psi(\mathbf{y})$ if and only if there exists an $\mathbf{s} \in \mathcal{S}^n(\Lambda)$ such that

1) $T_{\mathbf{x}_i\mathbf{s}\mathbf{y}} \in \mathcal{G}_\eta(\Lambda)$,
2) for every set of L, other distinct codewords $\{\mathbf{x}_j : j \in J,\ J \subset [N]\setminus\{i\},\ |J| = L\}$ such that there exists a set $\{\mathbf{s}_j : \mathbf{s}_j \in \mathcal{S}^n(\Lambda),\ j \in J\}$ with $T_{\mathbf{x}_j\mathbf{s}_j\mathbf{y}} \in \mathcal{G}_\eta(\Lambda)$ for all $j \in J$ we have
$$I\left(YX \wedge X^L \middle| S\right) \leq \eta \tag{46}$$
where P_{YXX^LS} is the joint type of $(\mathbf{y},\mathbf{x}_i,\{\mathbf{x}_j : j \in J\},\mathbf{s})$.

An interpretation of this rule is that the decoder outputs a list of codewords $\{\mathbf{x}_i\}$ each having a "good explanation" $\{\mathbf{s}_i\}$. A "good explanation" is a state sequence that plausibly could have generated the observed output $\mathbf{y}$ (condition 1) and makes all other L-tuples of codewords seem independent of the codeword and output (condition 2). It is clear that this decoder will output a list containing the correct codeword with high probability. The only thing to prove is that the list size is no larger than L. To do this, we show that no tuple of random variables (Y, X^{L+1}, S^{L+1}) can satisfy the conditions of the decoding rule. This in turn shows that for sufficiently large n, no set of $L+1$*codewords* can satisfy the conditions of the decoding rule. Therefore, for sufficiently large blocklengths, the decoding rule will only output L or fewer codewords.

For a vector $x^{M+1} = (x_1, x_2, \ldots, x_{M+1})$, define $x^{M+1}_{-\{i\}} = (x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_{M+1})$ to be the vector x^{M+1} with the i-the component removed.

Lemma 4: Let $\beta > 0$, $\mathcal{W}$ be an AVC with state cost function $l(\cdot)$ and constraint Λ, $P \in \mathcal{P}(\mathcal{X})$ with $I(P, \Lambda) > 0$ and $\min_x P(x) \ge \beta$, and $M = \tilde{L}_{\mathrm{sym}}(P, \Lambda) + 1$. For any $\alpha > 0$ and every collection of distributions $\{U_i \in \mathcal{P}(\mathcal{X}^M \times \mathcal{S}) : i = 1, 2, \ldots, M+1\}$ such that

$$\sum_{x^{M+1}, s} P(x_i) U_i(x^{M+1}_{-\{i\}}, s) l(s) \le \tilde{\lambda}_M(P) - \alpha \tag{47}$$

for all $i = 1, 2, \ldots, M+1$, there exists a $\zeta > 0$ such that

$$\max_{j \ne i} \sum_{y, x^{M+1}} \Bigg| \sum_s W(y|x_i, s) U_i(x^{M+1}_{-\{i\}}, s) P(x_i) - \sum_s W(y|x_j, s) U_j(x^{M+1}_{-\{j\}}, s) P(x_j) \Bigg| \ge \zeta \tag{48}$$

Proof: Note that the outer sum in (48) is over all x^{M+1}. Define the function $V_k : \mathcal{X}^{M+1} \times \mathcal{S} \to \mathbb{R}$ by

$$V_k(x^{M+1}, s) = U_k(x^{M+1}_{-\{k\}}, s)$$

Let Π_{M+1} be the set of all permutations of $[M+1]$ and for $\pi \in \Pi_{M+1}$ let π_i be the image of i under π. Then

$$\begin{aligned} &\max_{j \ne i} \sum_{y, x^{M+1}} \Bigg| \sum_s W(y|x_i, s) V_i(x^{M+1}, s) P(x_i) - \sum_s W(y|x_j, s) V_j(x^{M+1}, s) P(x_j) \Bigg| \\ &= \max_{j \ne i} \sum_{y, x^{M+1}} \Bigg| \sum_s W(y|x_i, s) V_{\pi_i}(\pi(x^{M+1}), s) P(x_i) - \sum_s W(y|x_j, s) V_{\pi_j}(\pi(x^{M+1}), s) P(x_j) \Bigg| \end{aligned}$$

