Finite Blocklength and Dispersion Bounds for the Arbitrarily-Varying Channel

Oliver Kosut and Jörg Kliewer

***Abstract*—Finite blocklength and second-order (dispersion) results are presented for the arbitrarily-varying channel (AVC), a classical model wherein an adversary can transmit arbitrary signals into the channel. A novel finite blocklength achievability bound is presented, roughly analogous to the random coding union bound for non-adversarial channels. This finite blocklength bound, along with a known converse bound, is used to derive bounds on the dispersion of discrete memoryless AVCs without shared randomness, and with cost constraints on the input and the state. These bounds are tight for many channels of interest, including the binary symmetric AVC. However, the bounds are not tight if the deterministic and random code capacities differ.**

I. Introduction

Active, malicious adversaries represent a potential threat against modern communication systems. This is particularly true of wireless systems, in which the inherently open nature of the communication medium allows for an intelligent jammer to transmit a damaging signal. The arbitrarily-varying channel (AVC) is a classical information-theoretic model that captures an active adversary in a point-to-point setting. Classical work on the AVC characterized the capacity with and without shared randomness between the encoder and decoder, and in which the input and state (or adversarial signal) are subject to cost constraints.

In this paper, we present finite blocklength and second-order results for the AVC under average probability of error and *without* shared randomness, including cases with cost constraints. We introduce a novel finite blocklength achievability bound, which is a strengthened form of the achievability bound used in [1] to derive the AVC capacity without shared randomness. We further show that in some cases, this achievability bound is strong enough to achieve both the capacity and the dispersion of discrete memoryless AVCs. The *dispersion* characterizes the asymptotic second-order behavior of a channel subject to a fixed probability of error constraint. Analysis of this sort dates back to Strassen [2], and has seen significant interest in recent years, particularly since [3]. The dispersion of the compound channel, which is closely related to the AVC—in fact, they are indistinguishable in the single-shot setting (see Remark 1)—was derived for discrete memoryless channels in [4]. We found the dispersion of AVCs with shared randomness between encoder and decoder in our prior work [5], although this result did not extend to channels with cost constraints. In the present paper, we provide the exact dispersion of discrete memoryless AVCs without shared randomness, and with or without cost constraints, provided certain conditions are satisfied. These conditions are satisfied for some channels of interest, such as binary symmetric AVCs, but not others, including parts of the parameter space for the binary adding AVC.

In the interest of space, several proofs have been omitted or abbreviated; see [6] for a full version of this paper.

O. Kosut is with the School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ 85287 USA (email: okosut@asu.edu).

J. Kliewer is with the Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, NJ 07102 USA (email: jkliewer@njit.edu).

This material is based upon work supported by the National Science Foundation under Grant No. CCF-1439465, CCF-1440014, CNS-1526547, CCF-1453718.

II. Preliminaries

A. Notation

Given a set $\mathcal{X}$, let $\mathcal{P}(\mathcal{X})$ be the set of random distributions with alphabet $\mathcal{X}$. For some $P \in \mathcal{P}(\mathcal{X})$, we write $X \sim P$ to mean that X is a random variable drawn from distribution P. The probability measure is denoted $\mathbb{P}$, and the expectation operator is denoted $\mathbb{E}$; the underlying distribution will be specified in context. Given a function $g : \mathcal{X} \to \mathbb{R}$ and a real number Γ, let $\mathcal{P}(\mathcal{X}, \Gamma)$ be the set of distributions $P \in \mathcal{P}(\mathcal{X})$ where $\mathbb{E}g(X) \le \Gamma$ if $X \sim P$. The underlying function g will be understood from the context. Also let

$$\mathcal{X}^n(\Gamma) = \{x^n \in \mathcal{X}^n : \textstyle\sum_{i=1}^n g(x_i) \le n\Gamma\}. \tag{1}$$

For alphabet $\mathcal{S}$, function $\ell : \mathcal{S} \to \mathbb{R}$ and real number Λ, we define $\mathcal{P}(\mathcal{S}, \Lambda)$ and $\mathcal{S}^n(\Lambda)$ similarly. Let $\mathcal{P}(\mathcal{Y}|\mathcal{X})$ be the set of conditional distributions $P_{Y|X}$ where $P_{Y|X}(\cdot|x) \in \mathcal{P}(\mathcal{Y})$ for all $x \in \mathcal{X}$. For any $P_X \in \mathcal{P}(\mathcal{X})$ and $P_{Y|X} \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$, we write $P_X P_{Y|X} \in \mathcal{P}(\mathcal{Y})$ where

$$(P_X P_{Y|X})(y) = \sum_{x \in \mathcal{X}} P_X(x) P_{Y|X}(y|x). \tag{2}$$

Similarly, given $P_S \in \mathcal{P}(\mathcal{S})$ and $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X} \times \mathcal{S})$, let $P_S W \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$ be given by

