Rateless Codes for AVC Models

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

***Abstract*—The arbitrarily varying channel (AVC) is a channel model whose state is selected maliciously by an adversary. Fixed-blocklength coding assumes a worst-case bound on the adversary's capabilities, which leads to pessimistic results. This paper defines a variable-length perspective on this problem, for which achievable rates are shown that depend on the realized actions of the adversary. Specifically, rateless codes are constructed which require a limited amount of common randomness. These codes are constructed for two kinds of AVC models. In the first the channel state cannot depend on the channel input, and in the second it can. As a by-product, the randomized coding capacity of the AVC with state depending on the transmitted codeword is found and shown to be achievable with a small amount of common randomness. The results for this model are proved using a randomized strategy based on list decoding.**

***Index Terms*—Adversarial models, arbitrarily varying channels (AVCs), randomization, rateless coding.**

I. Introduction

MODERN communication platforms such as sensor networks, wireless *ad hoc* networks, and cognitive radio involve communication in environments that are difficult to model. This difficulty may stem from the cost of measuring channel characteristics, the behavior of other users, or the interaction of heterogeneous systems using the same resources. These systems may use extra resources such as feedback on a low-rate control channel or common randomness to overcome this channel uncertainty.

Inspired by some of these challenges, we consider variable-length coding over arbitrarily varying channels (AVCs). The AVC is an adversarial channel model in which the channel is governed by a time varying state controlled by a *jammer* who wishes to maximize the decoding error probability. For fixed-blocklength coding, the capacity is the worst-case over all allowable actions of the jammer. However, in some cases the worst-case may be unduly pessimistic.

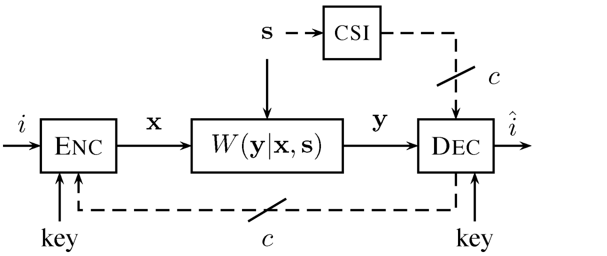

Fig. 1. Rateless communication system. The encoder and decoder share a source of common randomness. A single bit of feedback is available every c channel uses for the decoder to terminate transmission. Some partial information about the channel state is available at the decoder every c channel uses in a causal fashion.

In this paper, we study *randomized coding* for two different variable-length coding models based on the AVC. In a randomized code the encoder and decoder have a shared source of common randomness unknown to the jammer which acts as a shared *key* to mask the coding strategy from the jammer. The first model we study is the AVC under maximal error and randomized coding, in which the state sequence is chosen independently of the transmitted codeword. The second model is an AVC in which the jammer can choose the state sequence based on the transmitted codeword. This may be an appropriate model for a multi-hop network in which an internal node becomes compromised and tampers with transmitted packets. We call this situation an AVC with "nosy noise." Our first result is a formula for the randomized coding capacity of this AVC. Our proof uses results on list decoding for AVCs [3]–[5] with a partial derandomization technique used by Langberg [6].

The main focus of this paper is on the problem of *rateless coding* for these channels using *limited common randomness* and *partial channel state information*, as shown in Fig. 1. Rateless codes were first proposed for erasure channels [7], [8] and compound channels [9], [10], and a general model is discussed in [11]. They are strategies that allow a single-bit feedback signal (often called an ACK/NACK for "acknowledge"/"not acknowledge") every c channel uses to terminate transmission based on the observed channel output $\mathbf{y}$ and channel state information. In our model, the partial state information takes the form of estimates of the average channel induced by the channel state $\mathbf{s}$ over "chunks" of size c. In practice this channel information may come from exogenous measurements or from training information in the forward link, as in [12].

The arbitrarily varying channel was first studied in the seminal paper of Blackwell, Breiman, and Thomasian [13], who found a formula for the capacity under randomized coding and maximal error. Without randomized coding, the maximal-error problem is significantly harder [14]–[17] and is related to the zero-error capacity [18]. The AVC model was extended to include constraints on the jammer by Hughes and Narayan [19]

Manuscript received November 22, 2007; revised November 22, 2007. Current version published June 16, 2010. The work of A. D. Sarwate and M. Gastpar was supported 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). The material in this paper was presented at the IEEE Symposium on Information Theory, Nice, France, 2007 and at the IEEE Information Theory Workshop, Tahoe, CA, 2007.

A. D. Sarwate was with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720 USA. He is now with the Information Theory and Applications Center, University of California, San Diego, La Jolla, CA 92093 USA (e-mail: asarwate@alum.mit.edu).

M. Gastpar is with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, Berkeley, CA 94720 USA, and also with the Faculty of Electrical Engineering, Mathematics and Computer Science (EEMCS), Delft University of Technology, Delft, The Netherlands (e-mail: gastpar@eecs.berkeley.edu).

Communicated by I. Kontoyiannis, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2010.2048497

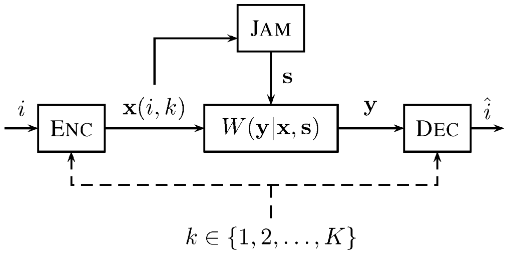
JAM
$\mathbf{s}$
i
ENC
$\mathbf{x}(i,k)$
$W(\mathbf{y}|\mathbf{x},\mathbf{s})$
$\mathbf{y}$
DEC
$\hat{i}$
$k \in \{1, 2, \ldots, K\}$

Fig. 2. The nosy noise error model—the jammer knows the codeword $\phi_k(i)$.

and Csiszár and Narayan [20]–[22]. For randomized coding, error exponents have also been studied [23]–[25]. Ahlswede's landmark paper [26] showed that the average error capacity under deterministic coding $\bar{C}_d$ is 0 or equal to the randomized coding capacity C_r. This result does not hold when there is a constraint on the channel state, but the method can be used to show that only $O(\log n)$ bits of common randomness is needed to achieve $C_r(\Lambda)$ for AVCs with cost constraint Λ.

The "nosy noise" model, shown in Fig. 2, has been discussed previously in the AVC literature [15], [27], where it is sometimes called the A*VC [28, Problem 2.6.21]. To our knowledge, for AVCs with a cost-constrained jammer the problem was not studied until Langberg [6] found the capacity for bit-flipping channels with randomized coding (see also [29]). Agarwal, Sahai and Mitter proposed a similar model with a distortion constraint [30], which is different than the AVC model considered here [5, p. 216].

II. Channel Models and Definitions

The time-varying channel is modeled by a set of channels $\mathcal{W} = \{W(y \,|\, x, s) : s \in \mathcal{S}\}$ with finite input alphabet $\mathcal{X}$ and finite output alphabet $\mathcal{Y}$. This is an arbitrarily varying channel (AVC) model. 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 observing the output $\mathbf{y}$ given the input $\mathbf{x}$ and state $\mathbf{s}$ over the AVC $\mathcal{W}$ without feedback is given by

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

In this paper, feedback is used only to terminate transmission, and we compare our achievable rates with those achievable without feedback (cf. [11]). The interpretation of (1) is that the channel state can change arbitrarily from time to time.

We will impose a cost constraint on the state sequences [20]. Let $l : \mathcal{S} \to \mathbb{R}^+$ be a cost function on the state set, where $\min_s l(s) = 0$ and $\max_{s \in \mathcal{S}} l(s) = \lambda^* < \infty$. The cost of the vector $\mathbf{s} = (s_1, s_2, \ldots, s_n)$ is the sum of the cost on the elements

$$l(\mathbf{s}) = \sum_{i=1}^{n} l(s_i). \tag{2}$$

In some cases, we will impose a total constraint Λ on the average cost, so that $l(\mathbf{s}) \leq n\Lambda$. If $\Lambda \geq \lambda^*$ we say the state is unconstrained. We will define the set $\mathcal{S}^n(\Lambda) = \{\mathbf{s} : l(\mathbf{s}) \leq n\Lambda\}$ to be the set of sequences with average cost less than or equal to Λ.

