Computationally Efficient Codes for Strongly Dobrushin-Stambler Nonsymmetrizable Oblivious AVCs

B. K. Dey
IIT Bombay
bikash@ee.iitb.ac.in

S. Jaggi
University of Bristol
sid.jaggi@bristol.ac.uk

M. Langberg
SUNY Buffalo
mikel@buffalo.edu

A. D. Sarwate
Rutgers University
ads221@rutgers.edu

Y. Zhang
IST Austria
zephyr.z798@gmail.com

***Abstract*—We propose a concatenated code construction for a class of discrete-alphabet oblivious arbitrarily varying channels (AVCs) with cost constraints. The code has time and space complexity polynomial in the blocklength n. It uses a Reed-Solomon outer code, logarithmic blocklength random inner codes, and stochastic encoding by permuting the codeword before transmission. When the channel satisfies a condition called strong DS-nonsymmetrizability (a modified version of nonsymmetrizability originally due to Dobrushin and Stambler), we show that the code achieves a rate that for a variety of oblivious AVCs (such as classically studied error/erasure channels) match the known capacities.**

I. Introduction

In this paper we describe a *computationally efficient* code construction for *arbitrarily varying channels* (AVCs) with an *oblivious adversary*. In order to make sense of the preceding sentence, a few definitions are in order. We consider a communication problem in which Alice wishes to send a message to Bob over a channel with discrete input alphabet $\mathcal{X}$ and discrete output alphabet $\mathcal{Y}$. The channel is modeled by an *AVC*, which is defined by a conditional distribution $W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(y|x,s)$ that has state variable s taking values in a discrete alphabet $\mathcal{S}$. We assume s can be chosen by a third party, James, who is an adversarial jammer.

We consider block codes for this channel so that Alice can encode a message into a codeword $\underline{x} \in \mathcal{X}^n$, James can choose a state sequence $\underline{s} \in \mathcal{S}^n$, and Bob observes an output $\underline{y} \in \mathcal{Y}^n$ formed by applying the channel $W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(y|x,s)$ letter-by-letter to $(\underline{x},\underline{s})$. We study general AVCs with discrete inputs, outputs, and states and are interested in coding strategies which are *computationally efficient*, by which we mean the encoding, decoding, and storage of the code uses time/space growing at most polynomially in the blocklength n. For a general discrete memoryless channel (DMC), *random codes* are capacity-achieving but are not computationally efficient. Designing computationally efficient codes is a first step towards more practical codes.

The AVC can model a variety of communication scenarios by assuming different limitations on James's allowable strategies and his access to the input to the channel. If James has no control over the state and $\underline{s}$ is drawn i.i.d. from a fixed distribution $Q_{\mathbf{s}}(s)$ then the AVC becomes a DMC. If James has to use the same $s \in \mathcal{S}$ for every channel use, the AVC becomes a compound channel. The more interesting cases are when James can choose $\underline{s}$ arbitrarily subject to some constraints. For example, if the AVC is an erasure channel where $\mathcal{S} = \{0,1\}$ and $s = 1$ indicates an erasure, James could be constrained to erasing only a certain fraction of inputs, which puts a constraint on the Hamming weight of $\underline{s}$.

We are interested in the case where Alice and Bob do not share any common randomness. We say James is *omniscient* if he can choose $\underline{s}$ based on the actual input $\underline{x}$. In the erasure example we just mentioned, this is the model used in a first course in coding theory: a code which corrects pn erasures must correct *every possible pattern* of pn erasures. In this paper we study the *oblivious* model where James does not have access to the transmitted codeword. That is, James must decide on the state $\underline{s}$ before Alice selects the codeword $\underline{x}$ to be transmitted. This means that Alice can potentially use *stochastic encoding* to limit the harm of any particular $\underline{s}$: in particular, we show that Alice can randomly permute the codeword positions such that (a) James cannot launch an effective attack and (b) Bob can successfully guess the permutation as part of the decoder. Unlike Ahlswede's "robustification" approach [1], Bob and James are both ignorant of the permutation chosen by Alice.

A. Related work

The AVC model was first proposed by Blackwell, Breiman, and Thomasian [2] in 1960 and many extensions and variations of the original model have been studied. The survey of Lapidoth and Narayan [3] covers many of these earlier results. One feature of many AVC models is that the capacity exhibits a "dichotomy": the capacity is either zero or some positive number [4]. Intuitively, if James can mimic or "spoof" Alice by choosing a valid codeword $\underline{x}'$ and using it to generate $\underline{s}$ that makes the AVC effectively a symmetric multiaccess channel (MAC) $W_{\mathbf{y}|\mathbf{x},\mathbf{x}'}(y|x,x')$, Bob will be unable to tell if Alice or James is the legitimate transmitter. When this scenario occurs we say the channel is *symmetrizable* [5]–[7]. Whether or not the AVC is symmetrizable can depend on the constraints on James's choice of $\underline{s}$.

The capacity of the AVC with an oblivious adversary was characterized by Csiszár and Narayan [8], who gave one definition of symmetrizability (CN-symmetrizability) based

979-8-3503-8284-6/24/$31.00 ©2024 IEEE

on that of Ericson [6]. Earlier, Dobrushin and Stambler provided a different definition of symmetrizability (DS-symmetrizability) [5]. The class of DS-*non*symmetrizable channels is a proper subset of the class of CN-*non*symmetrizable channels. In this paper, we define "strong DS nonsymmetrizability" which identifies a further subset of the DS-nonsymmetrizable class. We show that our codes achieve positive rate for AVCs which are strongly DS non-symmetrizable, and that this rate matches capacity for channels such as oblivious q-ary additive/erasure channels (widely studied in the literature) which fall within this class.