$$(P_S W)(y|x) = \sum_{s \in \mathcal{S}} P_S(s) W(y|x, s). \tag{3}$$

Note that $P_X P_S W \in \mathcal{P}(\mathcal{Y})$ is now also well defined. Given $P_X \in \mathcal{P}(\mathcal{X})$ or $P_{Y|X}(\mathcal{Y}|\mathcal{X})$, and any positive integer n, we write their stationary-memoryless extensions as $P_X^n(x^n) = \prod_{i=1}^n P_X(x_i)$ and $P_{Y|X}^n(y^n|x^n) = \prod_{i=1}^n P_{Y|X}(y_i|x_i)$. Given a sequence $x^n \in \mathcal{X}^n$, its *type* is $Q_{x^n}(x) = \frac{1}{n}|\{i : x_i = x\}|$. Let $\mathcal{P}_n(\mathcal{X})$ be the set of all types of sequences in $\mathcal{X}^n$. For $P \in \mathcal{P}_n(\mathcal{X})$, let $T(P)$ be the type class of P; i.e., the set of sequences $x^n \in \mathcal{X}^n$ with $Q_{x^n} = P$. Also, for $P \in \mathcal{P}_n(\mathcal{X})$,

978-1-5386-4780-6/18/$31.00©2018 IEEE

let U_{P_X} be the uniform distribution over type class $T(P)$. For any integer M, we write $[M] = \{1, \ldots, M\}$. Finally, log and exp are assumed to have base 2.

B. Problem Description

We first describe a single-shot AVC model, with the input, state, and output alphabets having arbitrary structure, and the channel itself represented by an arbitrary conditional probability measure. Subsequently, we specialize the model to the n-length stationary memoryless case.

A single-shot AVC is given by the tuple $(\mathcal{X}, \mathcal{S}, W(y|x,s), \mathcal{Y})$ where $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X} \times \mathcal{S})$. An (M, ϵ) code is given by an encoding function $\phi : [M] \to \mathcal{X}$ and a decoding function $\psi : \mathcal{Y} \to [M]$ where for any $s \in \mathcal{S}$, the average probability of error is at most ϵ; i.e.

$$\sup_{s \in \mathcal{S}} \frac{1}{M} \sum_{m=1}^{M} W(\psi^{-1}(m)^c | \phi(m), s) \leq \epsilon \tag{4}$$

where $\psi^{-1}(m)^c$ is the set of $y \in \mathcal{Y}$ such that $\psi(y) \neq m$. Let $M^\star(\epsilon)$ be the largest integer M for which there exists an (M, ϵ) code.

Given cost functions $g : \mathcal{X} \to \mathbb{R}$ and $\ell : \mathcal{S} \to \mathbb{R}$, an n-length cost-constrained AVC is given by the tuple

$$(\mathcal{X}^n(\Lambda), \mathcal{S}^n(\Gamma), W^n(y^n|x^n, s^n), \mathcal{Y}^n). \tag{5}$$

where Λ, Γ are real numbers. An (M, n, ϵ) code consists of a code for this channel with M messages and probability of error ϵ. Define $M^\star(n, \epsilon)$ similarly.

Remark 1: While in this paper we are primarily interested in the AVC, the above single-shot model is indistinguishable from a compound channel model, which differs from an AVC only in that the state must be held constant across the coding block, a distinction that only makes sense in the n-length setting. In fact, our finite blocklength achievability bound Thm. 1, which applies in the general single-shot setting, may be considered as an achievable bound for the compound channel as well as the AVC.

III. Finite Blocklength Achievability Bound

The following theorem is our new achievability bound for the AVC. As we will illustrate below, this bound is analogous to the random coding union (RCU) bound for non-state channels, as derived in [3].

Theorem 1: Fix P_X, and let $Z(x, \bar{x}, y) \in \{0, 1\}$ be a test such that

$$Z(x, \bar{x}, y) Z(\bar{x}, x, y) = 0 \text{ for all } x, \bar{x} \in \mathcal{X},\ y \in \mathcal{Y} \tag{6}$$

and let $\mathcal{A} \subseteq \mathcal{X} \times \mathcal{Y}$. For each $s \in \mathcal{S}$, let $(X, \bar{X}, Y_s) \sim P_X(x) P_X(\bar{x}) W(y|x,s)$. There exists an (M, ϵ) code such that

$$\begin{aligned} \epsilon \leq & \max_s \mathbb{P}((X, Y_s) \notin \mathcal{A}) \\ & + (2 \log e) M \, \mathbb{P}(Z(X, \bar{X}, Y_s) = 0,\ (X, Y_s) \in \mathcal{A}) \\ & + \operatorname{ess\,sup} 2 \log(3|\mathcal{S}|) \, \mathbb{P}(Z(X, \bar{X}, Y_s) = 0,\ (X, Y_s) \in \mathcal{A} | \bar{X}) \\ & + \sqrt{\frac{2 \ln(3|\mathcal{S}|)}{M}}. \end{aligned} \tag{7}$$

The test Z can be viewed as a test for whether x is more likely than $\bar{x}$ to be the transmitted codeword, given that y has been received by the decoder. Specifically, the proof of Thm. 1 uses the following decoding rule for codebook $\{c_1, \ldots, c_M\}$: *Given output y, decode to message i if $Z(c_i, c_j, y) = 1$ for all $j \neq i$. If there is no such message, declare an error.* Note that condition (6) ensures that two messages cannot simultaneously satisfy this criterion. The set $\mathcal{A}$ is arbitrary at this point; in our proof of the dispersion in Sec. IV we define it to be a jointly typical set of input-output pairs.