Point-to-Point Channel Coding: An (n, N) *deterministic code* $\mathcal{C}$ for the AVC $\mathcal{W}$ is a pair of maps (ϕ, ψ) with $\phi : [N] \to \mathcal{X}^n$ and $\psi : \mathcal{Y}^n \to [N]$. The *rate* of the code is $n^{-1} \log N$. The *decoding region* for message i is $D_i = \{\mathbf{y} : \psi(\mathbf{y}) = i\}$. A (n, N) *randomized code* $\mathbf{C}$ for the AVC $\mathcal{W}$ is a random variable taking on values in the set of deterministic codes. If $\mathbf{C} = (\Phi, \Psi)$ is uniformly distributed on a set of K codes, then we call this an (n, N, K) randomized code with *key size* $\log K$. Note that the realization of the code is shared by the encoder and decoder, so the key is known by both parties. The *rate* of the code is $R = n^{-1} \log N$. The *decoding region* is a random variable $\mathbf{D}_i = \{\mathbf{y} : \Psi(\mathbf{y}) = i\}$ and under key k we write $D_{i,k} = \{\mathbf{y} : \psi_k(\mathbf{y}) = i\}$. For a randomized code we require that the decoding error be small for each message *averaged over key values*. Randomization allows several different codewords to represent the same message. For maximal error, there are two cases to consider, depending on whether or not the state can depend on the actual *codeword*.

The *standard maximal error* for an (n, N) randomized code over an AVC $\mathcal{W}$ with cost constraint Λ is given by

$$\varepsilon = \max_{i} \max_{\mathbf{s} \in \mathcal{S}^n(\Lambda)} \mathbb{E}[1 - W(\mathbf{D}_i \,|\, \Phi(i), \mathbf{s})] \tag{3}$$

where the expectation is over the randomized code (Φ, Ψ). The *nosy maximal error* for an (n, N) randomized code over an AVC $\mathcal{W}$ with cost constraint Λ is given by

$$\hat{\varepsilon} = \max_{i} \max_{J : \mathcal{X}^n \to \mathcal{S}^n(\Lambda)} \mathbb{E}[1 - W(\mathbf{D}_i \,|\, \Phi(i), J(\Phi(i)))] \tag{4}$$

where the expectation is over the randomized code (Φ, Ψ). In these definitions, $\mathbf{D}_i$ and $\Phi(i)$, and $J(\Phi(i))$ correspond to the same key. We call an AVC under the nosy maximal error criterion an *AVC with nosy noise*. Fig. 2 shows the channel model under the nosy noise assumption. In the AVC with nosy noise, the jammer's strategies take the form of mappings $J : \mathcal{X}^n \to \mathcal{S}^n(\Lambda)$ from the codeword vectors to state sequences. Under randomized coding we will show that from a capacity standpoint all that matters is whether the jammer has access to the *current* input symbol.

A rate R is called achievable if for every $\epsilon > 0$ there exists a sequence of (n, N) codes of rate $R_n \geq R - \delta$ whose probability of error (maximal or nosy) is at most ϵ. For a given error criterion, the supremum of achievable rates is the capacity of the arbitrarily varying channel. We will write $C_r(\Lambda)$ for the randomized coding capacity under maximal error with constraint Λ, and $\hat{C}_r(\Lambda)$ for the randomized coding capacity with nosy noise and state constraints.

Information Quantities: For a fixed input distribution $P(x)$ and channel $V(y \,|\, x)$, we write $I(P, V)$ for the mutual information between input and output. For a finite or closed and convex set of channels $\mathcal{V}$ we use the shorthand

$$I(P, \mathcal{V}) = \min_{V \in \mathcal{V}} I(P, V). \tag{5}$$

We define the following sets:

$$\mathcal{Q}(\Lambda) = \left\{ Q \in \mathcal{P}(\mathcal{S}) : \sum_{s} Q(s) l(s) \leq \Lambda \right\} \tag{6}$$

$$\mathcal{U}(P, \Lambda) = \left\{ U \in \mathcal{P}(\mathcal{S} \,|\, \mathcal{X}) : \sum_{s,x} U(s \,|\, x) P(x) l(s) \leq \Lambda \right\}. \tag{7}$$

For an AVC $\mathcal{W} = \{W(y \,|\, x, s) : s \in \mathcal{S}\}$ with state constraint Λ we define two sets of channels:

$$\mathcal{W}_{\text{std}}(\Lambda) = \left\{ \sum_{s} W(y \,|\, x, s) Q(s) : Q \in \mathcal{Q}(\Lambda) \right\} \tag{8}$$

$$\mathcal{W}_{\text{dep}}(P, \Lambda) = \left\{ \sum_{s} W(y \,|\, x, s) U(s \,|\, x) : U \in \mathcal{U}(P, \Lambda) \right\}. \tag{9}$$

We will suppress the explicit dependence on Λ. The set in (8) is called the *convex closure* of $\mathcal{W}$, and the set in (9) is the *row-convex closure* of $\mathcal{W}$. In earlier works $\mathcal{W}_{\text{dep}}(P, \Lambda)$ is sometimes written as $\bar{\bar{\mathcal{W}}}$ [26].

Two information quantities of interest in randomized coding for AVCs are

$$C_{\text{std}}(\Lambda) = \max_{P} \min_{V \in \mathcal{W}_{\text{std}}(\Lambda)} I(P, V) \tag{10}$$

$$C_{\text{dep}}(\Lambda) = \max_{P} \min_{V \in \mathcal{W}_{\text{dep}}(P, \Lambda)} I(P, V). \tag{11}$$

Csiszár and Narayan [20] showed that the randomized coding capacity under maximal error $C_r(\Lambda)$ is equal to $C_{\text{std}}(\Lambda)$. In Theorem 1 we show that the randomized coding capacity under nosy noise $\hat{C}_r(\Lambda)$ is equal to $C_{\text{dep}}(\Lambda)$.

Partial Channel State Information: In the rateless codes we consider, the decoder selects an appropriate decoding time based on partial channel state observation. To simplify the analysis, we suppose that such information is available for each so-called *chunk* of $c(n)$ channel uses. Specifically, suppose that during the mth chunk of channel uses $\{(m-1)c+1, \ldots mc\}$ the channel inputs were $\mathbf{x}^{(mc)}$ and the state was $\mathbf{s}^{(mc)}$. Under the maximal error criterion, we define the average channel under $\mathbf{s}$ during the mth chunk by

$$V_m(y \,|\, x) = \frac{1}{c} \sum_{t=(m-1)c+1}^{mc} W(y \,|\, x, s_t).$$

Under the nosy noise criterion we define the average channel under $\mathbf{x}$ and $\mathbf{s}$ by

$$V_m(y \,|\, x) = \frac{1}{N(x \,|\, \mathbf{x}^{(mc)})} \times \sum_{t=(m-1)c+1}^{mc} W(y \,|\, x_t, s_t) \mathbf{1}(x_t = x). \tag{12}$$

A receiver with full side information would learn the channel V_m explicitly. We consider instead the case where the receiver is given a set $\mathcal{V}_m$ after the m-th chunk, where $\mathcal{V}_m$ is a subset of channels such that $V_m(y \,|\, x) \in \mathcal{V}_m$.

We denote the set of possible values for $\mathcal{V}_m$ by $\mathbb{V}(c)$. This is a collection of subsets of $\mathcal{W}_{\text{std}}(\Lambda) \cap \mathcal{P}_c(\mathcal{Y} \,|\, \mathcal{X})$ for maximal error and of $\mathcal{W}_{\text{dep}}(\Lambda) \cap \mathcal{P}_c(\mathcal{Y} \,|\, \mathcal{X})$ for nosy noise. We will assume a polynomial upper bound on the size of $\mathbb{V}(c)$

$$|\mathbb{V}(c)| \leq c^{v} \tag{13}$$

for some $v < \infty$.

We consider two models for $\mathcal{V}_m$: in the first the decoder gets an estimate of the empirical cost of the true state sequence, and in the second the decoder gets an estimate of the mutual information induced by the true channel. For rateless codes under maximal error we will assume that the receiver gets an estimate $\hat{\lambda}_m$ such that the true cost

$$\lambda_m = \frac{1}{c} \sum_{t=(m-1)c+1}^{mc} l(s_t) \tag{14}$$

satisfies $\lambda_m \leq \hat{\lambda}_m \leq \lambda_m + \epsilon$. The CSI set is then

$$\mathcal{V}_m = \left\{ V(y \,|\, x) = \frac{1}{c} \sum_{i=1}^{c} W(y \,|\, x, \hat{s}_i) \right. \\ \left. : l(\hat{\mathbf{s}}) \leq l\left(\mathbf{s}^{(mc)}\right) + c\epsilon \right\}. \tag{15}$$

We call such CSI ϵ*-cost-consistent.*

For rateless codes under nosy maximal error, we will say a CSI sequence is ϵ*-consistent* for input P if

$$I(P, V_m) - \min_{V \in \mathcal{V}_m} I(P, V) \leq \epsilon. \tag{16}$$

Our rateless codes for nosy maximal error will assume the CSI sequence is ϵ-consistent.