Work in the theoretical computer science community has examined models where James is computationally bounded, meaning James must produce $\underline{s}$ using a polynomial-time algorithm or "small" circuit [9]. Guruswami and Smith [10] studied efficient codes in this model and also showed efficient list decoding constructions, which were later strengthened [11], [12]. More recently, Shaltiel and Silbak showed explicit constructions of list- [13], [14] and uniquely-decodable codes for bit-flipping channels [15] that achieve rates beyond the Gilbert-Varshamov bound when James is computationally bounded. Yasunaga analyzed a class of computationally limited attacks [16] and proposed a different model which constrains the sampling complexity for James [17]. Rather than constraining James's actions, Ruzomberka et al. constrain the complexity of James' observations of the codeword [18]. Our results are close in spirit to Li et al., who studied code constructions for erasure models in the causal and myopic adversarial models [19].

B. Our contribution

We propose and analyze a code for general discrete AVCs with polynomial time encoding and decoding complexity. The construction uses a concatenated code and encoder randomization to limit how "targeted" James's attack can be. The code achieves positive rates under a new definition of symmetrizability based on that of Dobrushin and Stambler [5]. For known classes of AVCs for which efficient codes exist [10] our codes match their performance and are capacity-achieving, but our codes hold for larger classes of AVCs.

II. Problem statement and main results

Notation: We will generally use lowercase for scalars, underline for vectors, and boldface for random variables. The set $\mathbb{F}_q$ is the finite field of order q. For a positive integer n we define $[n] = \{1, 2, \dots, n\}$. For a finite alphabet $\mathcal{X}$ the set $\Delta(\mathcal{X})$ is the probability simplex on $\mathcal{X}$ (the set of all probability mass functions on $\mathcal{X}$). Distributions and conditional distributions will be given capital letters, with subscripts indicating the corresponding variables, so $P_{\mathbf{x},\mathbf{y}}(x,y)$ is a joint distribution on the pair of random variables $(\mathbf{x}, \mathbf{y})$ and $P_{\mathbf{y}|\mathbf{x}}(y|x)$ is the conditional distribution of $\mathbf{y}$ given $\mathbf{x}$. For a joint distribution $P_{\mathbf{x},\mathbf{s}} \in \Delta(\mathcal{X}) \times \Delta(\mathcal{S})$ the marginal distribution on $\mathcal{X}$ is denoted by $[P_{\mathbf{x},\mathbf{s}}]_{\mathbf{x}}$. For $P_{\mathbf{a}|\mathbf{b}}, Q_{\mathbf{a}|\mathbf{b}} \in \Delta(\mathcal{A}|\mathcal{B})$, define

$$\left\| P_{\mathbf{a}|\mathbf{b}} - Q_{\mathbf{a}|\mathbf{b}} \right\|_1 \triangleq \sum_{b \in \mathcal{B}} \left\| P_{\mathbf{a}|\mathbf{b}=b} - Q_{\mathbf{a}|\mathbf{b}=b} \right\|_1 \tag{1}$$

For a sequence $\underline{x} \in \mathcal{X}^n$, the type (empirical distribution) of $\underline{x}$ is denoted by $T_{\underline{x}}$. The set Σ_n is the set of permutations on $[n]$. For a sequence $\underline{x} = (x_1, \cdots, x_n) \in \mathcal{X}^n$ and a permutation $\pi \in \Sigma_n$, we write

$$\pi(\underline{x}) \triangleq \left(x_{\pi^{-1}(1)}, \cdots, x_{\pi^{-1}(n)}\right) \in \mathcal{X}^n.$$

where π^{-1} denotes the inverse of π. For a subset $\mathcal{I} \subset [n]$, denote by $\underline{x}_{\mathcal{I}} \in \mathcal{X}^{|\mathcal{I}|}$ the vector obtained by restricting $\underline{x}$ to indices in $\mathcal{I}$.

Logarithms to the base 2 and e are denoted by $\log$ and $\ln$, respectively. We will write $\exp_2(n) = 2^n$. The binary entropy function is $H(\cdot)$.

A. Channel model and constraints

Let $\mathcal{X}$, $\mathcal{S}$, and $\mathcal{Y}$ be finite discrete sets (alphabets). An *arbitrarily varying channel* (AVC) is a collection of channels $\{W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(y|x,s) \colon s \in \mathcal{S}\}$ indexed by $\mathcal{S}$. If $\underline{x} = (x_1, x_2, \dots, x_n)$, $\underline{y} = (y_1, y_2, \dots, y_n)$ and $\underline{s} = (s_1, s_2, \dots, s_n)$ are length n vectors, the probability of observing the output $\underline{y}$ given the input $\underline{x}$ and state $\underline{s}$ over the AVC $\mathcal{W}$ is given by:

$$W_{\underline{\mathbf{y}}|\underline{\mathbf{x}},\underline{\mathbf{s}}}(\underline{y}|\underline{x},\underline{s}) = \prod_{i=1}^{n} W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(y_i|x_i,s_i). \tag{2}$$

We consider sets of vectors $\underline{x}$ and $\underline{s}$ that are subject to cost constraints. Let γ and ξ be nonnegative bounded scalar-valued cost functions on $\mathcal{X}$ and $\mathcal{S}$ respectively, and let g and p be scalars. Then define the sets

$$\Gamma(g) = \{P_{\mathbf{x}}(x) \colon \mathbb{E}_{P_{\mathbf{x}}}[\gamma(x)] \le g\} \subseteq \Delta(\mathcal{X}) \tag{3}$$

$$\Xi(p) = \{Q_{\mathbf{s}}(s) \colon \mathbb{E}_{Q_{\mathbf{s}}}[\xi(s)] \le p\} \subseteq \Delta(\mathcal{S}). \tag{4}$$

We define an AVC as the tuple $\mathcal{A} = (\Gamma, \Xi, W_{\mathbf{y}|\mathbf{x},\mathbf{s}})$ consisting of the input constraint set Γ, the state constraint set Ξ, and the channel law $W_{\mathbf{y}|\mathbf{x},\mathbf{s}}$.