Remark 2: From Thm. 1, one can recover a bound similar to the RCU bound of [3] as follows. Given a channel without state (i.e., $|\mathcal{S}| = 1$), we may choose

$$Z(x, \bar{x}, y) = \mathbf{1}\big(\imath(x; y) > \imath(\bar{x}; y)\big) \tag{8}$$

where $\imath(x; y)$ is the information density. This test clearly satisfies (6). One can now see that the optimal choice for $\mathcal{A}$ to minimize the first two terms in (7) is

$$\mathcal{A} = \{(x, y) : (2 \log e) M \, \mathbb{P}(\imath(\bar{X}, y) \geq \imath(x; y)) \leq 1\}. \tag{9}$$

Thus the first two terms in (7) become

$$\mathbb{E} \min\big\{1, (2 \log e) M \, \mathbb{P}\big(\imath(\bar{X}, Y) \geq \imath(X; Y) \big| X, Y\big)\big\}. \tag{10}$$

This expression is nearly identical to the standard RCU bound, except that $M - 1$ has been replaced by $(2 \log e) M$. This difference constitutes less than 2 bits. Furthermore, the last two terms in (7) are vanishingly small. These terms appear as a consequence of a Chernoff bound (Lemma 2) being applied in the achievability proof. As outlined below, the Chernoff bound is applied twice; each application produces one of the two extra terms in (7).

The proof of Thm. 1 relies on the following lemma, which is a sharpened version of [1, Lemma A1]. The lemma is a Chernoff bound that holds even for variables that are not i.i.d., provided they have a bounded conditional expectation.

Lemma 2: Let $X_1, \ldots, X_M$ be random variables and let $f_i(x_1, \ldots, x_i)$ be a set of M functions where

$$\mathbb{E}[f_i(X_1, \ldots, X_i) | X_1, \ldots, X_{i-1}] \leq \mu \quad \text{a.s.} \tag{11}$$

and $f_i(X_1, \ldots, X_i) \in [0, \gamma]$ a.s. Then for all $t \in [\mu, \gamma]$,

$$\begin{aligned} \mathbb{P}\left(\frac{1}{M} \sum_{i=1}^{M} f_i(X_1, \ldots, X_i) > t \right) \\ < \min\left\{ 2^{-M\left(\frac{t-\mu}{\gamma} \log_2 e\right)}, e^{-2M\left(\frac{t-\mu}{\gamma}\right)^2} \right\}. \end{aligned} \tag{12}$$

Proof of Thm. 1: Applying the decoding rule described above, given codebook $\{c_1, \ldots, c_M\}$ and state s, the average probability of error is

$$\begin{aligned} & P_e(c_1, \ldots, c_M | s) \\ & \quad = \frac{1}{M} \sum_i \mathbb{P}\big(Z(c_i, c_j, Y_s) = 0 \text{ for some } j \neq i \big| X = c_i\big). \end{aligned} \tag{13}$$

Let $\mathcal{A}$ be an arbitrary subset of $\mathcal{X} \times \mathcal{Y}$. We may upper bound the probability of error by

$$P_e(c_1, \ldots, c_M|s) \le \frac{1}{M} \sum_i \Bigg[\mathbb{P}\Big((c_i, Y_s) \notin \mathcal{A} \text{ or } Z(c_i, c_j, Y_s) = 0 \text{ for some } j < i \Big) + \mathbb{P}\Big((c_i, Y_s) \in \mathcal{A},\, Z(c_i, c_j, Y_s) = 0 \text{ for some } j > i \Big) \Bigg]. \tag{14}$$

Let $C_1, \ldots, C_M$ be independent random variables, each drawn from P_X. We proceed to show that with some positive probability, $P_e(C_1, \ldots, C_M|s)$ exceeds the quantity in the RHS of (7) for all $s \in \mathcal{S}$. Let

$$q(\bar{x}, s) = \mathbb{P}(Z(X, \bar{x}, Y_s) = 0 | (X, Y_s) \in \mathcal{A}). \tag{15}$$

Now let $f_i(c_1, \ldots, c_i|s) = 0$ if $\sum_{j<i} q(c_j, s) > M t_{1s}$ (where t_{1s} is a constant to be determined), and otherwise

$$f_i(c_1, \ldots, c_i|s) = \mathbb{P}\Big((c_i, Y_s) \notin \mathcal{A} \text{ or } Z(c_i, c_j, Y_s) = 0 \text{ for some } j < i \Big). \tag{16}$$