Rateless Codes: Based on the partial channel state information, the decoder selects an appropriate decoding time such as to ensure reliable decoding. To simplify the analysis, we will assume that the decoding time is an integer multiple of the chunk size $c(n)$, denote by $\mathbf{M}$. Thus, the empirical rate is simply given by

$$R_{\text{emp}} = \frac{1}{\mathbf{M}c} \log_2 N \tag{17}$$

where N is the number of codewords in the codebook.

Formally, a (c, N, K) randomized rateless code is set of maps $\{(\Phi_m, \tau_m, \Psi_m) : m = 1, 2, \ldots\}$

$$\Phi_m : [N] \times [K] \to \mathcal{X}^c \tag{18}$$

$$\tau_m : \mathcal{Y}^{mc} \times \mathbb{V}(c)^m \times [K] \to \{0, 1\} \tag{19}$$

$$\Psi_m : \mathcal{Y}^{mc} \times \mathbb{V}(c)^m \times [K] \to [N]. \tag{20}$$

To encode chunk m, the encoding function Φ_m uses the message in $[N]$ and key in $[K]$ to choose a vector of c channel inputs.

The decision function τ_m defines a random variable, called the *decoding time* $\mathbf{M}$ of the rateless code

$$\mathbf{M} = \min \left\{ m : \tau_m \left(\mathbf{y}_1^{mc}, \mathcal{V}_1^m, k \right) = 1 \right\}. \tag{21}$$

Let $\mathcal{M} = \{M_*, M_* + 1, \ldots, M^*\}$ be the smallest interval containing the support of $\mathbf{M}$. The set of possible (empirical) rates for the rateless code are given by $\{(mc)^{-1} \log N : m \in \mathcal{M}\}$.

We can define decoding regions for the rateless code at a decoding time $\mathbf{M} = M$. Note that if $\mathbf{M} = M$ we have $\tau_M(\mathbf{y}_1^{Mc}, \mathcal{V}_1^M, k) = 1$. For message i, key k and side information vector $\mathcal{V}_1^M$ we can define a decoding region

$$D_{i,k}\left(\mathcal{V}_1^M\right) = \left\{\mathbf{y}_1^{Mc} : \tau_M\left(\mathbf{y}_1^{Mc}, \mathcal{V}_1^M, k\right) = 1, \right.\\ \left. \Psi_M\left(\mathbf{y}_1^{Mc}, \mathcal{V}_1^M, k\right) = i\right\}. \quad (22)$$

For a given state $\mathbf{s}_1^{Mc}$ (for maximal error) or jammer strategy $J_M(i, \Phi_1^M(i,k))$ (for nosy noise) the probability of the decoding regions are

$$\Delta_{i,k} = W^{Mc}\left(D_{i,k}\left(\mathcal{V}_1^M\right)\middle|\, \Phi_1^M(i,k), \mathbf{s}_1^{Mc}\right)$$
$$\hat{\Delta}_{i,k} = W^{Mc}\left(D_{i,k}\left(\mathcal{V}_1^M\right)\middle|\, \Phi_1^M(i,k), J_M\left(i, \Phi_1^M(i,k)\right)\right).$$

The *maximal* and *nosy noise error* for a (c, N, K) rateless code at decoding time $\mathbf{M} = M$ are, respectively

$$\varepsilon\left(M, \mathbf{s}, \mathcal{V}_1^M\right) = \max_{i \in [N]} \frac{1}{K} \sum_{k=1}^{K} \left(1 - \Delta_{i,k}\left(M, \mathbf{s}_1^{Mc}, \mathcal{V}_1^M\right)\right) \quad (23)$$

$$\hat{\varepsilon}\left(M, J, \mathcal{V}_1^M\right) = \max_{i \in [N]} \frac{1}{K} \sum_{k=1}^{K} \left(1 - \hat{\Delta}_{i,k}\left(M, J_M, \mathcal{V}_1^M\right)\right). \quad (24)$$

Here $J = (J_1, \ldots, J_M)$ and $J_M : [N] \times \mathcal{X}^{Mc} \to \mathcal{S}^{Mc}$ is the adversary's strategy. Note that in these error definitions we do not take the maximum over all $\mathbf{s}$ or $\mathbf{J}$, because the rate and error at which we decode will depend on the realized state sequence, in contrast to the point-to-point AVC errors in (3) and (4).

III. Main Results and Contributions

Our first main result is Theorem 1, which is a characterization of the capacity of the AVC with nosy noise. The proof is given in Section IV.

Theorem 1: Let $\mathcal{W}$ be an AVC with state cost function $l(\cdot)$ and cost constraint Λ. Then $C_{\text{dep}}(\Lambda)$ is the randomized coding capacity of the AVC with nosy noise

$$\hat{C}_r(\Lambda) = C_{\text{dep}}(\Lambda). \quad (25)$$

Furthermore, for any $\epsilon > 0$, there exists an n sufficiently large such that the sequence of rate-key size pairs $(R, K(n))$ is achievable with nosy maximal error $\hat{\varepsilon}_r(n)$, where $n^2 \leq K(n) \leq \exp(n\epsilon)$ and

$$R = C_{\text{dep}}(\Lambda) - \epsilon \quad (26)$$

$$\hat{\varepsilon}(n) \leq \exp(-n\hat{E}(\epsilon)) + \frac{12nC_{\text{dep}}(\Lambda)\log|\mathcal{Y}|}{\epsilon\sqrt{K(n)}\log K(n)} \quad (27)$$

where $\hat{E}(a) > 0$ for $a > 0$.

This theorem is proved by first constructing list-decodable codes with constant list size for cost-constrained AVCs. These list-decodable codes can be combined with a message-authentication scheme due to Langberg [6] in Lemma 2, which shows that the a secret key can be used to disambiguate the list. Because $\mathcal{W}_{\text{std}}(\Lambda) \subseteq \mathcal{W}_{\text{dep}}(\Lambda)$, in general we have $C_{\text{dep}}(\Lambda) \leq C_{\text{std}}(\Lambda)$. In some cases equality can hold, as in the following example.

Example 1 (Bit-Flipping): Consider an AVC with input alphabet $\mathcal{X} = \{0,1\}$, state alphabet $\mathcal{S} = \{0,1\}$ and output alphabet $\mathcal{Y} = \{0,1\}$, with $y = x \oplus s$, where $\oplus$ denotes addition modulo two and $l(s) = s$. It has been shown [6], [20] $C_{\text{std}}(\Lambda) = C_{\text{dep}}(\Lambda) = 1 - h_b(\Lambda)$, where $h_b(t)$ is the binary entropy function. Furthermore, the capacities $C_r(\Lambda)$ and $\hat{C}_r(\Lambda)$ are both equal to $1 - h_b(\Lambda)$.

In general, the capacities under maximal error and nosy noise are different.

Example 2 (Real Adder): Consider an AVC with input alphabet $\mathcal{X} = \{0,1\}$, state alphabet $\mathcal{S} = \{0,1\}$ and output alphabet $\mathcal{Y} = \{0,1,2\}$, with $y = x + s$, and the addition is taken over the real numbers. When $l(s) = s$, if $\Lambda \geq 1/2$ Csiszár and Narayan [20] showed that $C_r(\Lambda) = 1/2$ and is achieved with $P = (1/2, 1/2)$. However, in the case of nosy noise the capacity is lower when $\Lambda > 1/2$ because the jammer can see the codeword, it can selectively set the output to be 1 if $P = (1/2, 1/2)$. A straightforward calculation [5] shows that $\hat{C}_r(\Lambda) = C_{\text{dep}}(\Lambda) < 1/2$.

Theorems 2 and 3 provide achievable strategies for rateless coding over channels with input-independent and input-dependent state, respectively. Their proofs are given in Sections V-A–1 and V-C. To state our results in a way that makes the tradeoff between error probability and blocklength clearer, we will assume

$$c(n) = n^{1/4} \quad (28)$$
$$M^*(n) = n/c(n) = n^{3/4}. \quad (29)$$

For maximum and minimum rates $R_{\max}$ and $R_{\min}$ the number of messages is $N(n) = \exp(nR_{\min})$ and $M_* = \frac{R_{\min}}{R_{\max}} n^{3/4}$.