B. Codes and jamming strategies

We model a scenario in which a transmitter (named Alice) wishes to send a message to a receiver (named Bob) over the AVC whose state s is controlled by an adversary (James the jammer). A (n, M) code of blocklength n for the AVC $\mathcal{A}$ is a pair of maps $(\boldsymbol{\phi}, \psi)$ where $\boldsymbol{\phi} \colon [M] \to \mathcal{X}^n$ and $\psi \colon \mathcal{Y}^n \to [M]$. We allow the encoder $\boldsymbol{\phi}$ to be a stochastic map. The code must satisfy the input cost constraint so $T_{\boldsymbol{\phi}(m)} \in \Gamma$ almost surely over the encoder randomness. We denote the decoding region for message m as $\mathcal{D}_m = \{\underline{y} : \psi(\underline{y}) = m\}$. The rate of the code is $R = \frac{1}{n}\log_2(M)$ bits per channel use.

In the *oblivious* AVC model, Alice encodes the message $m \in [M]$ into a (possibly random) codeword $\underline{\mathbf{x}}(m) \in \mathcal{X}^n$ of blocklength n. James, who knows the maps ϕ and ψ but not the realization of $\underline{\mathbf{x}}(m)$, uses a (possibly stochastic) jamming strategy $J_{\underline{\mathbf{s}}}$ (a distribution on state vectors in Ξ) to choose a state sequence $\underline{\mathbf{s}}$. The inputs $\underline{\mathbf{x}}(m)$ and $\underline{\mathbf{s}}$ produce the output $\underline{\mathbf{y}}$ according to the distribution in (2). The jamming strategy must satisfy the state cost constraint so $T_{\underline{s}} \in \Xi$ almost surely over the jammer's randomness. We let $\mathfrak{J}(\Xi)$ denote the set of jamming strategies which satisfy the cost constraint.

The probability of correct decoding for a fixed $(m, \underline{x}, \underline{s})$ is

$$\nu(m, \underline{x}, \underline{s}) = \sum_{\underline{y} \in \mathcal{D}_m} W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(\underline{y}|\underline{x}, \underline{s}) J_{\underline{\mathbf{s}}}(\underline{s}) \mathbb{1}\{\boldsymbol{\phi}(m) = \underline{x}\}.$$

The error for message $m \in [M]$ and state sequence $\underline{s}$ is

$$\varepsilon(m, (\boldsymbol{\phi}, \psi), \underline{s}) = 1 - \mathbb{E}_{\boldsymbol{\phi}} \left[\sum_{\underline{x}, \underline{s}} \nu(m, \underline{x}, \underline{s}) \right]. \tag{5}$$

The maximal probability of error is then

$$\hat{\varepsilon}((\boldsymbol{\phi}, \psi)) = \sup_{J_{\underline{\mathbf{s}}} \in \mathfrak{J}(\Xi)} \max_{m \in [M]} \sum_{\underline{s} \in \Xi} \varepsilon(m, (\boldsymbol{\phi}, \psi), \underline{s}) J_{\underline{\mathbf{s}}}(\underline{s}). \tag{6}$$

A rate R is achievable under maximal error if there is a sequence of (n, M) codes $\{(\boldsymbol{\phi}_n, \psi_n)\}$ such that $\frac{1}{n} \log_2 M \geq R$ and $\hat{\varepsilon}((\boldsymbol{\phi}_n, \psi_n)) \to 0$.

In this paper we are interested in the rates achievable using *computationally efficient* codes. We say the sequence of codes achieving rate R is computationally efficient if the encoder $\boldsymbol{\phi}(\cdot)$ and decoder $\psi(\cdot)$ can be computed in time and space polynomial in the blocklength n and the encoder uses a number of uniform random bits that is also polynomial in n. To do this we will use a concatenated coding structure in which message bits are encoded using a Reed-Solomon code over a finite field $\mathbb{F}_q$ and the symbols $\mathbb{F}_q$ are mapped to sequences of symbols in $\mathcal{X}$.

C. Types of symmetrizability

In the AVC literature, the notion of *symmetrizability* captures whether James can fool Bob into decoding incorrectly. There are several different types of symmetrizability for AVC models. Dobrushin and Stambler [5] introduced a definition in which James can symmetrize the channel if he can use a memoryless channel $U_{\mathbf{s}|\mathbf{x}}(s|x)$ to generate a state sequence $\underline{s}$ satisfying the cost constraint from a codeword $\underline{x}' \neq \underline{x}$ such that the received vector $\underline{y}$ is likely to be jointly typical with $\underline{x}'$ for some feasible (single-letter) state distribution. They showed that AVCs which are *not* symmetrizable in this manner admit a positive rate via a single-step typicality-based decoder.

Let $\mathcal{A} = (\Gamma(g), \Xi(p), W_{\mathbf{y}|\mathbf{x},\mathbf{s}})$ be an oblivious arbitrarily varying channel (AVC) with input constraints $\Gamma(g) \subset \Delta(\mathcal{X})$, state constraints $\Xi(p) \subset \Delta(\mathcal{S})$ and channel law $W_{\mathbf{y}|\mathbf{x},\mathbf{s}} \in \Delta(\mathcal{Y}|\mathcal{X}, \mathcal{S})$.

