Non-coherent Network Coding: An Arbitrarily Varying Channel Approach

Mahdi Jafari Siavoshani, Shenghao Yang, Raymond Yeung

Abstract

In this paper, we propose an "arbitrarily varying channel" (AVC) approach to study the capacity of non-coherent transmission in a network that employs randomized linear network coding. The network operation is modeled by a matrix channel over a finite field where the transfer matrix changes arbitrarily from time-slot to time-slot but up to a known distribution over its rank. By extending the AVC results to this setup, we characterize the capacity of such a non-coherent transmission scheme and show that subspace coding is optimal for achieving the capacity.

By imposing a probability distribution over the state space of an AVC, we obtain a channel which we called "partially arbitrarily varying channel" (PAVC). In this work, we characterize the "randomized", "stochastic" as well as the "deterministic" code capacity of a PAVC under the average error probability criterion. Although we introduce the PAVC to model the non-coherent network coding, this extension to an AVC might be of its own interest as well.

I. Introduction

Randomized linear network coding [1] is an efficient and practical approach to implement network coding [2], [3] in large dynamically changing networks because it does not require a priori the knowledge of the network topology. However, in order to enable the receivers to decode, to each packet a coding vectors is appended to learn the channel while the packet passes through the network. So in other words, use of coding vectors is akin to use of training symbols to learn the transformation induced by a network.

A different approach, than using coding vectors, is to assume a non-coherent scenario for communication, as proposed in [4], where neither the source(s) nor the receiver(s) have any knowledge of the network topology or the network nodes operations. Non-coherent communication allows creation of end-to-end systems that are completely oblivious to the network state. In that work, the authors propose communication via choosing subspaces and they introduce a subspace channel called "operator channel" (a channel which has subspaces as input and output symbols). Then, they focused on algebraic subspace code constructions over a Grassmannian for the operator channel.

Following [4], different probabilistic models have been proposed so far to model the non-coherent randomized linear network coding channel where these models enable us to define and characterize the capacity for such a channel. In all of these works, when there is no error in the network, the non-coherent linear network coding channel is modeled by a multiplicative matrix channel.

Montanari *et al.* [5] introduced a probabilistic model to capture the end-to-end functionality of non-coherent network coding operation, with a focus on the case of error correction capabilities. Their model does not examine

coding schemes defined over multiple blocks, but instead, allows the packets length to increases to infinity, with the result that in the large packet length limit the scheme essentially becomes coherent.

Jafari *et al.* [6], [7], [8] modeled the non-coherent network coding channel by assuming that the transfer matrix has i.i.d. entries selected uniformly at random in every time-slot. They showed that coding over subspace is sufficient to achieve the capacity for all range of channel parameters. Then, they obtained the channel capacity as a solution of a convex optimization problem over $O(\min[M, N])$ variables. Moreover, when the field size is greater than a threshold, they characterize the capacity by solving the optimization problem.

Silva *et al.* [9] derived the capacity of the multiplicative finite field matrix channel under the assumption that the transfer matrix is square and chosen uniformly at random among all full-rank matrices. Similarly, in this model they obtained that coding over subspaces is sufficient to achieve the capacity.

Yang *et al.* [12], [13] (see also [10], [11]) considered a completely general scenario, making no assumption on the distribution of the transfer matrix. They obtained upper and lower bounds on the channel capacity, and give a sufficient condition on the distribution of the transfer matrix such that coding over subspaces is capacity achieving. They also studied the achievable rates of coding over subspaces.

Nobrega *et al.* [14] considered the case where the probability distribution of the rank of the transfer matrix is arbitrary; however all matrices with the same rank are equiprobable. Then, they followed a similar approach to [8] to write the capacity as a solution of a convex optimization problem over $O(\min[M, N])$ variables. They also observed that by using subspace codes we do not loose anything in terms of rate optimality and finally they provided some upper and lower bounds for the capacity.