Theorem 2: Let $\mathcal{W}$ be an AVC with state cost function $l(\cdot)$. Fix $\epsilon > 0$, $R_{\min} > 0$, and input type $P \in \mathcal{P}(\mathcal{X})$ with $\min_x P(x) > 0$. Then there is an n_0 sufficiently large such that for all $n > n_0$ there exists a $(c(n), \exp(nR_{\min}), K(n))$ randomized rateless code with $K(n)/n \to \infty$ whose decoding time satisfies

$$\mathbf{M} = \min_{M_* \leq M \leq M^*} \left\{ \frac{nR_{\min}}{Mc} < I\left(P, \mathcal{W}_{\text{std}}\left(\frac{1}{Mc} l\left(\mathbf{s}_1^{Mc}\right)\right)\right) - g(\epsilon) \right\} \quad (30)$$

where $g(\epsilon) \to 0$ as $\epsilon \to 0$. The maximal error of the code at this decoding time satisfies

$$\varepsilon\left(\mathbf{s}, \mathcal{V}_1^{\mathbf{M}}\right) = O\left(\frac{n}{K(n)}\right)$$

for state sequences $\mathbf{s}$ and ϵ-cost-consistent CSI $\mathcal{V}_1^{\mathbf{M}}$.

Theorem 3: Let $\mathcal{W}$ be an AVC with state cost function $l(\cdot)$. Fix $R_{\min} > 0$, $\epsilon > 0$, input type $P \in \mathcal{P}(\mathcal{X})$ with $\min_x P(x) > 0$. Then there is an n_0 sufficiently large such that for all $n > n_0$,

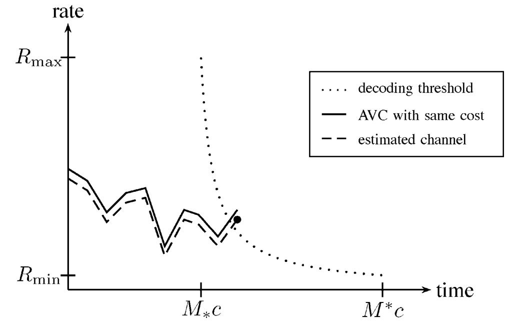

Fig. 3. Decoding rate versus time in a rateless code. The empirical mutual information corresponding to the AVC with the true cost (solid line) varies, and the δ-consistent channel estimates (dashed line) can track it. Once the channel estimates cross the decoding threshold (dotted line), the receiver terminates transmission and tries to decode.

there exists a $(c(n), \exp(nR_{\min}), K(n))$ rateless code whose decoding time satisfies

$$\mathbf{M} = \min_{M_* \le M \le M^*} \left\{ M : \frac{nR_{\min}}{Mc} < \frac{1}{M} \sum_{m=1}^{M} I(P, V_m) - 2\epsilon \right\}$$

where V_m is the average channel in (12). The nosy maximal error at this decoding time satisfies

$$\hat{\varepsilon}(J, \mathcal{V}_1^{\mathbf{M}}) \le O\left(\frac{n}{\epsilon \sqrt{K} \log K} \right)$$

for $K(n) = O(\exp(c))$, state sequences $\mathbf{s}$ and ϵ-consistent side information given by (16).

Theorems 2 and 3 say that if the CSI estimates "good," then the decoder will decode when the empirical mutual information of the channel exceeds the empirical rate $\frac{nR_{\min}}{Mc}$. The two models differ in how they measure the empirical mutual information. The channel tracking is illustrated in Fig. 3. The solid line represents the mutual information of the AVC corresponding to the true channel as measured under the maximal error or nosy noise criterion. The dashed line represents the mutual information corresponding to the estimated channel. The dotted line is the empirical rate, so once the estimate crosses the threshold then the decoder will decode. For Theorem 2 the error decays as $n/K(n)$, whereas it decays like $n/\sqrt{K} \log K$ in Theorem 3.

Both codebooks begin with a constant composition code that is good and manipulate it into a rateless code with the desired properties. The code in Theorem 2 is a fully randomized code whose randomness is reduced by Lemma 1. In Theorem 3, the decoder decodes each chunk of c channel uses into a list of possible messages. As more chunks are received, the list size shrinks and the decoding time $\mathbf{M}$ is chosen to guarantee that the list size is bounded by a constant. Lemma 2 shows that this code can be used as part of a randomized code in which the key disambiguates the list at the decoder.

Example 3 [Bit-Flipping (Mod-Two Adder)]: Consider the mod-two additive AVC described in Example 1 on page 4 where the partial side information $\mathcal{V}_m$ is an estimate $\hat{\lambda}_m$ of the empirical Hamming weight of the state sequence $\mathbf{s}^{(mc)}$. The receiver tracks the empirical weight of the state sequence to compute an estimate $\hat{\Lambda}_M$ of the crossover probability. Theorems 2 and 3 both give rateless codes that can decode as soon as the estimated empirical mutual information $Mc(1-h_b(\hat{\Lambda}_M))$ exceeds the size of the message ($\log N$ bits). As $R_{\min}$ can be as small as we like, these codes can work for empirical state sequences with Hamming weight arbitrarily close to 1/2. The realized rate is within ϵ of $1 - h_b(\hat{\Lambda}_M)$, but the two codes differ greatly in the dependence of the error probability on the amount of common randomness. When the bit-flips cannot depend on the transmitted codeword, the error decays with K^{-1}, and when they can it decays with $(\sqrt{K} \log K)^{-1}$.

Remark: For the bit-flipping example, the rates guaranteed by both theorems are close to the *capacity* of the AVC with the corresponding cost constraint. However, in general this may not be the case. Both coding schemes use a fixed input type P, which is is a common feature of rateless coding strategies [9], [12], [31] but may result in some loss in rate [32] with respect to an input distribution chosen with knowledge of the empirical state distribution. It may be possible to adapt the channel input distribution, perhaps using ideas from universal prediction [33] but we leave that for future work.

This scheme can also be used with more general settings for the parameters of the scheme beyond those in (28) and (29). It is also possible to relax the ϵ-consistency requirement. However, in that setting it is hard to quantify how close the rate at the decoding time is to the empirical mutual information of the channel.

IV. Two Partial Derandomization Techniques

We now describe two lemmas which can be used to reduce the common randomness needed for our code constructions: the "elimination technique" [26], and a message authentication technique [6], [34]. The former is applied to randomized codes [24], [25] to limit the key size, and the second to list-decodable codes; both yield randomized codes with key sizes of $O(\log n)$ bits.

Lemma 1 (Elimination Technique [26]): Let J be a positive integer and let $\mathbf{C}$ be an (n, N, J) randomized code with $N = \exp(nR)$ whose expected maximal error satisfies

$$\max_{\mathbf{s} \in \mathcal{S}^n(\Lambda)} \max_i \mathbb{E}_{\mathbf{C}}[\varepsilon(i, \mathbf{s})] \le \delta(n)$$

for an AVC $\mathcal{W}$ with cost function $l(\cdot)$ and cost constraint Λ. Then for all μ satisfying

$$\mu \log \delta(n)^{-1} - h_b(\mu) \log 2 > \frac{n}{K}(R \log 2 + \log |\mathcal{S}|)$$

where $h_b(\mu)$ is the binary entropy function, with probability exponentially small in n, the (n, N, K) randomized code uniformly distributed on K i.i.d. copies from $\mathbf{C}$ will have with maximal probability of error less than μ.

The proof follows directly from the arguments in [26] and is omitted. In particular, if there is a sequence of randomized codes whose errors decay exponentially, so $\delta(n) \le \exp(-\alpha n)$, then a little algebra shows that we can choose the key size $K(n)$ and the error μ to satisfy $\mu \le n/K(n)$. The code of [24], [25] has exponentially decaying error probability, so Lemma 1 shows that the randomized coding capacity $C_r(\Lambda)$ is achievable with common randomness $K(n)$ polynomial in n, which corresponds to $O(\log n)$ bits.

For AVCs with nosy noise, the state can depend on the transmitted codeword. By combining these list-decodable codes with a message authentication scheme used by Langberg [6], we can construct randomized codes for this channel with limited common randomness. The relationship between the key size, list size, and error is given by the following Lemma.

Lemma 2 (Message Authentication [6]): Let $\mathcal{W}$ be an AVC and suppose we are given an (n, N, L) deterministic list-decodable code and probability of error ϵ. For key size $K(n)$ where $K(n)$ is a power of a prime there exists an $(n, N/\sqrt{K(n)}, K(n))$ randomized code with nosy maximal error $\hat{\varepsilon}(\mathbf{s})$ such that

$$\max_{\mathbf{s}} \hat{\varepsilon}(\mathbf{s}) \leq \epsilon + \frac{2L \log N(n)}{\sqrt{K(n)} \log K(n)}. \tag{31}$$

By choosing the appropriate input distribution we can obtain our first new result: a formula for the randomized coding capacity for the AVC under nosy noise.