Similarly, let $g_i(c_i, \ldots, c_M|s) = 0$ if $\sum_{j>i} q(c_j, s) > M t_{1s}$, and otherwise

$$g_i(c_i, \ldots, c_M|s) = \mathbb{P}\Big((c_i, Y_s) \in \mathcal{A},\, Z(c_i, c_j, Y_s) = 0 \text{ for some } j > i \Big). \tag{17}$$

We now define three classes of error events (again t_{2s}, t_{3s} are to be determined):

$$\mathcal{E}_{1s} = \left\{ \frac{1}{M} \sum_i q(C_i, s) > t_{1s} \right\}, \tag{18}$$

$$\mathcal{E}_{2s} = \left\{ \frac{1}{M} \sum_i f_i(C_1, \ldots, C_i|s) > t_{2s} \right\}, \tag{19}$$

$$\mathcal{E}_{3s} = \left\{ \frac{1}{M} \sum_i g_i(C_i, \ldots, C_M|s) > t_{3s} \right\}. \tag{20}$$

Note that if $\mathcal{E}_{1s}$ does not occur, then RHS of (14) is equal to $\frac{1}{M} \sum_i [f_i(c_1, \ldots, c_M|s) + g_i(c_1, \ldots, c_M|s)]$. We proceed to find constants t_{1s}, t_{2s}, t_{3s} such that the probability that each of these events is less than $(3|\mathcal{S}|)^{-1}$, thus proving that there exists at least one code that does not fall into any of these events. Define

$$\alpha_s = \mathbb{E} q(\bar{X}, s) = \mathbb{P}(Z(X, \bar{X}, Y_s) = 0 | (X, Y_s) \in \mathcal{A}), \tag{21}$$

$$\gamma_s = \operatorname{ess\,sup} q(\bar{X}, s). \tag{22}$$

Note that in (22), the essential supremum corresponds to a supremum over the support set of $\bar{X}$. If we choose

$$t_{1s} = \alpha_s \log e + \frac{\gamma_s \log(3|\mathcal{S}|)}{M}. \tag{23}$$

then by Lemma 2

$$\mathbb{P}(\mathcal{E}_{1s}) = \mathbb{P}\left(\frac{1}{M} \sum_i q(C_i, s) > t_{1s} \right) < 2^{-M\left(\frac{t_{1s} - \alpha_s \log e}{\gamma_s}\right)} = (3|\mathcal{S}|)^{-1}. \tag{24}$$

If $\sum_{j<i} q(c_j, s) \le M t_{1s}$ then for any fixed $c_1, \ldots, c_{i-1}$,

$$\mathbb{E} f_i(c_1, \ldots, c_{i-1}, C_i) \tag{25}$$

$$\le \mathbb{P}((X, Y_s) \notin \mathcal{A}) + \sum_{j<i} \mathbb{P}(Z(X, c_j, Y_s) = 0,\, (X, Y_s) \in \mathcal{A}) \tag{26}$$

$$= \mathbb{P}((X, Y_s) \notin \mathcal{A}) + \mathbb{P}((X, Y_s) \in \mathcal{A}) \sum_{j<i} q(c_j, s) \tag{27}$$

$$\le \mathbb{P}((X, Y_s) \notin \mathcal{A}) + \mathbb{P}((X, Y_s) \in \mathcal{A}) M t_{1s}. \tag{28}$$

Moreover, the upper bound in (28) holds for all $(c_1, \ldots, c_{i-1})$, since when $\sum_{j<i} q(c_j, s) > M t_{1s}$ the function is identically zero. If we choose

$$t_{2s} = \mathbb{P}((X, Y_s) \notin \mathcal{A}) + \mathbb{P}((X, Y_s) \in \mathcal{A}) M t_{1s} + \sqrt{\frac{\ln(3|\mathcal{S}|)}{2M}}. \tag{29}$$

then by Lemma 2 and the fact that $f_i \in [0, 1]$,

$$\mathbb{P}(\mathcal{E}_{2s}) = \mathbb{P}\left(\frac{1}{M} \sum_i f_i(C_1, \ldots, C_i) > t_{2s} \right) \tag{30}$$

$$< e^{-2M(t_{2s} - \mathbb{P}((X,Y_s) \notin \mathcal{A}) + \mathbb{P}((X,Y_s) \in \mathcal{A}) M t_{1s})^2} = (3|\mathcal{S}|)^{-1}. \tag{31}$$

By a similar argument, $\mathbb{P}(\mathcal{E}_{3s}) < (3|\mathcal{S}|)^{-1}$ if

$$t_{3s} = \mathbb{P}((X, Y_s) \in \mathcal{A}) M t_{1s} + \sqrt{\frac{\ln(3|\mathcal{S}|)}{2M}}. \tag{32}$$