In most of the previous works, only certain probability models for the channel transfer matrix have been discussed. However, in practice a complete probabilistic characterization of the matrix channel is difficult and the network may not follow a given probability model. Instead of assuming a complete probability model, we consider in this paper that only a partial knowledge about the probabilistic model of the channel is known.

More precisely, we assume that the rank distribution of the transfer matrix is known a priori, but the distribution of matrices among each rank is unknown and arbitrary. Though very similar to the arbitrarily varying channel (AVC) model introduced in [15] (refer to [16] and the references therein), but this non-coherent network coding model is not exactly an AVC. We introduce a "partially arbitrary varying channel" (PAVC) to capture the statistical property of this non-coherent network coding model.

By extending results for the AVC, we obtain the capacities of the PAVC for deterministic, stochastic, and randomized codes (Theorem 1, Theorem 2, and Theorem 3). We further show that the randomized and the deterministic code capacities of the non-coherent network coding model are the same (Theorem 4), and that subspace coding is sufficient to achieve the capacity (Corollary 3). This AVC approach to the non-coherent network coding provides a justification for the optimality of subspace coding in a more general setting.

The paper is organized as follows. In §II, we describe the non-coherent network coding model and introduce the PAVC. In §III, we state the main results of the paper; the capacity of a PAVC and as a corollary we will state the capacity of the non-coherent network coding under having constraint over the rank distribution. The proofs of the

PAVC capacity results are stated in Appendix A, Appendix B and Appendix C.

II. Problem Setup and Notation

A. Notation

Let $\mathsf{Uni}(\mathcal{M})$ denote the uniform distribution over the set $\mathcal{M}$. For example, we use $\mathsf{Uni}(\mathbb{F}_q^\ell)$ to denote the uniform distribution over vectors of length ℓ that are defined over finite field $\mathbb{F}_q$. For $m \times n$ matrices over $\mathbb{F}_q$, we use $\mathsf{Uni}(\mathbb{F}_q^{m\times n}, r)$ to denote the uniform distribution over $m \times n$ matrices with rank r.

We use bold letters to denote vectors and matrices. For the convenience of notation, we use $[i : j]$ to denote the set $\{i, i+1, \ldots, j-1, j\}$ where $i, j \in \mathbb{Z}$.

B. Non-coherent Network Coding Channel Model

Consider a unicast communication over a network where the relay nodes perform random linear network coding over a finite field $\mathbb{F}_q$. Suppose that time is slotted and the channel is block time-varying. At every time-slot, the source injects M packets $\boldsymbol{X}_1[t], \ldots, \boldsymbol{X}_M[t]$ of length T symbols from $\mathbb{F}_q$ into the network, *i.e.,* $\boldsymbol{X}_i[t] \in \mathbb{F}_q^T$. The receiver collects N packets $\boldsymbol{Y}_1[t], \ldots, \boldsymbol{Y}_N[t]$ and aims to decode the transmitted packets.

We use matrices $\boldsymbol{X}[t]$ and $\boldsymbol{Y}[t]$ to denote respectively, the transmitted and received packets, *i.e.,* the ith column of these matrices represent the ith transmitted and received packets, respectively. For a unicast communication, at time-slot (block) t, the receiver observes

$$\boldsymbol{Y}[t] = \boldsymbol{X}[t]\boldsymbol{H}[t], \tag{1}$$

where $\boldsymbol{X}[t] \in \mathbb{F}_q^{T\times M}$, $\boldsymbol{Y}[t] \in \mathbb{F}_q^{T\times N}$, and $\boldsymbol{H}[t] \in \mathbb{F}_q^{M\times N}$. We assume that the channel transfer matrix $\boldsymbol{H}[t]$ is unknown to both the transmitter and the receiver and it changes arbitrarily from one block to another block with a constraint on its rank. More precisely, the ranks of $\boldsymbol{H}[t]$, $t = 1, 2, \ldots$, are independent and follow the same distribution of a random variable R. The conditional distribution of $\boldsymbol{H}[t]$ given $\mathrm{rk}\left(\boldsymbol{H}[t]\right)$ is unknown and changes arbitrarily for different t. However, we assume that the distribution of the random variable R is known. We may consider the channel transfer matrix as the channel state. For given $\boldsymbol{h}[1 : n]$ the channel transition probability is