Proof of Theorem 1: To show the converse, note that the jammer can choose a memoryless strategy $U(s \,|\, x) \in \mathcal{U}(P, \Lambda)$. Choosing the worst U yields a discrete memoryless channel whose capacity is $C_{\mathrm{dep}}(\Lambda)$, and therefore the randomized coding capacity for this channel is given by $C_{\mathrm{dep}}(\Lambda)$.

To show that rates below $C_{\mathrm{dep}}(\Lambda)$ are achievable, we first fix $K(n)$ and let P be the input distribution maximizing $C_{\mathrm{dep}}(\Lambda)$. In [5] it is shown using methods from [3], [4] that for any $\epsilon_1 > 0$, there is an n sufficiently large and a list-decodable code with codewords of type P, rate $R = C_{\mathrm{dep}}(\Lambda) - \epsilon_1$, list size

$$L < \left\lfloor \frac{6 \log |\mathcal{Y}|}{\epsilon_1} \right\rfloor + 1 \tag{32}$$

and error $\epsilon_1 \leq \exp(-nE(\epsilon_1))$, where $E(\epsilon_1) > 0$. We can use Lemma 2 to construct an $(n, N(n)/\sqrt{K(n)}, K(n))$ randomized code with error probability

$$\hat{\varepsilon} < \exp(-nE(\epsilon_1)) + \frac{12 n C_{\mathrm{dep}}(\Lambda) \log |\mathcal{Y}|}{\epsilon_1 \sqrt{K(n)} \log K(n)}.$$

The rate of this randomized code is

$$R = C_{\mathrm{dep}}(\Lambda) - \epsilon_1 - \frac{1}{n} \log \frac{\sqrt{K(n)}}{L}.$$

For any $\epsilon > 0$ and $K(n) \leq \exp(n\epsilon)$ we can choose ϵ_1 small enough so that $R = C_{\mathrm{dep}}(\Lambda) - \epsilon$. ■

V. Rateless Codes With Limited Common Randomness

The rateless codes for maximal error and nosy noise are very similar and can be described by the following algorithm.

1) The encoder and decoder choose a key $k \in [K(n)]$ using common randomness. The encoder chooses a message $i \in [N(n)]$ to transmit and maps it into a codeword $\mathbf{x}(i, k) \in \mathcal{X}^n$.
2) For $m = 1, 2, \ldots, M_* - 1$ the encoder transmits $\mathbf{x}^{(mc)}(i, k)$ in the mth chunk and the decoder sets the feedback bit $\tau_m(\mathbf{y}_1^{(m-1)c}, \hat{\lambda}_1^{m-1}, k) = 0$.
3) For $m = M_*, \ldots, M^* = n/c$, if $\tau_{m-1}(\mathbf{y}_1^{(m-1)c}, \hat{\lambda}_1^{m-1}, k) = 0$, the encoder transmits $\mathbf{x}^{(mc)}(i, k)$ in channel uses $(m-1)c+1, (m-1)c+2, \ldots, mc$.
4) The decoder receives channel outputs $\mathbf{y}^{(mc)}$. Under maximal error it also gets an estimate $\hat{\lambda}_m$ of the state cost in the mth chunk. Under nosy noise it gets a channel state information set $\mathcal{V}_m$. The decision function τ_m takes the form of a threshold test comparing the empirical rate $\frac{\log N}{mc}$ with a mutual information calculated from the channel output and partial CSI. If $\tau_m(\,\cdot\,) = 1$ then the decoder attempts to decode the received sequence, sets $\hat{i} = \Psi_m(\mathbf{y}_1^{mc}, k)$, and feeds back a 1 to terminate transmission. Otherwise, the decoder feeds back a 0 and we return to step 3) to send chunk $m+1$.

Our schemes use a fixed maximum blocklength n and we will express other parameters as functions of n. For a fixed minimum rate $R_{\min}$, input distribution P, and key size $K(n)$ we will construct a randomized rateless code with chunk size $c(n) = n^{1/4}$ and decoding time $M^*(n) = n^{3/4}$ (see (28) and (29)). Our codebooks will consist of codewords drawn uniformly from the set

$$(\mathcal{T}_c(P))^{n/c} = \underbrace{\mathcal{T}_c(P) \times \mathcal{T}_c(P) \times \cdots \mathcal{T}_c(P)}_{n/c \text{ times}}. \tag{33}$$

That is, the codewords are formed by concatenating constant-composition chunks of length c.

A. Rateless Codes for Maximal Error

We now prove Theorem 2 on rateless coding for AVCs under maximal error. Our code is built up from the code of [24] and we use Lemma 1 to partially derandomize the construction. In this section, we will assume the CSI takes the form of (14)–(15) and that it is ϵ-cost-consistent. Define

$$\Lambda_M = \frac{1}{M} \sum_{m=1}^{M} \lambda_m \tag{34}$$

$$\hat{\Lambda}_M = \frac{1}{M} \sum_{m=1}^{M} \hat{\lambda}_m. \tag{35}$$

These are the true and estimated cost for the state sequence $\mathbf{s}_1^{Mc}$. The number of possible values for λ_m is at most $(c+1)^{|\mathcal{S}|}$, which is an upper bound on the number of types on $\mathcal{S}$ with denominator c. Without loss of generality we can assume $\hat{\lambda}_m$ takes values in the same set as λ_m.

The decision function $\tau_m(\mathbf{y}_1^{mc}, \hat{\lambda}_1^m, k)$ for this code is given by

$$\mathbf{1}\left(\frac{\log N}{mc} < I(P, \mathcal{W}_{\mathrm{std}}(\hat{\Lambda}_m)) - \delta \right) \tag{36}$$

where $\hat{\Lambda}_m$ is given by (35).

Our code relies on the existence of a set of codewords $\{\mathbf{x}(i, k)\}$ which, when truncated to blocklength mc, form a good randomized code for an AVC satisfying a given cost constraint. The condition checked by the decision function (36) is sufficient to guarantee that the decoding error will be small.

We will also use a parameter $R_{\min}$ which is the minimum rate of the code, so $N(n) = \exp(nR_{\min})$.

Lemma 3 (Fully Randomized Rateless Codebook): Let $\mathcal{W}$ be an AVC with cost function $l(\cdot)$. For any $\delta > 0$, $R_{\min} > 0$ and input distribution $P \in \mathcal{P}(\mathcal{X})$ with $\min_x P(x) > 0$, for sufficiently large blocklength n there exists a randomized rateless code with $N(n) = \exp(nR_{\min})$ messages whose decoding time $\mathbf{M}$ satisfies (36) and whose rate at $\mathbf{M} = M$ satisfies

$$\frac{\log N}{Mc} < I\left(P, \mathcal{W}_{\mathrm{std}}\left(\frac{1}{Mc}\sum_{i=1}^{Mc} l(s_i)\right)\right) - f(\delta) \qquad (37)$$

for all $\mathbf{s}$ and δ-cost-consistent partial state information sequences $\mathcal{V}_1^M$, where $f(\delta) \to 0$ as $\delta \to 0$. The error at decoding time satisfies

$$\varepsilon(M, \mathbf{s}, \mathcal{V}_1^M) = O(\exp(-E_3(\delta)Mc)) \qquad (38)$$

where $E_3(\delta) > 0$.

1) Proof of Theorem 2: We are now ready to prove the Theorem 2.

Proof: Fix $\epsilon > 0$, $R_{\min} > 0$ and $P \in \mathcal{P}(\mathcal{X})$. Choose n sufficiently large so that the codebook-valued random variable $\mathbf{C}_{M^*}$ that is the randomized code from Lemma 3 satisfies (38) with $\epsilon = \delta$ under the conditions on the state and side information in (15) and (30). For each M, let $\mathbf{C}_M$ be the the codebook truncated to blocklength Mc.

We can now draw $K(n)$ codebooks sampled uniformly from $\mathbf{C}_{M^*}$. Since $\mathbf{C}_{M^*}$ truncated to blocklength Mc is $\mathbf{C}_M$, this sampling induces a sampling on $\mathbf{C}_M$ for each M. Each of these truncated codebooks has error probability exponentially small in Mc, so by Lemma 1 we can choose n sufficiently large and chunk size $c(n)$ so that with probability going to 1, the error probability is at most $O(n/K(n))$ for each of the truncated codes. Therefore a code satisfying the conditions of the theorem exists. ■

One case in which we can obtain δ-cost-consistent state information is in the scheme proposed by Eswaran *et al.* [12] for coding over a channel with individual state sequence. The codes from this section can be used as a component in that coding scheme, which is an iterated rateless coding strategy using zero-rate feedback and unlimited common randomness. One drawback of the scheme in [12] is that the amount of common randomness needed to choose the rateless code is very large. By using the rateless code constructed in Theorem 2 the amount of common randomness can be reduced and can be accommodated in the zero-rate feedback link.