Therefore, there exists a codebook $\{c_1, \ldots, c_M\}$ falling into no error events for any s. In particular, since $\mathcal{E}_{1s}$ does not occur, the functions f_i, g_i are equal to the expressions in (16)–(17) (rather than zero), so we may rewrite the RHS of (13) to conclude that for all s, the probability of error is at most $t_{2s} + t_{3s}$. Applying the definitions of $t_{2s}, t_{3s}, \alpha_s, \gamma_s$ yields (7). ∎

IV. Dispersion Bounds

Consider an n-length cost-constrained AVC with finite alphabets, given by the single-letter conditional distribution $W(y|x, s)$. Given $P_X \in \mathcal{P}(\mathcal{X})$ and $P_S \in \mathcal{P}(\mathcal{S})$, let $(X, S, Y) \sim P_X(x) P_S(s) W(y|x, s)$. Now we define the following information quantities:

$$\imath(x; y) = \log \frac{(P_S W)(y|x)}{(P_X P_S W)(y)}, \tag{33}$$

$$I(P_X, P_{Y|X}) = \mathbb{E} \imath_{P_{Y|X} \| (P_X P_{Y|X})}(X; Y), \tag{34}$$

$$\tilde{\imath}(x; s; y) = \imath(x; y) - \mathbb{E}(\imath(X; Y)|X = x) - \mathbb{E}(\imath(X; Y)|S = s) + I(P_X, P_S W), \tag{35}$$

$$V(P_X, P_S, W) = \mathbb{E}\, \tilde{\imath}(X; S; Y)^2, \tag{36}$$

$$T(P_X, P_S, W) = \mathbb{E}\, |\tilde{\imath}(X; S; Y)|^3. \tag{37}$$

For any $P_X \in \mathcal{P}(\mathcal{X})$, let

$$\Lambda_0(P_X) = \min_{P_{S|X}} \sum_{x \in \mathcal{X},\, s \in \mathcal{S}} P_X(x) P_{S|X}(s|x) \ell(s) \tag{38}$$

where the minimum is over distributions $P_{S|X} \in \mathcal{P}(\mathcal{S}|\mathcal{X})$ such that, for all x, x' in the support of P_X,

$$\sum_s P_{S|X}(s|x)W(y|x',s) = \sum_s P_{S|X}(s|x')W(y|x,s), \quad (39)$$

and $\Lambda_0(P_X) = \infty$ if there is no distribution satisfying (39). An AVC is said to be *symmetrizable* if $\Lambda_0(P_X) \le \Lambda$ for all $P_X \in \mathcal{P}(\mathcal{X}, \Gamma)$, in which case the capacity is zero. For non-symmetrizable AVCs, the capacity was found in [1] to be

$$C = \max_{\substack{P_X \in \mathcal{P}(\mathcal{X},\Gamma):\\ \Lambda_0(P_X) \ge \Lambda}} \min_{P_S \in \mathcal{P}(\mathcal{S},\Lambda)} I(P_X, P_S W). \quad (40)$$

Note that the feasible sets for both the maximum and minimum in (40) are convex sets. Moreover, mutual information is concave in the input distribution and convex in the channel distribution, so the maximum and minimum in (40) can be exchanged without changing the value. We may define $\Pi_X(\Gamma)$ and $\Pi_S(\Lambda)$ to be the sets of optimal distributions for P_X and P_S respectively. Let

$$V_+ = \min_{P_X \in \Pi_X(\Gamma)} \max_{P_S \in \Pi_S(\Lambda)} V(P_X, P_S, W) \quad (41)$$

For a cost-constrained AVC, the *random code capacity*—defined as the capacity when the encoder and decoder have access to an unlimited amount of shared randomness, unknown to the adversary—is given by [7]

$$C_r = \max_{P_X \in \mathcal{P}(\mathcal{X},\Gamma)} \min_{P_S \in \mathcal{P}(\mathcal{S},\Lambda)} I(P_X, P_S W). \quad (42)$$

Let $\Pi_X^{(r)}(\Gamma)$ and $\Pi_S^{(r)}(\Lambda)$ be the set of optimal distributions for P_X and P_S in (42). Let

$$V_- = \max_{P_S \in \Pi_S^{(r)}(\Lambda)} \min_{P_X \in \Pi_X^{(r)}(\Gamma)} V(P_X, P_S, W). \quad (43)$$

Let Q be the complementary CDF of the standard Gaussian distribution, and Q^{-1} its inverse.

The following theorems give upper and lower bounds on the normal approximation for discrete-memoryless AVCs.