Definition 1 (DS-symmetrizability [5])**.** *For an oblivious AVC $\mathcal{A} = (\Gamma(g), \Xi(p), W_{\mathbf{y}|\mathbf{x},\mathbf{s}})$, an input distribution $P_{\mathbf{x}} \in \Gamma(g)$ is* Dobrushin–Stambler (DS) symmetrizable *if there exists $Q_{\mathbf{s}} \in \Xi(p)$ and $U_{\mathbf{s}|\mathbf{x}} \in \Delta(\mathcal{S}|\mathcal{X})$ with $\left[U_{\mathbf{s}|\mathbf{x}} P_{\mathbf{x}}\right]_{\mathbf{s}} \in \Xi(p)$ such that for all $(x', y) \in \mathcal{X} \times \mathcal{Y}$,*

$$[P_{\mathbf{x}} W_{\mathbf{y}|\mathbf{x},\mathbf{s}} U_{\mathbf{s}|\mathbf{x}'}]_{\mathbf{y}|\mathbf{x}'} = [W_{\mathbf{y}|\mathbf{x}',\mathbf{s}} Q_{\mathbf{s}}]_{\mathbf{y}|\mathbf{x}'}. \tag{7}$$

We call an input distribution $P_{\mathbf{x}}$ weakly DS-symmetrizable *if the above holds in absence of the state constraint $\left[U_{\mathbf{s}|\mathbf{x}} P_{\mathbf{x}}\right]_{\mathbf{s}} \in \Xi(p)$.*

Let $\Delta_{\mathrm{DS}}(\mathcal{A}) \subset \Delta(\mathcal{X})$ denote the set of all weakly DS-symmetrizable input distributions on $\mathcal{X}$.

Definition 2 (Strong DS-nonsymmetrizability)**.** *We say that $P_{\mathbf{x}} \in \Gamma(g)$ is strongly* DS-nonsymmetrizable *if there exists a $\delta > 0$ such that for every $(Q_{\mathbf{s}}, U_{\mathbf{s}|\mathbf{x}}) \in \Delta(\mathcal{S}) \times \Delta(\mathcal{S}|\mathcal{X})$ with $Q_{\mathbf{s}} \in \Xi(p)$,*

$$\left\| \left[P_{\mathbf{x}} W_{\mathbf{y}|\mathbf{x},\mathbf{s}} U_{\mathbf{s}|\mathbf{x}'}\right]_{\mathbf{y}|\mathbf{x}'} - \left[W_{\mathbf{y}|\mathbf{x}',\mathbf{s}} Q_{\mathbf{s}}\right]_{\mathbf{y}|\mathbf{x}'} \right\|_1 \geq \delta. \tag{8}$$

The class of strongly DS-nonsymmetrizable channels is non-empty: consider a bit-flipping adversarial channel under no input constraint and a Hamming weight constraint $\mathrm{wt}(\underline{s}) \leq np$ (for some $p < 1/2$) on the state vector. It is easy to check that this channel is strongly DS-nonsymmetrizable.

We contrast this notion and the symmetrizability definition used by Csiszár and Narayan [7], which we call CN-symmetrizability. Their intuition is as follows: James can symmetrize the channel if he can use a memoryless channel $U_{\mathbf{s}|\mathbf{x}}(s|x)$ to generate a state sequence $\underline{s}$ satisfying the cost constraint from a codeword $\underline{x}' \neq \underline{x}$ such that Bob cannot tell the difference between Alice sending $\underline{x}$ and James choosing $\underline{x}'$ or Alice sending $\underline{x}'$ and James choosing $\underline{x}$. .

Definition 3 (CN-symmetrizability [7])**.** *For an oblivious AVC $\mathcal{A} = (\Gamma(g), \Xi(p), W_{\mathbf{y}|\mathbf{x},\mathbf{s}})$, an input distribution $P_{\mathbf{x}} \in \Gamma(g)$ is* Csiszár–Narayan (CN) symmetrizable *if there exists a $U_{\mathbf{s}|\mathbf{x}'} \in \Delta(\mathcal{S}|\mathcal{X})$ with $\left[U_{\mathbf{s}|\mathbf{x}} P_{\mathbf{x}}\right]_{\mathbf{s}} \in \Xi(p)$ such that for all $(x, x', y) \in \mathcal{X} \times \mathcal{X} \times \mathcal{Y}$,*

$$\left[W_{\mathbf{y}|\mathbf{x},\mathbf{s}} U_{\mathbf{s}|\mathbf{x}'}\right]_{\mathbf{y}|\mathbf{x},\mathbf{x}'} = \left[W_{\mathbf{y}|\mathbf{x}',\mathbf{s}} U_{\mathbf{s}|\mathbf{x}}\right]_{\mathbf{y}|\mathbf{x}',\mathbf{x}}. \tag{9}$$

Let $\Delta_{\mathrm{CN}}(\mathcal{A}) \subset \Delta(\mathcal{X})$ be the set of all CN-symmetrizable $P_{\mathbf{x}}$.

Csiszár and Narayan [7] showed that the capacity of the oblivious AVC is given by

$$C \triangleq \max_{P_{\mathbf{x}} \in \Gamma(g) \setminus \Delta_{\mathrm{CN}}(\mathcal{A})} \min_{Q_{\mathbf{s}} \in \Xi(p)} I(\mathbf{x}; \mathbf{y}), \tag{10}$$

where the mutual information is evaluated using $P_{\mathbf{x},\mathbf{s},\mathbf{y}} = P_{\mathbf{x}} Q_{\mathbf{s}} W_{\mathbf{y}|\mathbf{x},\mathbf{s}}$. If the capacity-achieving input distribution is not DS-symmetrizable, then the decoder can use typicality-based decoding rule [8]. However, there are channels which are DS-symmetrizable but not CN-symmetrizable [8, Appendix I].