B. Rateless Coding for Channels With Input-Dependent State

We now prove Theorem 3 on rateless coding for AVCs under nosy maximal error. The idea is to build rateless codes which are *list-decodable* with constant list size at the decoding time $\mathbf{M}$. Lemma 2 can be used with these list decodable codes to construct a randomized code with small key size.

We explicitly use information about the output sequence $\mathbf{y}$ at the decoder together with the side information $\mathcal{V}_m$. For $\delta > 0$ and distribution $P \in \mathcal{P}(\mathcal{X})$, given the mth chunk of channel outputs $\mathbf{y}^{(mc)}$ and the side information set $\mathcal{V}_m$, define

$$\mathcal{V}_m\left(\mathbf{y}^{(mc)}, \epsilon\right) = \left\{V \in \mathcal{V}_m : d_{\max}\left(T_{\mathbf{y}^{(mc)}}, \sum_x P(x)V(y \,|\, x)\right) < \epsilon\right\}$$

where $d_{\max}(\cdot,\cdot)$ is the total variational distance. Although $\mathcal{V}_m(\mathbf{y}^{(mc)}, \epsilon)$ depends on P, in our construction P is fixed so we do not make this dependence explicit. Define the decision function $\tau_m(\mathbf{y}_1^{mc}, \mathcal{V}_1^m, k)$ as

$$\mathbf{1}\left(\frac{\log N}{mc} < \frac{1}{m}\sum_{i=1}^{m} I\left(P, \mathcal{V}_m\left(\mathbf{y}^{(mc)}, \delta\right)\right) - \epsilon\right). \qquad (39)$$

Once the decision threshold $\mathbf{M}$ is reached, the decoder list-decodes the received codeword and produces a list of candidate message-key pairs. From Lemma 2, with high probability there will be only one message-key pair in the list consistent with the key used to encode the message.

The key lemma is Lemma 4, which shows that a codebook of concatenated constant-composition chunks can be list-decoded in each chunk using a channel estimate for that chunk. The overall list size is exponential in Mc. The decoding condition (39) can be used to bound the list size at decoding. The proof is omitted for space reasons (see [5]). The final step is to sample codewords from $(\mathcal{T}_c(P))^{n/c}$. The subsampling ensures a constant bound on the list size for all decoding times.

Lemma 4 (Concatenated Exponential List Codes): Let $\mathcal{W}$ be an AVC. For any $\delta > 0$ and $\xi > 0$, $P \in \mathcal{P}(\mathcal{X})$ with $\min_x P(x) > 0$, there is a c sufficiently large such that the set $(\mathcal{T}_c(P))^M$ is a list-decodable code with blocklength Mc, N_M messages and list size $L(\mathbf{y}_1^{Mc}, \mathcal{V}_1^M)$ for $\mathcal{V}_1^M = (\mathcal{V}_1, \mathcal{V}_2, \ldots, \mathcal{V}_M) \in \mathbb{V}(c)^M$, where $N_M \geq \exp(Mc(H(X) - \xi))$ and

$$L\left(\mathbf{y}_1^c, \mathcal{V}_1^M\right) \leq \exp\left(c\left(\sum_{m=1}^{M} \max_{V \in \mathcal{V}_m(\mathbf{y}^{(mc)}, \delta)} H \times (X_m \,|\, Y_m) + M\xi\right)\right). \qquad (40)$$

The maximal probability of error is

$$\varepsilon_L \leq M \exp(-cE_2(\xi)) \qquad (41)$$

where $H(X)$ is calculated with respect to the distribution $P(x)$ and for a channel $V \in \mathcal{V}_m(\mathbf{y}^{(mc)}, \delta)$ the conditional entropy $H(X \,|\, Y)$ is with respect to the distribution $P(x)V(y \,|\, x)$, and $E_2(\xi) > 0$.

Our codebook is constructed by sampling codewords from the codebook $(\mathcal{T}_c(P))^{n/c} = (\mathcal{T}_c(P))^{M^*}$. Truncating this set to blocklength Mc gives $(\mathcal{T}_c(P))^M$. We want to show that for each M the sampled codewords can be used in a list decodable code with constant list size L. We can define for each truncation M, output sequence $\mathbf{y}_1^{Mc}$, and side information sequence $(\mathcal{V}_1, \ldots, \mathcal{V}_M)$ a "decoding bin"

$$B\left(M, \mathbf{y}_1^{Mc}, \mathcal{V}_1^M\right) \subset \mathcal{X}^{Mc}$$

which is the list given by the code in Lemma 4. The size of each bin can be upper bounded by (40).

Lemma 5 (Constant List Size): Let $\mathcal{W}$ be an AVC with cost function $l(\cdot)$. For any $\epsilon > 0$, $P \in \mathcal{P}(\mathcal{X})$ with $\min_x P(x) > 0$, for sufficiently large blocklength n there exists a set of $N(n) = \exp(nR_{\min})$ codewords $\{\mathbf{x}(j) : j \in [N]\} \subset (\mathcal{T}_c(P))^{n/c}$ such that for any CSI sequence $(\mathcal{V}_1, \mathcal{V}_2, \ldots, \mathcal{V}_{M^*})$ and channel output $\mathbf{y}$ with decoding time $\mathbf{M}$ given by (39), the truncated codebook $\{\mathbf{x}_1^{Mc}(j) : j \in [N]\}$ is a list decodable code with list size L satisfying

$$L \geq \frac{12 \log |\mathcal{Y}|}{\epsilon}$$

and maximal probability of decoding error $\varepsilon_L(M) \leq M \exp(-cE(\epsilon))$, where $E(\epsilon) > 0$.

C. Proof of Theorem 3

Proof: We will use the codebook from Lemma 5. Since the set of messages is of fixed size N, we use the construction of Lemma 2. This makes the code, when decoded at after $\mathbf{M} = M$ chunks, an $(Mc, \exp(nR_{\min})/\sqrt{K(n)}, K(n))$ randomized code with probability of error

$$\hat{\varepsilon}(M, \mathbf{s}) \leq M \exp(-cE(\epsilon)) + \frac{2LnR_{\min}}{\sqrt{K} \log K}.$$

Then we can use choose $L = 12(\log |\mathcal{Y}|)/\epsilon$ to get

$$\hat{\varepsilon}(M, \mathbf{s}) \leq M \exp(-cE(\epsilon)) + \frac{24nR_{\min} \log |\mathcal{Y}|}{\epsilon \sqrt{K} \log K}.$$

Finally, we must show that the loss in rate is small, assuming ϵ-consistent state information. But this follows because by (16), for all m

$$I(P, V_m) - I(P, \mathcal{V}_m) \leq \epsilon.$$

Therefore, the average of mutual information terms in (39) is at most ϵ smaller than the averages with the true channels and hence we get the bound on the decoding time. ■

VI. Discussion

In this paper, we constructed rateless codes for two different channel models with time varying state based on AVCs. In the first model, the state cannot depend on the transmitted codeword, and in the second model it can. By adapting previously proposed derandomization strategies, we showed that a sublinear amount of common randomness is sufficient, which means that a secure control channel of small rate is sufficient to enable reliable communication. Our codes can partially derandomize the construction proposed in [12] for communicating over channels with individual state sequences.

We also showed the capacity formula $\hat{C}_r(\Lambda) = C_{\text{dep}}(\Lambda)$ for AVCs with "nosy noise" in which the jammer has knowledge of the transmitted codeword. Although in some examples $\hat{C}_r(\Lambda)$ may equal the capacity under maximal error $C_r(\Lambda)$, in general it is smaller. It is interesting to note that the jammer's worst strategy for nosy noise is to make a "memoryless attack" on the input by choosing the state s_t according the the minimizing conditional distribution $U(s|x_t)$ in (11). In constrast, if the jammer is given strictly causal knowledge of the input sequence, Blackwell *et al.* [13] showed that the capacity is given by C_{std}, which is also the capacity when the jammer has no knowledge of the input sequence. Thus from the jammer's perspective, causal information about $\mathbf{x}$ is as good as no knowledge, and full knowledge is as good as knowledge of the current input.

Appendix

A. Proof of Lemma 3

Proof: Fix $\delta > 0$, $R_{\min} > 0$ and P. We will prove that for each $M \in \mathcal{M} = \{M_*, \ldots, M^*\}$ there exists a randomized codebook $\mathbf{C}_M$ of blocklength Mc with rate $R_M = nR_{\min}/(Mc)$. Let $\tilde{\Lambda}_M$ be defined by

$$R_M = I(P, \mathcal{W}_{\text{std}}(\tilde{\Lambda}_M)) - \delta. \tag{42}$$

The distribution of the codebook $\mathbf{C}_M$ will be the same as the distribution of the codebook $\mathbf{C}_{M^*}$ of blocklength M^*c truncated to blocklength c.