Theorem 3: Consider an n-length, cost-constrained AVC. For any $\epsilon \in (0, 1/2)$,

$$\log M^\star(n,\epsilon) \le nC_r - \sqrt{nV_-}\,\mathsf{Q}^{-1}(\epsilon) + (|\mathcal{X}| + |\mathcal{S}| - \tfrac{3}{2}) \log n + O(1). \quad (44)$$

Theorem 4: Consider a cost-constrained AVC for which there exists a distribution $P_X^\star \in \Pi_X(\Gamma)$ that achieves the minimum in (41) such that $\Lambda_0(P_X^\star) > \Lambda$. Then for any $\epsilon \in (0, 1/2)$,

$$\log M^\star(n,\epsilon) \ge nC - \sqrt{nV_+}\,\mathsf{Q}^{-1}(\epsilon) - (|\mathcal{X}| + |\mathcal{S}| - \tfrac{3}{2}) \log n - O(1). \quad (45)$$

While our bounds do not match even to first order when the random code capacity exceeds the capacity, the following corollary gives a sufficient condition for the bounds to hold up to second order.

Corollary 5: Consider a cost-constrained non-symmetrizable AVC such that: (i) there exists a distribution $P_X^\star \in \Pi^{(r)}(\Lambda)$ where $\Lambda_0(P_X^\star) > \Lambda$, and (ii) at least one of the sets $\Pi_X(\Gamma)$ and $\Pi_S(\Lambda)$ contain only a single element. Then $C_r = C$, $V_+ = V_-$, and for any $\epsilon \in (0, 1/2)$,

$$\log M^\star(n,\epsilon) = nC - \sqrt{nV_+}\,\mathsf{Q}^{-1}(\epsilon) + O(\log n). \quad (46)$$

We now consider two examples, illustrating cases in which the sufficient condition in Corollary 5 does or does not hold. The capacity of both of these examples was originally found in [1].

Example 1 (Binary symmetric AVC): Let $\mathcal{X}, \mathcal{Y}, \mathcal{S} = \{0, 1\}$, and $W(y|x,s) = 1$ if $y = x \oplus s$, where $\oplus$ is addition modulo 2. Let $g(x) = x$ and $\ell(s) = s$. If $P_X = [1-p, p]$, then $\Lambda_0(P_X) = \min\{p, 1-p\}$. Thus, the channel is symmetrizable if $\Lambda \ge \min\{\Gamma, 1/2\}$. Otherwise, the capacity and the random code capacity are both $H(\Gamma(1-\Lambda)+(1-\Gamma)\Lambda) - H(\Lambda)$, where $H(\cdot)$ is the binary entropy function. Moreover, the optimal input and state distributions in both (40) and (42) are unique, so this channel satisfies the conditions of Corollary 5. The dispersion is given by

$$V_+ = \begin{cases} 4\Gamma(1-\Gamma)\Lambda(1-\Lambda)\log^2 \frac{\Lambda+\Gamma-2\Lambda\Gamma}{1-\Lambda-\Gamma+2\Lambda\Gamma}, & \Gamma \le 1/2 \\ 0 & \Gamma > 1/2. \end{cases}$$

Of particular note is that, even though the capacity is the same as a non-adversarial binary symmetric channel with crossover probability Λ, the dispersion is strictly smaller.

Example 2 (Binary adding AVC): Let $\mathcal{X}, \mathcal{S} = \{0, 1\}$, $\mathcal{Y} = \{0, 1, 2\}$, and $W(y|x,s) = 1$ if $y = x + s$, where we are using real-valued addition. Again let $g(x) = x$ and $\ell(s) = s$. If $P_X = [1-p, p]$, then $\Lambda_0(P_X) = p$. Thus, the channel is symmetrizable if $\Gamma \le \Lambda$. If $\Gamma > \Lambda$ and $\Lambda \le 1/2$, then the capacity and the random code capacity are equal (although with no simple closed form), and moreover the optimal input and state distributions are unique, so the sufficient conditions of Corollary 5 are satisfied. However, if $\Gamma > \Lambda > 1/2$, then the capacity and random code capacity differ, in which case our results do not give tight bounds on the dispersion.

The following lemma, which is a slight restatement of [8, Thm. 3], will be used in the proofs of Thms. 3 and 4. It a Berry-Esseen-type result for interacting constant-composition distributions, which was itself derived from a result on Latin hypercube sampling in [9]. This lemma is key to deriving the dispersion of the AVC under input and state constraints, just as it was in [8] to derive the dispersion of constant-composition codebooks for the multiple-access channel.