We target a rate with a similar expression except we maximize over input distributions which are not weakly DS-symmetrizable.

$$R_{\mathrm{DS}} \triangleq \max_{P_{\mathbf{x}} \in \Gamma(g) \setminus \Delta_{\mathrm{DS}}(\mathcal{A})} \min_{Q_{\mathbf{s}} \in \Xi(p)} I(\mathbf{x}; \mathbf{y}), \tag{11}$$

where the mutual information is evaluated using $P_{\mathbf{x},\mathbf{s},\mathbf{y}} = P_{\mathbf{x}} Q_{\mathbf{s}} W_{\mathbf{y}|\mathbf{x},\mathbf{s}}$. Let $P^*_{\mathbf{x}} \in \Gamma(g) \setminus \Delta_{\mathrm{DS}}(\mathcal{A})$ be a maximizer. In some cases, in particular for classically studied error/erasure channels, the two rates in (10) and (11) are the same.

D. Main Result

Theorem 1. *Let $\mathcal{A} = (\Gamma, \Xi, W_{\mathbf{y}|\mathbf{x},\mathbf{s}})$ be an AVC for which $R_{\mathrm{DS}} > 0$. The encoding and decoding processes described in Sections III-A-III-C have encoding/decoding complexity that is $\mathrm{poly}(n)$, and attain any rate below R_{DS} with probability of error at most $1/\mathrm{poly}(n)$.*

III. Computationally efficient codes

We now describe our code construction. Further details and block diagrams can be found in the extended manuscript [20].

A. Codebook

For a blocklength n, let the chunk-length be $\ell \triangleq c_\ell \log(n)$ for some constant $c_\ell > 0$. The number of chunks is $N \triangleq \frac{n}{\ell} = \frac{n}{c_\ell \log(n)}$. Let the rate of the outer and inner codes be $R_{\text{out}} \triangleq 1-\varepsilon_{\text{out}}$ and $R_{\text{in}} \triangleq R_{\text{DS}} - \varepsilon_{\text{in}}$, respectively, where $\varepsilon_{\text{out}}, \varepsilon_{\text{in}} > 0$ are arbitrarily small positive constants. The overall rate is $R = R_{\text{out}} R_{\text{in}}$. Let $P \triangleq n^{c_{\mathcal{P}}}$ for a constant $c_{\mathcal{P}} > 0$ chosen as part of code-design. Our code then comprises of the following three components, known *a priori* to all parties:

1) **Outer code:** Let $q \triangleq 2^{R_{\text{in}}\ell} = 2^{R_{\text{in}} c_\ell \log(n)} = n^{R_{\text{in}} c_\ell}$. The outer code $\mathcal{C}_{\text{out}}$ is an (N, NR_{out}) Reed–Solomon (RS) code over $\mathbb{F}_q$.
2) **Inner code:** For each $i \in [N]$ and $j \in [P]$, the (i,j)-th inner code is $\mathcal{C}_i^j = \left\{\underline{x}_i^j(u) : u \in [2^{\ell R_{\text{in}}}]\right\} \subset \mathcal{X}^\ell$, where each entry $\underline{x}_i^j(u)$ is drawn i.i.d. according to $P_{\mathbf{x}} = P_{\mathbf{x}}^*$.
3) **Set of permutations:** Let $\mathcal{P} = \left\{\pi^1, \cdots, \pi^P\right\} \subset \Sigma_n$ be a set of permutations on $[n]$, sampled i.i.d. from $\mathsf{Unif}(\Sigma_n)$.

B. Encoder

1) **Outer encoder:** Let $\mathbf{m} \sim \mathsf{Unif}([2^{nR}])$. View $\mathbf{m} \in \{0,1\}^{nR}$ as a binary string of length nR and divide nR bits of $\mathbf{m}$ into chunks of size ℓR_{in}. Note that there are $\frac{nR}{\ell R_{\text{in}}} = NR_{\text{out}}$ chunks. View each chunk as a symbol in $\mathbb{F}_q$ (recall $q = 2^{\ell R_{\text{in}}}$). Encode these NR_{out} symbols over $\mathbb{F}_q$ using $\mathcal{C}_{\text{out}} = \mathsf{RS}_q(N, NR_{\text{out}})$. We obtain a vector $\underline{\mathbf{u}} = (\mathbf{u}_1, \cdots, \mathbf{u}_N) \in \mathbb{F}_q^N$.
2) **Inner encoder:** For each $i \in [N]$, view $\mathbf{u}_i \in \mathbb{F}_q$ as a message $\mathbf{u}_i \in [q] = [2^{\ell R_{\text{in}}}]$ for the i-th inner code. Select $\mathbf{j} \sim \mathsf{Unif}([P])$. Encode $\mathbf{u}_i$ to $\underline{x}_i^{\mathbf{j}}(\mathbf{u}_i) \in \mathcal{C}_i^{\mathbf{j}}$. Concatenate these N inner encodings and obtain
$$\underline{x}^{\mathbf{j}}(\underline{\mathbf{u}}) \triangleq \left(\underline{x}_1^{\mathbf{j}}(\mathbf{u}_1), \underline{x}_2^{\mathbf{j}}(\mathbf{u}_2), \cdots, \underline{x}_N^{\mathbf{j}}(\mathbf{u}_N)\right) \in (\mathcal{X}^\ell)^N = \mathcal{X}^n.$$
3) **Permutation:** Finally, permute the above vector using $\pi^{\mathbf{j}} \in \mathcal{P}$ and obtain $\pi^{\mathbf{j}}(\underline{x}^{\mathbf{j}}(\underline{\mathbf{u}})) \in \mathcal{X}^n$.