We can lower bound this by averaging over all $\pi \in \Pi_{M+1}$:

$$\max_{j \ne i} \sum_{y, x^{M+1}} \frac{1}{(M+1)!} \sum_{\pi \in \Pi_{M+1}} \Bigg| \sum_s W(y|x_i, s) V_{\pi_i}(\pi(x^{M+1}), s) P(x_i) - \sum_s W(y|x_j, s) V_{\pi_j}(\pi(x^{M+1}), s) P(x_j) \Bigg| \tag{49}$$

Define the average

$$\begin{aligned} \bar{V}(x^{M+1}_{-\{i\}}, s) &= \frac{1}{(M+1)!} \sum_{\pi \in \Pi_{M+1}} V_{\pi_i}(\pi(x^{M+1}), s) \\ &= \frac{1}{(M+1)!} \sum_{l=1}^{M+1} \sum_{\pi \in \Pi_{M+1} : \pi_i = l} U_l(\pi(x^{M+1})_{-\{\pi_i\}}, s) \\ &= \frac{1}{(M+1)!} \sum_{l=1}^{M+1} \sum_{\sigma \in \Pi_M} U_l(\sigma(x^{M+1}_{-\{i\}}), s) \end{aligned}$$

Note that for each $s \in \mathcal{S}$, $\bar{V}(x^{M+1}_{-\{i\}}, s)$ is a symmetric function of $x^{M+1}_{-\{i\}}$.

Now, we lower bound (49) by using the convexity of $|\cdot|$ to pull the averaging inside the absolute value and substituting $\bar{V}$. We arrive at the following expression:

$$F(\bar{V}, P) = \max_{j \ne i} \sum_{y, x^{M+1}} \Bigg| \sum_s W(y|x_i, s) \bar{V}(x^{M+1}_{-\{i\}}, s) P(x_i) - \sum_s W(y|x_j, s) \bar{V}(x^{M+1}_{-\{j\}}, s) P(x_j) \Bigg| \tag{50}$$

The function $F(\bar{V}, P)$ is a continuous function on the compact set of symmetric distributions $\{\bar{V}\}$ and the set of distributions P with $\min_x P(x) \ge \beta$, so it has a minimum $\zeta = F(\bar{V}^*, P^*)$ for some $(\bar{V}^*, P^*)$. We will prove that $\zeta > 0$ by contradiction.

Suppose $F(\bar{V}^*, P^*) = 0$. Then

$$\sum_s W(y|x_i, s) \bar{V}^*(x^{M+1}_{-\{i\}}, s) P^*(x_i) = \sum_s W(y|x_j, s) \bar{V}^*(x^{M+1}_{-\{j\}}, s) P^*(x_j)$$

So

$$\sum_y \sum_s W(y|x_i, s) \bar{V}^*(x^{M+1}_{-\{i\}}, s) P^*(x_i) = \sum_y \sum_s W(y|x_j, s) \bar{V}^*(x^{M+1}_{-\{j\}}, s) P^*(x_j)$$

and

$$\bar{V}^*(x^{M+1}_{-\{i\}}) P^*(x_i) = \bar{V}^*(x^{M+1}_{-\{j\}}) P^*(x_j),$$

which implies (see [15, Lemma A3]) that for all j:

$$\bar{V}^*(x^{M+1}_{-\{j\}}) P^*(x_j) = P^{*(M+1)}(x^{M+1})$$

Therefore

$$\sum_s W(y|x_1, s) \bar{V}^*(s|x_2^{M+1}) \tag{51}$$

is symmetric in $(x_1, x_2, \ldots, x_{M+1})$. Therefore, $\bar{V}^*(s|x_2^{M+1}) \in \mathcal{U}_{\mathrm{sym}}(M+1)$. From the definition of $\tilde{\lambda}_M(P)$ in (8), we see that

$$\sum_{x^{M+1}, s} \bar{V}^*(x_{-\{i\}}^{M+1}, s) P(x_i) l(s) \geq \tilde{\lambda}_M(P)$$

But from (47), and the definition of $\bar{V}$, we see that the $\{U_i\}$ must be chosen such that

$$\sum_{x^{M+1}, s} \bar{V}^*(x_{-\{i\}}^{M+1}, s) P(x_i) l(s) \leq \tilde{\lambda}_M(P) - \alpha \qquad (52)$$

Therefore, we have a contradiction and the minimum ζ of $F(\bar{V}, P)$ must be greater than 0. Equation (48) follows. ■

The next lemma shows that for a sufficiently small choice of the threshold η in the decoding rule, there are no random variables that can force the decoding rule to output a list that is too large. The proof follows from Lemma 4 in the same way as in [15].