Lemma 6: Given $P_X \in \mathcal{P}_n(\mathcal{X})$ and $P_S \in \mathcal{P}_n(\mathcal{S})$, let

$$(X^n, S^n, Y^n) \sim U_{P_X}(x^n)\, U_{P_S}(s^n)\, W^n(y^n|x^n, s^n). \quad (47)$$

Let $Z_n = \sum_{i=1}^n \imath(X_i; Y_i)$ and let $\Sigma_n = \frac{1}{n}\operatorname{Var}(Z_n)$. If $V(P_X, P_S, W) > 0$, then for all γ,

$$\left| \mathbb{P}\left(\frac{Z_n - \mathbb{E}Z_n}{\sqrt{n\Sigma_n}} > \gamma \right) - \mathsf{Q}(\gamma) \right| \le \frac{K\, T(P_X, P_S, W)}{\Sigma_n^{3/2}\sqrt{n}} \quad (48)$$

where K is an absolute constant. Moreover,

$$0 \le \Sigma_n - V(P_X, P_S, W) \le \frac{3}{n-1} \operatorname{Var}(\imath(X;Y)). \quad (49)$$

Proof of Thm. 3: Let $P_S^\star \in \Pi_S^{(r)}(\Lambda)$ achieve the maximum in (43). Let $P_S \in \mathcal{P}_n(\mathcal{S}) \cap \mathcal{P}(\mathcal{S}, \Lambda)$ be such that $\|P_S - P_S^\star\|_\infty \leq 1/n$. The adversary may randomly choose the state sequence from U_{P_S}, inducing the non-adversarial channel $U_{P_S}W^n$. Thus, an upper bound on the achievable rate for this non-adversarial channel is also an upper bound on the underlying AVC. From here on, we only consider this non-adversarial channel. We first bound the number of messages in constant composition codes. Specifically, for any $P_X \in \mathcal{P}_n(\mathcal{X}, \Gamma)$, consider an (M, n, ϵ) code with codewords entirely in T_{P_X}. Applying the finite blocklength non-adversarial converse bound [10, Proposition 4.4], for any $\delta > 0$,

$$\begin{aligned}
&\epsilon + \delta \\
&\geq \sup_{Q_{Y^n}} \max_{x^n \in T_{P_X}} \mathbb{P}\left\{\log \frac{(U_{P_S}W^n)(Y^n|x^n)}{Q_{Y^n}(Y^n)} \leq \log(M\delta)\right\} && (50)\\
&\geq \max_{x^n \in T_{P_X}} \mathbb{P}\left\{\log \frac{(U_{P_S}W^n)(Y^n|x^n)}{(P_X P_S W)^n(Y^n)} \leq \log(M\delta)\right\} && (51)\\
&= \mathbb{P}\left\{\log \frac{(U_{P_S}W^n)(Y^n|X^n)}{(P_X P_S W)^n(Y^n)} \leq \log(M\delta)\right\} && (52)\\
&\geq \mathbb{P}\left\{\log \frac{(P_S W)^n(Y^n|X^n)}{(P_X P_S W)^n(Y^n)} \leq \log(M\delta) - \log|\mathcal{P}_n(\mathcal{S})|\right\} && (53)
\end{aligned}$$

where in (50)–(51), $Y^n \sim (U_{P_S}W^n)(y^n|x^n)$, whereas in (52)–(53), $(X^n, Y^n) \sim U_{P_X}(x^n)(U_{P_S}W^n)(y^n|x^n)$; in (51) we have chosen $Q_{Y^n} = (P_X P_S W)^n$; (52) holds since the quantity in (51) depends only on the type of x^n; and (53) holds because $U_{P_S}(s^n) \leq P_S^n(s^n)|\mathcal{P}_n(\mathcal{S})|$ for all s^n. From (53), we may apply an argument similar to that of [3, Thm. 49], with Lemma 6 in place of the Berry-Esseen theorem, to conclude the proof. (See [6] for more details.) ∎

Proof of Thm. 4: Let $P_X^\star \in \Pi_X(\Gamma)$ achieve the minimum in (41), with $\Lambda_0(P_X^\star) > \Lambda$, the existence of which is assumed in the statement of the theorem. Let $P_X \in \mathcal{P}_n(\mathcal{X}) \cap \mathcal{P}(\mathcal{X}, \Gamma)$ be such that $\|P_S - P_S^\star\|_\infty \leq 1/n$. By continuity, for sufficiently large n we have $\Lambda_0(P_X) > \Lambda$. Let

$$\mathcal{A} = \left\{(x^n, y^n) : \log \frac{(P_S W)^n(y^n|x^n)}{(U_{P_X} P_S^n W^n)(y^n)} \geq \gamma \right. \\ \left. \text{for some } P_S \in \mathcal{P}_n(\mathcal{S})\right\} \quad (54)$$

where we define with hindsight $\gamma = \log\left[\sqrt{n}\,|\mathcal{P}_n(\mathcal{S})\,M\right]$. Fix $\eta > 0$, let $\mathcal{D}_\eta$ be the set of joint distributions $Q_{XX'SY}$ such that $Q_S \in \mathcal{P}(\mathcal{S}, \Lambda)$ and $D(Q_{XX'SY}\|P_X \times Q_{X'S} \times W) \leq \eta$ where