C. Decoder

Bob receives $\underline{y} \in \mathcal{Y}^n$. For each $j' = 1, 2, \cdots, P$, Bob attempts to decode $\underline{y}$ as follows:

1) **De-permutation:** Compute $\underline{y}^{j'} \triangleq (\pi^{j'})^{-1}(\underline{y}) \in \mathcal{Y}^n$.
2) **Inner decoder:** Write $\underline{y}^{j'}$ as
$$\underline{y}^{j'} = \left(\underline{y}_1^{j'}, \underline{y}_2^{j'}, \cdots, \underline{y}_N^{j'}\right),$$
where $\underline{y}_i^{j'} \in \mathcal{Y}^\ell$ for each $i \in [N]$.
   a) **Validating the guess of j':** For each type $P_{\mathbf{s}} \in \Xi(p)$, $i \in [N]$, and pair $(u_i, \underline{s}_i^{j'})$, define
$$\xi(u_i, \underline{s}_i^{j'}) \triangleq \left\| P_{\mathbf{x}} P_{\mathbf{s}} W_{\mathbf{y}|\mathbf{x},\mathbf{s}} - T_{\underline{x}_i^{j'}(u_i), \underline{s}_i^{j'}, \underline{y}_i^{j'}} \right\|_1$$
   Then construct the set $\mathcal{U}_i^{j'}$ below:
$$\mathcal{U}_i^{j'}(P_{\mathbf{s}}) \triangleq \left\{(u_i, \underline{s}_i^{j'}) \in [2^{\ell R_{\text{in}}}] \times \mathcal{S}^\ell : \xi(u_i, \underline{s}_i^{j'}) \le \delta_{\text{in}}\right\}, \tag{12}$$
   for some constant $\delta_{\text{in}} > 0$.
$$\mathcal{F}^{j'}(P_{\mathbf{s}}) \triangleq \left\{i \in [N] : \mathcal{U}_i^{j'}(P_{\mathbf{s}}) \neq \emptyset\right\}. \tag{13}$$
   If there does not exist a type $P_{\mathbf{s}} \in \Xi(p)$ such that
$$\frac{1}{N}\left|\mathcal{F}^{j'}(P_{\mathbf{s}})\right| \ge 1 - \Delta, \tag{14}$$
   for $\Delta \triangleq 1/\ell^2$, then declare the guess j' as incorrect. Increase the value of j' by 1 and start from step 1. Otherwise, declare that the guess j' is correct.
   b) **Inner decoding:** Let $P_{\mathbf{s}} \in \Xi(p)$ be an arbitrary distribution and $j' \in [P]$ such that (14) holds. Define $\widehat{\underline{\mathbf{u}}} \triangleq (\widehat{\mathbf{u}}_1, \cdots, \widehat{\mathbf{u}}_N) \in ([2^{\ell R_{\text{in}}}] \cup \{\mathsf{e}\})^N$ as follows. For each $i \in [N]$,
      - if $i \in \mathcal{F}^{j'}(P_{\mathbf{s}})$, let $\widehat{\mathbf{u}}_i$ (together with some $\underline{s}_i^{j'}$) be an arbitrary element from $\mathcal{U}_i^{j'}$.
      - otherwise, $\widehat{\mathbf{u}}_i = \mathsf{e}$, where e denotes an erasure.
3) **Outer decoder:** Decode $\widehat{\underline{\mathbf{u}}}$ to $\widehat{m} \in [2^{nR}]$ using the standard error-erasure Reed–Solomon decoder for $\mathcal{C}_{\text{out}}$.

D. Intuition behind our scheme

To get a sense of the role played by different parts of our encoder/decoder, consider the well-studied bit-flipping adversary with binary inputs/outputs and James can flip pn bits. As in classic concatenated code designs, to avoid the exponential decoding complexity of random codes, our encoder uses an inner code with blocklength $\ell = \Theta(\log(n))$ along with a Reed-Solomon outer code. The inner codes are random codes with codebooks of size $n^{c_\ell R_{\text{in}}}$ which can be encoded by a lookup table, so encoding $\Theta(\frac{n}{\log(n)})$ chunks can also be done in time $n^{c_\ell R_{\text{in}}+1}$. The outer code's Reed-Solomon Encoder Step 1 can be done in quasilinear time. Finally, Alice's permutation can also be applied in linear time. Hence the encoding process takes time $\mathcal{O}(n^{c_\ell R_{\text{in}}+1})$.

Without Alice's encoding permutation in Encoder step 3, James could allocate more than $p\ell$ bit flips to certain chunks, which poses a challenge for decoding. If Bob knew which permutation Alice used, then he could de-permute the received codeword. We show that regardless of James's choice of $\underline{s}$, after de-permuting, with high probability the type of $\underline{s}$ in most chunks will be approximately the same. Bob can decode those chunks using typicality decoding [5], and the outer code can compensate for the errors in the remaining chunks.

The set of permutations $\mathcal{P}$ that Alice can use is known by all parties and we choose $P = |\mathcal{P}| = n^{c_{\mathcal{P}}}$. Because Alice and Bob do not share common randomness, Bob must deduce which permutation Alice used based on the output of the channel. He does this by exhaustively trying every permutation $\pi \in \mathcal{P}$ and every type in $P_{\mathbf{s}} \in \Xi(p)$. He applies π^{-1} to de-permute in Decoder Step 1 and for each type $P_{\mathbf{s}}$ uses typicality-based

list-decoding (equation (12)) on each chunk in Step 2a. If every type $P_{\mathbf{s}}$ returns more than a Δ-fraction of chunks with empty lists (equation (14)), then Bob moves on to the next permutation. Since the number of permutations is $n^{c_{\mathcal{P}}}$, the number of chunks is $\mathcal{O}(n/\log(n))$, the number of types in a chunk is $\mathcal{O}(\log(n))$, each chunk's inner codebook has size $n^{c_\ell R_{\text{in}}}$, and the number of state sequences in a given chunk is $\mathcal{O}(n^{c_\ell})$, the overall decoding complexity is $n^{c_{\mathcal{P}}+1+c_\ell(R_{\text{in}}+1)}$.