Standard randomized codebook. Fix M and let $\mathbf{A}_M$ be a randomized codebook of A codewords drawn uniformly from the constant-composition set $\mathcal{T}_{Mc}(P)$ with maximum mutual information (MMI) decoding. Choose A such that

$$\frac{1}{Mc} \log A < I(P, \mathcal{W}_{\text{std}}(\tilde{\Lambda}_M)) - \delta/2.$$

From Hughes and Thomas [24, Theorem 1] the following exponential error bound holds for all messages i and state sequences $\mathbf{s} \in \mathcal{S}^{Mc}$ with $l(\mathbf{s}) \leq (Mc)\tilde{\Lambda}_M$

$$\delta_M(\mathbf{A}_M, i, \mathbf{s}) \leq \exp\left(-Mc\left(E_r\left(\frac{1}{Mc} \times \log A + \delta/2, P, \tilde{\Lambda}_M\right) - \delta/2\right)\right). \tag{43}$$

Let ζ_M denote this upper bound. The exponent $E_r(\cdot)$ is positive as long as the first argument is smaller than $I(P, \mathcal{W}_{\text{std}}(\tilde{\Lambda}_M))$. Therefore

$$\frac{1}{A} \sum_{i=1}^{A} \delta_M(\mathbf{A}_M, i) \leq \zeta_M.$$

Thinning. Let $\mathbf{B}_M$ be a random codebook formed by selecting B piecewise constant-composition codewords uniformly from $\mathcal{A}_M^* = \mathbf{A}_M \cap (\mathcal{T}_c(P))^M$. We declare an encoding error if $|\mathcal{A}_M^*| < B$. We have [12]

$$\frac{|\mathcal{T}_c(P)|^M}{|\mathcal{T}_{Mc}(P)|} \geq \exp(-M \log(Mc)\eta(P)) \triangleq \gamma_M \tag{44}$$

where $\eta(P) < \infty$ is a positive constant. Since $\mathbf{A}_M$ is formed by iid draws from $\mathcal{T}_{Mc}(P)$, the event that codeword i from $\mathbf{A}_M$ is also in $(\mathcal{T}_c(P))^M$ is a Bernoulli random variable with parameter at least γ_M. The size of $|\mathcal{A}_M^*|$ can be upper bounded using Sanov's Theorem [35]

$$\mathbb{P}(|\mathcal{A}_M^*| < B) \leq (A+1)^2 \exp(-A \cdot D(B/A \,|\, \gamma_M)).$$

Choose $B = \gamma_{M^*}^2 A$. Then we can make the probability that $|\mathcal{A}_M^*| < B$ as small as we like and much smaller than the decoding error bound. Furthermore, this bound holds for all $M \in \mathcal{M}$. Therefore a subcodebook of B piecewise constant-composition codewords exists with high probability.

The encoder draws $\mathbf{B}_M$ and declares an error if $|\mathbf{B}_M| < B$. If there is no error it transmits the ith codeword in the codebook for message $i \in [B]$. The average error on the fraction $B/A = \gamma_{M^*}^2$ of preserved codewords can be at most A/B times the original average error

$$\frac{1}{B} \sum_{i=1}^{B} \delta_M(\mathbf{B}_M, i) \leq \frac{\zeta_M}{\gamma_{M^*}^2}.$$

Permutation. Define a randomized code $\mathbf{C}_M$ by drawing a permutation π uniformly from all permutations on $[B]$ and code $\mathcal{B}$ according to $\mathbf{B}_M$ and encoding message i with the codeword $\pi(i)$ from $\mathcal{B}$. The maximal error of $\mathbf{C}_M$ is the same as average error for $\mathbf{B}_M$, so $\delta_M(\mathbf{C}_M, i) < \zeta_M/\gamma_{M^*}^2$. For each M we can construct a randomized codebook $\mathbf{C}_M$ as aforedescribed.

Nesting. Now consider the codebook $\mathbf{C}_{M^*}$ of blocklength $n = M^* c$ and set the size of the codebook to B to equal $N(n) = \exp(n R_{\min})$. We must guarantee that the errors will still be small. Since $B = \gamma_{M^*}^2 A$, the rate of the codebook $\mathbf{A}_{M^*}$ is $\rho_{M^*} = (M^* c)^{-1} \log N/\gamma_{M^*}^2$. If we truncate $\mathbf{C}_{M^*}$ to blocklength Mc, the resulting randomized code is identically distributed to $\mathbf{C}_M$. The rate for the corresponding $\mathbf{A}_M$ can be bounded using (44), (28), and (29)

$$\rho_M \leq R_M + 2\eta(P) \frac{R_{\max}}{R_{\min}} \cdot \frac{\log n}{n^{1/4}}.$$

Therefore, we can choose n sufficiently large so that the gap between ρ_M and R_M can be made smaller than $\delta/2$, so $\rho_M < R_M + \delta/2$. Therefore using the definition of $\tilde{\Lambda}_M$ in (42) and the fact that $\tilde{\Lambda}_M \geq \hat{\Lambda}_M$ we have

$$\rho_M < I(P, \mathcal{W}_{\text{std}}(\tilde{\Lambda}_M)) - \delta/2$$

and the exponent in (43) is positive. Now, for $(\mathbf{s}, \{\hat{\lambda}_m\})$ such that (37) holds, the error is less than $\frac{\zeta_M}{\gamma_{M^*}^2}$, so we have (38).

Rate loss. The last step is to compare $I(P, \mathcal{W}_{\text{std}}(\hat{\Lambda}_M))$ to the empirical mutual information induced by the true state sequence. By assumption, the partial CSI is δ-cost-consistent, so by (16), $\Lambda_M \leq \hat{\Lambda}_M \leq \Lambda_M + \delta$. Therefore

$$I(P, \mathcal{W}_{\text{std}}(\Lambda_M)) - I(P, \mathcal{W}_{\text{std}}(\hat{\Lambda}_M)) = O(\delta \log \delta^{-1}).$$

By the triangle inequality and (36)

$$I(P, \mathcal{W}_{\text{std}}(\Lambda_M)) - \frac{\log N}{mc} = O(\delta \log \delta^{-1}).$$

This proves (37). ■

B. Proof of Lemma 5

Proof: Fix $\epsilon > 0$ and consider the codebook $(\mathcal{T}_c(P))^{M^*}$. The truncation of this codebook to blocklength Mc for $M \in \mathcal{M}$ is the codebook defined in Lemma 4. Let $\{\mathbf{Z}_j : j \in [N]\}$ be $N = \exp(n R_{\min})$ random variables distributed uniformly on the set $(\mathcal{T}_c(P))^{M^*}$. The decoder will operate in two steps: first it will decode the received sequence into the exponential size list $B(M, \mathbf{y}_1^{Mc}, \mathcal{V}_1^M)$ given by the decoder of Lemma 4, and then it will output only those codewords in the list which match one of the sampled codewords $\{\mathbf{Z}_j\}$. Note that the decoder for Lemma 4 has error satisfying (41).

For any $\delta > 0$ and $\xi > 0$ we can choose $c(n)$ sufficiently large so that for any fixed M, $\mathbf{y}_1^{Mc}$, and $\mathcal{V}_1^M \in \mathbb{V}(c)^M$ that satisfy the conditions of the decoding rule in (39) the probability that $\mathbf{Z}_j$ lands in the list $B(M, \mathbf{y}_1^{Mc}, \mathcal{V}_1^M)$ output by the decoder of Lemma 4 is upper bounded

$$\mathbb{P}\left(\mathbf{Z}_j \in B\left(M, \mathbf{y}_1^{Mc}, \mathcal{V}_1^M\right)\right) \leq \exp\left(-c \sum_{m=1}^{M} I\left(P, \mathcal{V}_m\left(\mathbf{y}^{(mc)}, \delta\right)\right) + 2Mc\xi\right).$$

The random variable $\mathbf{1}(\mathbf{Z}_j \in B(M, \mathbf{y}_1^{Mc}, \mathcal{V}_1^M))$ is Bernoulli with parameter smaller than G, so we can bound the probability that L of the N codewords $\{\mathbf{Z}_j\}$ land in the set $B(M, \mathbf{y}_1^{Mc}, \mathcal{V}_1^M)$ using Sanov's Theorem [35]

$$\mathbb{P}\left(\frac{1}{N} \sum_{j=1}^{N} \mathbf{1}\left(\mathbf{Z}_j \in B\left(M, \mathbf{y}_1^{Mc}, \mathcal{V}_1^M\right)\right) > L/N\right) \leq (N+1)^2 \exp(-N D(L/N \,\|\, G)).$$