Lemma 5: Let $\beta > 0$, $\mathcal{W}$ be an AVC with state cost function $l(\cdot)$ and constraint Λ, $P \in \mathcal{P}(\mathcal{X})$ with $\min_x P(x) \geq \beta$, and $M = \tilde{L}_{\mathrm{sym}}(P, \Lambda) + 1$. Then, there exists an $\eta > 0$ sufficiently small such that no tuple of random variables (Y, X^{M+1}, S^{M+1}) can simultaneously satisfy

$$\min_x P(x) \geq \beta \qquad (53)$$

$$P_{X_i} = P \qquad (54)$$

$$P_{YX_iS_i} \in \mathcal{G}_\eta(\Lambda) \qquad (55)$$

$$I\left(YX_i \wedge X_{-\{i\}}^{M+1} \middle| S_i\right) \leq \eta, \quad 1 \leq i \leq M+1 \qquad (56)$$

Given Lemma 5, the following lemma shows that given an input distribution P and a list size $M = \tilde{L}_{\mathrm{sym}}(P, \Lambda) + 1$, there exists a list code with list size L and small error probability.

Lemma 6 (see [15, Lemma 3]): Let P be a type satisfying $\min_x P(x) \geq \beta$ and let $L = \tilde{L}_{\mathrm{sym}}(P, \Lambda) + 1$. For any $\delta > 0$, there exists a list code of list size L with codewords of constant type P such that

$$\frac{1}{n} \log \left(\frac{N}{L} \right) > I(P, \Lambda) - \delta, \qquad \bar{\varepsilon}_L \leq \exp(-n\gamma)$$

for all $n \geq n_2$, where $\gamma > 0$ and n_2 depend only on β, δ, and $\mathcal{W}$.

The code in Lemma 6 is a code whose codewords are all of a constant type P. This lemma is proved in [15] by selecting N codewords uniformly from the set of codewords with constant composition P and showing that with high probability, the result codebook satisfies a set of joint typicality and conditional joint typicality conditions universally over all state sequences $\mathbf{s}$.

Proof of Theorem 3: Lemma 4 implies Lemma 5, which allows us to use Lemma 6 to show that a code exists with rate close to $I(P, \Lambda)$ and small error. Since $I(P, \Lambda)$ is continuous in P, for a fixed list size L, the rate $I(P, \Lambda)$ is achievable for all P such that $\tilde{L}_{\mathrm{sym}}(P, \Lambda) < L$. ■

Acknowledgment

The authors would like to thank the anonymous reviewers for their careful and detailed comments, Prof. Venkat Anantharam for his help in clarifying the arguments, and the associate editor, Prof. Hirosuke Yamamoto, for his kind guidance through the review process.

References

[1] R. Ahlswede, "A note on the existence of the weak capacity for channels with arbitrarily varying channel probability functions and its relation to Shannon's zero-error capacity," *Ann. Math. Statist.*, vol. 41, no. 3, pp. 1027–1033, 1970.

[2] R. Ahlswede, "Channel capacities for list codes," *J. Appl. Probabil.*, vol. 10, no. 4, pp. 824–836, 1973.

[3] R. Ahlswede, "Elimination of correlation in random codes for arbitrarily varying channels," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 44, no. 2, pp. 159–175, 1978.

[4] R. Ahlswede, "The maximal error capacity of arbitrarily varying channels for constant list sizes," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1416–1417, Jul. 1993.

[5] D. Blackwell, L. Breiman, and A. Thomasian, "The capacities of certain channel classes under random coding," *Ann. Math. Statist.*, vol. 31, pp. 558–567, 1960.

[6] V. Blinovsky, P. Narayan, and M. Pinsker, "Capacity of the arbitrarily varying channel under list decoding," *Probl. Inf. Transmiss.*, vol. 31, no. 2, pp. 99–113, 1995.

[7] V. Blinovsky and M. Pinsker, G. Cohen, S. Litsyn, A. Lobstein, and G. Zémor, Eds., "Estimation of the size of the list when decoding over an arbitrarily varying channel," in *Proc. 1st French-Israeli Workshop Algebraic Coding*, Berlin, Germany, Jul. 1993, pp. 28–33.

[8] T. Cover and J. Thomas*, Elements of Information Theory*. New York: Wiley, 1991.

[9] I. Csiszár and J. Körner, "On the capacity of the arbitrarily varying channel for maximum probability of error," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 57, pp. 87–101, 1981.

[10] I. Csiszár and J. Körner*, Information Theory: Coding Theorems for Discrete Memoryless Systems*. Budapest, Hungary: Akadémi Kiadó, 1982.