To understand the decoder better, consider the case where Alice uses a permutation π^* and Bob guesses an incorrect permutation $\pi \neq \pi^*$. If Bob applies π^{-1} to $\underline{y}$, the chunks of $\pi^{-1}(\underline{y})$ will contain different symbols from the original chunks encoded by Alice. The success of our decoder relies, in part, on the chunk-wise decoding of the inner codes failing to return *any* message at all for a significant fraction of chunks. This failure will signal to Bob that π was not correct. This style of decoder is different than the standard decoding for AVCs [7], where the correct transmitted codeword is the "legitimate" codeword, which is expected to "defeat" alternative codewords in the decoder. Since we cannot rely on the "correct" codeword "defeating" alternate candidates, we use only the first stage, namely a simpler "typicality" decoder. While a two-stage decoder [7] works for any CN-nonsymmetrizable channel, it is known that a typicality decoder works [8] for a subclass of channels, namely DS-nonsymmetrizable channels. In order to show that Bob will reject incorrect permutations during decoding, our arguments require the input distribution to satisfy the strong DS-nonsymmetrizability condition in Definition 2.

IV. Proof sketch

Full proofs can be found in the extended version [20]. There are two types of errors in our scheme. The first type of error occurs when Bob guesses and attempts to decode w.r.t. the correct permutation but the decoding fails. We show that for any state vector $\underline{s}$ James chooses, with high probability over the set $\mathcal{P}$ of permutations chosen , "most" permutations in $\mathcal{P}$ "spread" $\underline{s}$ "quasi-uniformly" over most chunks, and therefore with high probability over the chosen permutation, design of inner codes and transmitted message, Bob decodes the correct message. More precisely:

- With probability at least $1-2^{-\Omega(n^{c_{\mathcal{P}}-1.5})}$ over the random choice of permutations $\mathcal{P}$, for any $\underline{s}$ there exists a $1-n^{-c_{\mathcal{P}}/2}$ fraction of permutations in $\mathcal{P}$, such that at most $1/\ell^3$ fraction of the N chunks contain $\underline{s}$ sub-vectors (of length-ℓ) that are atypical w.r.t. $T_{\underline{s}}$, the type of the overall $\underline{s}$ vector.
- Further, for any chunk where the $\underline{s}$ sub-vector is typical, with probability $n^{-\Omega(1)}$ over random inner code design and choice of transmitted inner codeword, the corresponding $(\underline{x},\underline{s})$ sub-vectors are jointly atypical w.r.t. $P^*_{\mathbf{x}}T_{\underline{s}}$. Then, with probability at least $1-2^{-\Omega(n/\ell^2)}$ over the design of the inner codes (the inner codes are designed i.i.d. across chunks), at most another $1/\ell^3$ fraction of chunks are such that the corresponding $(\underline{x},\underline{s})$ sub-vectors are jointly atypical w.r.t. $P^*_{\mathbf{x}}T_{\underline{s}}$.
- Finally, by standard concentration arguments/Csiszár-Narayan-type analysis [7], [8], the probability of there being more than yet another $1/\ell^3$ fraction of chunks such that the corresponding $(\underline{x},\underline{s},\underline{y})$ sub-vectors are jointly atypical w.r.t. $P^*_{\mathbf{x}}T_{\underline{s}}W_{\mathbf{y}|\mathbf{x},\mathbf{s}}$, *or* there is more than one subcodeword that is jointly typical w.r.t. the corresponding sub-vector $\underline{y}$, is $2^{-\Omega(n/\ell^2)}$.

A slight technical challenge in proving the concentration inequalities above results from the fact that Alice's permutation makes the error events across chunks dependent. However, they are negatively correlated [21], and hence tail bounds assuming independence serve as upper bounds for the error events we are interested in [22]. Overall, conditioned on Bob guessing the correct permutation, the probability of one of the error-events above is $\mathcal{O}(n^{-c_{\mathcal{P}}/2})$ – conditioning on their complement, setting $\varepsilon_{\text{out}} = 6/\ell^3$ enables Bob's Reed-Solomon outer code to decode the transmitted message correctly.

The second type of error comes from Bob failing to reject an incorrect permutation j'. We show that with high probability Bob's decoder detects this. More precisely:

- Since the AVC is strongly DS-nonsymmetrizable, for any state vector $\underline{s}$ the jammer chooses, there exists a $\delta > 0$ and a constant (depending on δ) fraction of the N chunks such that such that $P_{\mathbf{x}}$ is DS-nonsymmetrizable.
- With probability at least $1-n^{-\Omega(\log(n))}$ over the random design of $\mathcal{P}$ and inner codes, for any pair of distributions $j \neq j'$, the expected distribution (over Alice's transmitted message) of Alice's transmitted symbols in any set of N/ℓ^2 chunks of $(\pi^{j'})^{-1}(\underline{y})$ are $\left(1 \pm \frac{1}{\text{poly}(n)}\right)(P^*_{\mathbf{x}})^{\otimes N/\ell}$. That is, despite being drawn from inner codebooks, for any set of sufficiently few (at most N/ℓ^2) chunks the channel inputs are polynomially close to i.i.d. draws from $P^*_{\mathbf{x}}$.
- Using a modification of arguments in [7], for $j' \neq j$ the probability that Bob's decoder outputs non-empty lists in more than $N(1/\ell^2)$ chunks is $2^{-\Omega(N/\ell)}$, and hence by a union bound with high probability each incorrect permutation will be detected.

V. Conclusion

We described a code construction for AVCs with oblivious adversaries which has polynomial-time encoding and decoding complexity. Our achievable rates match the capacity-achieving performance of existing computationally-efficient constructions for additive-error/erasure channels. For general AVCs, we can achieve positive rates for AVCs using a DS-nonsymmetrizable $P_{\mathbf{x}}$. The reason our codes may not achieve oblivious capacity for general AVCs is since they are restricted to a potentially smaller subset of input distributions than in (10). We leave open the question of computationally tractable schemes attaining the capacity for *all* AVCs.