$$(P_X \times Q_{X'S} \times W)(x, x', s, y) = P_X(x)Q_{X'S}(x', s)W(y|x, s).$$

Define a test where $Z(x^n, \bar{x}^n, y^n) = 1$ iff $(x^n, y^n) \in \mathcal{A}$, and either $(\bar{x}^n, y^n) \notin \mathcal{A}$ or there exists s^n such that $Q_{x^n, \bar{x}^n, s^n, y^n} \in \mathcal{D}_\eta$. Since $\Lambda_0(P_X) > \Lambda$, by the continuity of relative entropy on discrete alphabets, for sufficiently small η this test satisfies the uniqueness condition in (6). Thus, we may apply Thm. 1 with $X^n \sim U_{P_X}$ to find that there exists an (M, n, ϵ) code where

$$\begin{aligned}
\epsilon \leq{}& \max_{s^n \in \mathcal{S}^n(\Lambda)} \mathbb{P}((X^n, Y_{s^n}^n) \notin \mathcal{A}) \\
&+ (2\log e)M\,\mathbb{P}(Z(X^n, \bar{X}^n, Y_{s^n}^n) = 0,\, (X^n, Y_{s^n}^n) \in \mathcal{A}) \\
&+ \max_{\bar{x}^n} 2\log(3n|\mathcal{S}|)\,\mathbb{P}(Z(X^n, \bar{x}^n, Y_{s^n}^n) = 0, (X^n, Y_{s^n}^n) \in \mathcal{A}) \\
&+ \sqrt{\frac{2\ln(3n|\mathcal{S}|)}{M}} && (55)
\end{aligned}$$

where $Y_{s^n}^n$ indicates the channel output sequence with state sequence s^n. We may bound the first term in (55) by

$$\begin{aligned}
&\mathbb{P}((X^n, Y_{s^n}^n) \notin \mathcal{A}) \\
&= \mathbb{P}\left(\log \frac{(P_S W)^n(Y_{s^n}^n|X^n)}{(U_{P_X} P_S^n W^n)(Y_{s^n}^n)} < \gamma \text{ for all } P_S \in \mathcal{P}_n(\mathcal{S})\right) && (56)\\
&\leq \mathbb{P}\left(\log \frac{(Q_{s^n} W)^n(Y_{s^n}^n|X^n)}{(P_X Q_{s^n} W)^n(Y_{s^n}^n)} < \gamma + \log|\mathcal{P}_n(\mathcal{X})|\right) && (57)\\
&\leq \mathsf{Q}\left(\frac{nI(P_X; Q_{s^n}W) - \gamma - \log|\mathcal{P}_n(\mathcal{X})|}{\sqrt{n\Sigma_n}}\right) + O\left(\frac{1}{\sqrt{n}}\right) && (58)
\end{aligned}$$

where (57) follows because $U_{P_X}(x^n) \leq |\mathcal{P}_n(\mathcal{X})|P_X^n(x^n)$ for all x^n, and (58) follows from Lemma 6, where Σ_n satisfies (49). The second term in (55) can be shown to be $O(1/\sqrt{n})$ with an application of Markov's inequality. The third term in (55) can be shown to be exponentially vanishing for any $\eta > 0$ by the definition of $\mathcal{D}_\eta$ and basic results from the method of types. The fourth term in (55) is also insignificant if M is exponentially large. These bounds are sufficient to prove the theorem. (See [6] for more details.) ∎

REFERENCES

[1] I. Csiszár and P. Narayan, "The capacity of the arbitrarily varying channel revisited: positivity, constraints," *IEEE Transactions on Information Theory*, vol. 34, no. 2, pp. 181–193, 1988.

[2] V. Strassen, "Asymptotic approximations in Shannon's information theory," http://www.math.cornell.edu/ pmlut/strassen.pdf, Aug. 2009, english translation of original Russian article in Trans. Third Prague Conf. on Inform. Th., Statistics, Decision Functions, Random Processes (Liblice, 1962), Prague, 1964.

[3] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inform. Theory*, vol. 56, pp. 2307–2359, 2010.

[4] Y. Polyanskiy, "On dispersion of compound DMCs," in *2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Oct 2013, pp. 26–32.

[5] O. Kosut and J. Kliewer, "Dispersion of the discrete arbitrarily-varying channel with limited shared randomness," in *2017 IEEE International Symposium on Information Theory (ISIT)*, June 2017, pp. 1242–1246.

[6] ——, "Finite blocklength and dispersion bounds for the arbitrarily-varying channel," [Online] arXiv:1801.03594, Jan. 2018.

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

[8] J. Scarlett, A. Martinez, and A. G. i Fàbregas, "Second-order rate region of constant-composition codes for the multiple-access channel," *IEEE Transactions on Information Theory*, vol. 61, no. 1, pp. 157–172, Jan 2015.

[9] W. Loh, "On Latin hypercube sampling," *Annals of Stats.*, vol. 24, no. 5, pp. 2058–2080, 1996.

[10] V. Y. F. Tan, "Asymptotic estimates in information theory with non-vanishing error probabilities," *Found. Trends Commun. Inf. Theory*, vol. 11, no. 1-2, pp. 1–184, Sep. 2014.