[11] I. Csiszár and P. Narayan, "Arbitrarily varying channels with constrained inputs and states," *IEEE Trans. Inf. Theory*, vol. 34, no. 1, pp. 27–34, Jan. 1988.

[12] I. Csiszár and P. Narayan, "The capacity of the arbitrarily varying channel revisited: Positivity, constraints," *IEEE Trans. Inf. Theory*, vol. 34, no. 2, pp. 181–193, Mar. 1988.

[13] D. P. Dubhashi and A. Panconesi*, Concentration of Measure for the Analysis of Randomized Algorithms*. Cambridge, U.K.: Cambridge Univ. Press, 2009.

[14] T. Ericson, "Exponential error bounds for random codes on the arbitrarily carying channel," *IEEE Trans. Inf. Theory*, vol. 31, no. 1, pp. 42–48, Jan. 1985.

[15] B. Hughes, "The smallest list for the arbitrarily varying channel," *IEEE Trans. Inf. Theory*, vol. 43, no. 3, pp. 803–815, May 1997.

[16] M. Langberg, "Private codes or succinct random codes that are (almost) perfect," in *Proc. 45th Annu. IEEE Symp. Found. Comput. Sci.*, Rome, Italy, 2004, pp. 325–334.

[17] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *IEEE Trans. Inf. Theory*, vol. 44, no. 10, pp. 2148–2177, Oct. 1998.

[18] S. Nishimura, "The strong converse theorem in the decoding scheme of list size L," *Kodai Math. Semin. Rep.*, vol. 21, no. 4, pp. 418–425, 1969.

[19] S. Nitinawarat, "On the deterministic code capacity region of an arbitrarily varying multiple-access channel under list decoding," in *Proc. IEEE Int. Symp. Inf. Theory*, Austin, TX, Jun. 2010, pp. 290–294.

[20] A. Sarwate and M. Gastpar, "Rateless codes for AVC models," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3105–3114, Jul. 2010.

[21] S. S. Shitz and S. Verdú, "The empirical distribution of good codes," *IEEE Trans. Inf. Theory*, vol. 43, no. 3, pp. 836–846, May 1997.

[22] J. Wolfowitz, "Simultaneous channels," *Arch. Rational Mechan. Anal.*, vol. 4, no. 4, pp. 378–386, 1960.

Anand D. Sarwate (S'99–M'09) received B.S. degrees in electrical engineering and computer science and mathematics from the Massachusetts Institute of Technology (MIT), Cambridge, in 2002 and the M.S. and Ph.D. degrees in electrical engineering in 2005 and 2008, respectively, from the University of California, Berkeley. From 2008–2011 he was a postdoctoral researcher at the Information Theory and Applications Center at the University of California, San Diego.

Since October 2011, he is a Research Assistant Professor at the Toyota Technological Institute at Chicago. His research interests include information theory, distributed signal processing, machine learning, and privacy.

Dr. Sarwate received the Laya and Jerome B. Wiesner Student Art Award from MIT, and the Samuel Silver Memorial Scholarship Award and Demetri Angelakos Memorial Achievement Award from the EECS Department at University of California at Berkeley. He was awarded an NDSEG Fellowship from 2002 to 2005. He is a member of Phi Beta Kappa and Eta Kappa Nu.

Michael Gastpar received the Dipl. El.-Ing. degree from the Swiss Federal Institute of Technology (ETH), Zurich, in 1997, the M.S. degree from the University of Illinois at Urbana-Champaign, Urbana, in 1999, and the Doctoratès Science degree from the Swiss Federal Institute of Technology (EPFL), Lausanne, in 2002, all in electrical engineering. He was also a student in engineering and philosophy at the Universities of Edinburgh and Lausanne.

He is currently a Professor in the School of Computer and Communication Sciences,École Polytechnique Fédérale (EPFL), Lausanne, Switzerland, and an Adjunct Associate Professor with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley. He also holds a faculty position at Delft University of Technology, The Netherlands, and he was a Researcher with the Mathematics of Communications Department, Bell Labs, Lucent Technologies, Murray Hill, NJ. His research interests are in network information theory and related coding and signal processing techniques, with applications to sensor networks and neuroscience. Dr. Gastpar won the 2002 EPFL Best Thesis Award, an NSF CAREER Award in 2004, an Okawa Foundation Research Grant in 2008, and an ERC Starting Grant in 2010.

He is an Information Theory Society Distinguished Lecturer (2009–2011). He was an Associate Editor for Shannon Theory for the IEEE TRANSACTIONS ON INFORMATION THEORY (2008–2011), and he has served as Technical Program Committee Co-Chair for the 2010 International Symposium on Information Theory, Austin, TX.