Acknowledgements

The work of M. Langberg and A. D. Sarwate was supported in part by the US NSF under awards CCF-1909451 and CCF-1909468. B. K. Dey was supported in part by the Bharti Centre for Communication in IIT Bombay.

REFERENCES

[1] R. Ahlswede, "Coloring Hypergraphs : A New Approach fo Multi-user Source Coding – II," *Journal of Combinatorics, Information and System Sciences*, vol. 5, no. 3, pp. 220–268, 1980.

[2] D. Blackwell, L. Breiman, and A. J. Thomasian, "The Capacity of a Class of Channels under Random Coding," *Annals of Mathematical Statistics*, vol. 31, no. 3, pp. 558–567, 1960.

[3] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2148–2177, 1998.

[4] R. Ahlswede, "Elimination of correlation in random codes for arbitrarily varying channels," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 44, pp. 181–193, 1978.

[5] R. L. Dobrušin and S. Z. Stambler, "Coding theorems for classes of discrete memoryless arbitrarily time-variable channels," *Problemy Peredači Informacii*, vol. 11, no. 2, pp. 3–22, 1975.

[6] T. Ericson, "Exponential error bounds for random codes in the arbitrarily varying channel," *IEEE Transactions on Information Theory*, vol. 31, no. 1, pp. 42–48, 1 1985. [Online]. Available: https://dx.doi.org/10.1109/TIT.1985.1056995

[7] I. Csiszár and P. Narayan, "The capacity of the arbitrarily varying channel revisited: positivity, constraints," *IEEE Trans. Inform. Theory*, vol. 34, no. 2, pp. 181–193, 1988. [Online]. Available: https://doi.org/10.1109/18.2627

[8] I. Csiszár and P. Narayan, "Capacity and decoding rules for classes of arbitrarily varying channels," *IEEE Transactions on Information Theory*, vol. 35, no. 4, pp. 752–769, 7 1989. [Online]. Available: https://doi.org/10.1109/18.32153

[9] R. J. Lipton, "A new approach to information theory," in *STACS 94*, P. Enjalbert, E. W. Mayr, and K. W. Wagner, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1994, pp. 699–708.

[10] V. Guruswami and A. Smith, "Optimal rate code constructions for computationally simple channels," *J. ACM*, vol. 63, no. 4, sep 2016. [Online]. Available: https://doi.org/10.1145/2936015

[11] R. Shaltiel and J. Silbak, "Explicit List-Decodable Codes with Optimal Rate for Computationally Bounded Channels," in *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2016)*, ser. Leibniz International Proceedings in Informatics (LIPIcs), K. Jansen, C. Mathieu, J. D. P. Rolim, and C. Umans, Eds., vol. 60. Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2016, pp. 45:1–45:38. [Online]. Available: https://doi.org/10.4230/LIPIcs.APPROX-RANDOM.2016.45

[12] J. Silbak, S. Kopparty, and R. Shaltiel, "Quasilinear time list-decodable codes for space bounded channels," in *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, 2019, pp. 302–333. [Online]. Available: https://doi.org/10.1109/FOCS.2019.00028

[13] R. Shaltiel and J. Silbak, "Explicit uniquely decodable codes for space bounded channels that achieve list-decoding capacity," in *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, ser. STOC 2021. New York, NY, USA: Association for Computing Machinery, 2021, p. 1516–1526. [Online]. Available: https://doi.org/10.1145/3406325.3451048

[14] ——, "Explicit list-decodable codes with optimal rate for computationally bounded channels," *computational complexity*, vol. 30, no. 3, 2021. [Online]. Available: https://doi.org/10.1007/s00037-020-00203-w

[15] ——, "Error correcting codes that achieve BSC capacity against channels that are poly-size circuits," in *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, 2022, pp. 13–23. [Online]. Available: https://doi.org/10.1109/FOCS54457.2022.00009

[16] K. Yasunaga, "Error-correcting codes against chosen-codeword attacks," in *Information Theoretic Security*, A. C. Nascimento and P. Barreto, Eds. Cham: Springer International Publishing, 2016, pp. 177–189.

[17] ——, "Error Correction by Structural Simplicity: Correcting Samplable Additive Errors," *The Computer Journal*, vol. 62, no. 9, pp. 1265–1276, 10 2018. [Online]. Available: https://doi.org/10.1093/comjnl/bxy100

[18] E. Ruzomberka, C.-C. Wang, and D. J. Love, "Channel capacity for adversaries with computationally bounded observations," *IEEE Transactions on Information Theory*, vol. 70, no. 1, pp. 75–92, 2024. [Online]. Available: https://doi.org/10.1109/TIT.2023.3330811

[19] S. Li, P. Krishnan, S. Jaggi, M. Langberg, and A. D. Sarwate, "Computationally efficient codes for adversarial binary-erasure channels," in *2023 IEEE International Symposium on Information Theory (ISIT)*, 2023, pp. 228–233. [Online]. Available: https://doi.org/10.1109/ISIT54713.2023.10206731

[20] B. Kumar Dey, S. Jaggi, M. Langberg, A. D. Sarwate, and Y. Zhang, "Computationally efficient codes for strongly Dobrushin-Stambler non-symmetrizable oblivious AVCs," 2024, in preparation.

[21] K. Joag-Dev and F. Proschan, "Negative association of random variables, with applications," *Ann. Statist.*, vol. 11, no. 1, pp. 286–295, 1983. [Online]. Available: https://doi.org/10.1214/aos/1176346079

[22] D. Dubhashi and D. Ranjan, "Balls and bins: a study in negative dependence," *Random Structures & Algorithms*, vol. 13, no. 2, pp. 99–124, 1998. [Online]. Available: https://doi.org/10.1002/(SICI)1098-2418(199809)13:2<99::AID-RSA1>3.0.CO;2-M