We can bound the exponent by

$$L\left(-n R_{\min} + c \sum_{m=1}^{M} I(P, \mathcal{V}_m(\mathbf{y}^{(mc)}, \delta)) - 2Mc\xi\right) + L \log L - 2L. \tag{45}$$

From the rule in (39), we know that $(M, \mathbf{y}_1^{Mc}, \mathcal{V}_1^M)$ satisfies

$$n R_{\min} < c \sum_{m=1}^{M} I\left(P, \mathcal{V}_m(\mathbf{y}^{(mc)}, \delta)\right) - Mc\epsilon. \tag{46}$$

Substituting this into (45) we see that

$$N D(L/N \,\|\, G) > L(Mc\epsilon - 2Mc\xi) + L \log L - 2L.$$

For large enough n we have the bound $(N+1)^2 \leq 2n\rho + L$. For large enough L, $L \log L > 3L$, so we can ignore those terms as well. This gives the following bound:

$$\mathbb{P}\left(\frac{1}{N} \sum_{j=1}^{N} \mathbf{1}\left(\mathbf{Z}_j \in B(M, \mathbf{y}_1^{Mc}, \mathcal{V}_1^M)\right) > L/N\right) \leq \exp(-LMc(\epsilon - 2\xi) + 2n R_{\min}). \tag{47}$$

The number of decoding bins $B(M, \mathbf{y}_1^{Mc}, \mathcal{V}_1^M)$ can be bounded by $|\mathcal{Y}|^{Mc} c^{Mv}$. Therefore we can take a union bound over all the decoding bins in (47) to get an upper bound of

$$\exp(-LMc(\epsilon - 2\xi) + Mc \log |\mathcal{Y}| + Mv \log c + 2n R_{\min}).$$

Since $\frac{n R_{\min}}{Mc} \leq \log |\mathcal{Y}|$ for all $M \geq M_*$, we can choose n and c sufficiently large such that the upper bound becomes

$$\exp(-LMc(\epsilon - 2\xi) + 4Mc \log |\mathcal{Y}|).$$

If $\epsilon > 2\xi$ then we can choose $L > 4\log|\mathcal{Y}|/(\epsilon - 2\xi)$ to guarantee that subsampling will yield a good list-decodable code for all $M \in \{M_*, \ldots, M^*\}$. Choosing $\xi = \epsilon/3$ and $E(\epsilon) = E_2(\epsilon/3)$, where $E_2(\cdot)$ is from (41), yields the result. ■

Acknowledgment

The authors would like to thank the anonymous reviewers for their extensive comments and suggestions which greatly clarified and simplified the manuscript. They would also like to thank the Associate Editor, I. Kontoyiannis, for his many efforts and constructive comments. Finally, they also benefited from fruitful discussions with K. Eswaran and A. Sahai.

References

[1] A. Sarwate and M. Gastpar, "Channels with nosy "noise"," in *Proc. 2007 Int. Symp. Inf. Theory*, Nice, France, 2007, submitted for publication.

[2] A. Sarwate and M. Gastpar, "Rateless coding with partial CSI at the decoder," in *Proc. 2007 Inf. Theory Workshop*, Lake Tahoe, CA, Sep. 2007.

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

[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, 1993.

[5] A. Sarwate, "Robust and adaptive communication under uncertain interference," Ph.D. dissertation, Univ. Calif., Berkeley, Jul. 2008.

[6] M. Langberg, "Private codes or succinct random codes that are (almost) perfect," in *Proc. 45th Ann. IEEE Symp. Found. Comput. Sci. (FOCS 2004)*, Rome, Italy, 2004.

[7] M. Luby, "LT codes," in *Proc. 43rd Ann. IEEE Symp. Found. Comput. Sci.*, 2002.

[8] A. Shokrollahi, "Fountain codes," in *Proc. 41st Allerton Conf. Commun., Control, Comput.*, Oct. 2003, pp. 1290–1297.

[9] N. Shulman, "Communication over an unknown channel via common broadcasting," Ph.D. dissertation, Tel Aviv Univ., Tel Aviv, Israel, 2003.

[10] A. Tchamkerten and I. E. Telatar, "Variable length coding over an unknown channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2126–2145, May 2006.

[11] S. S. (Shitz), I. E. Telatar, and S. Verdú, "Fountain capacity," *IEEE Trans. Inf. Theory*, vol. 53, no. 11, pp. 4372–4376, Nov. 2007.

[12] K. Eswaran, A. Sarwate, A. Sahai, and M. Gastpar, "Zero-rate feedback can achieve the empirical capacity," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 25–39, Jan. 2010.

[13] 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.

[14] J. Kiefer and J. Wolfowitz, "Channels with arbitrarily varying channel probability functions," *Inf. Control*, vol. 5, no. 1, pp. 44–54, 1962.

[15] R. Ahlswede and J. Wolfowitz, "The capacity of a channel with arbitrarily varying channel probability functions and binary output alphabet," *Zeitschrift für Wahrscheinlichkeit und verwandte Gebiete*, vol. 15, no. 3, pp. 186–194, 1970.

[16] R. Ahlswede, "A method of coding and an application to arbitrarily varying channels," *J. Combinator., Inf. Syst. Sci.*, vol. 5, pp. 10–35, 1980.

[17] 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.

[18] 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, pp. 1027–1033, 1970.

[19] B. Hughes and P. Narayan, "Gaussian arbitrarily varying channels," *IEEE Trans. Inf. Theory*, vol. 33, no. 2, pp. 267–284, 1987.

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

[21] 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, 1988.

[22] I. Csiszár and P. Narayan, "Capacity of the Gaussian arbitrarily varying channel," *IEEE Trans. Inf. Theory*, vol. 37, no. 1, pp. 18–26, 1991.

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

[24] B. Hughes and T. Thomas, "On error exponents for arbitrarily varying channels," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 87–98, 1996.

[25] T. Thomas and B. Hughes, "Exponential error bounds for random codes on Gaussian arbitrarily varying channels," *IEEE Trans. Inf. Theory*, vol. 37, no. 3, pp. 643–649, 1991.

[26] 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.

[27] R. Ahlswede, "Channels with arbitrarily varying channel probability functions in the presence of noiseless feedback," *Zeitschrift für Wahrscheinlichkeit und verwandte Gebiete*, vol. 25, pp. 239–252, 1973.

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

[29] A. Smith, "Scrambling adversarial errors using few random bits, optimal information reconciliation, and better private codes," in *Proc. 2007 ACM-SIAM Symp. Discrete Algorithms (SODA 2007)*, 2007.

[30] M. Agarwal, A. Sahai, and S. Mitter, "Coding into a source: A direct inverse rate-distortion theorem," in *Proc. 45th Ann. Allerton Conf. Commun., Contr. Computat.*, 2006.

[31] S. Draper, B. Frey, and F. Kschischang, "Rateless coding for non-ergodic channels with decoder channel state information," *IEEE Trans. Inf. Theory*, submitted for publication.

[32] N. Shulman and M. Feder, "The uniform distribution as a uniform prior," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 1356–1362, Jun. 2004.

[33] N. Merhav and N. Feder, "Universal prediction," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2124–2147, Oct. 1998.

[34] P. Erdös, P. Frankl, and Z. Füredi, "Families of finite sets in which no set is covered by the union of r others," *Israel J. Math.*, vol. 51, no. 1–2, pp. 79–89, 1985.

[35] T. Cover and J. Thomas*, Elements of Information Theory*. Hoboken, NJ: Wiley, 1991.

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 (UC Berkeley).

He is currently a Postdoctoral Researcher with the Information Theory and Applications Center, University of California, San Diego. His research interests include information theory, distributed signal processing, machine learning, communications, and randomized algorithms for communications and signal processing in sensor networks.

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, UC Berkeley. He was awarded an NDSEG Fellowship from 2002 to 2005. He is a member of Phi Beta Kappa and Eta Kappa Nu.

Michael C. Gastpar (M'04) received the Dipl. El.-Ing. degree from the Swiss Federal Institute of Technology (ETH), Zürich, in 1997, the M.S. degree from the University of Illinois at Urbana-Champaign, Urbana, in 1999, and the Doctorat ès 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 University of Edinburgh, Edinburgh, U.K., and the University of Lausanne.

He is currently an Associate Professor at the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, and a Full Professor in the Faculty of Electrical Engineering, Mathematics and Computer Science (EEMCS), Delft University of Technology, Delft, The Netherlands. He was a summer researcher at 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, and an Okawa Foundation Research Grant in 2008.