$$W_{\mathsf{m}}^n\left(\boldsymbol{y}[1:n]|\boldsymbol{x}[1:n];\boldsymbol{h}[1:n]\right) = \prod_{t=1}^{n} W_{\mathsf{m}}\left(\boldsymbol{y}[t]|\boldsymbol{x}[t];\boldsymbol{h}[t]\right),$$

where $W_{\mathsf{m}}(\boldsymbol{y}|\boldsymbol{x};\boldsymbol{h}) \triangleq \mathbb{1}_{\{\boldsymbol{y}=\boldsymbol{x}\boldsymbol{h}\}}$ is a stochastic matrix.

The above model is very similar to an arbitrarily varying channel (AVC) model (refer to [16] for more information about AVC) but it does not completely fit into that model. In this work, we will show that it is indeed possible to extend the AVC concepts and results for the above channel model and characterize its capacity.

C. Partially Arbitrarily Varying Channel (PAVC)

Before defining a partially arbitrarily varying channel (PAVC), let us first consider an AVC model. Let $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ denote the input and output symbol of a channel where $\mathcal{X}$ and $\mathcal{Y}$ are finite sets denoting the channel

input and output alphabets, respectively. Let us consider a transmission scenario where the channel parameters vary arbitrarily from symbol to symbol during the course of a transmission. More precisely, for the channel transition matrix, we can write

$$W^n(\boldsymbol{y}|\boldsymbol{x};\boldsymbol{s}) \triangleq \prod_{t=1}^{n} W(y_t|x_t;s_t), \tag{2}$$

where $\boldsymbol{s}=(s_1,\ldots,s_n)$, $s_i \in \mathcal{S}$, and $W : \mathcal{X}\times\mathcal{S}\to\mathcal{Y}$ is a given stochastic matrix. $\mathcal{S}$ is a finite set, often referred to as the state space. This model, called a "discrete memoryless arbitrarily varying channel," will be referred to as an AVC.

Now, we define a PAVC as an AVC with a probability constraint over the state space $\mathcal{S}$. Define a function $\mathfrak{q} : \mathcal{S}\to\mathcal{Q}$ where $\mathcal{Q}\triangleq\{0,\ldots,m\}$ and define a random variable Q with alphabet $\mathcal{Q}$ whose distribution is known by the encoder and the decoder. For a PAVC, we have $\mathfrak{q}(S_t)$, $t=1,2\ldots$, are independent and follow the same distribution of Q. In other words,

$$P_{\mathfrak{q}(\boldsymbol{S})}(q_1,\ldots,q_n) = \prod_{t=1}^{n} P_Q(q_t), \tag{3}$$

where $\mathfrak{q}(\boldsymbol{S})\triangleq(\mathfrak{q}(S_1),\ldots,\mathfrak{q}(S_n))$. We call this model a "discrete memoryless partially arbitrarily varying channel," and will refer to it as a PAVC.

In this work, we are interested in characterizing the capacity of a PAVC. However, we first have to define the capacity. As there are different notions of capacity for an AVC based on different error criteria, the same is true for a PAVC (for more information refer to [16]).

Suppose that the message set of a code is identified as the set $\mathcal{M}=\{1,\ldots,K\}$, so that a length-$n$ block code is given by a pair of mapping (ψ,ϕ), where $\psi:\ \mathcal{M}\mapsto\mathcal{X}^n$ is the encoder, and $\phi:\ \mathcal{Y}^n\mapsto\mathcal{M}\cup\{0\}$ is the decoder, where the output 0 counts for an error. Let us define

$$e(i,\boldsymbol{s},\psi,\phi) \triangleq \sum_{\boldsymbol{y}:\ \phi(\boldsymbol{y})\neq i} W^n(\boldsymbol{y}|\psi(i);\boldsymbol{s}). \tag{4}$$

Then, the error probability for message i, when this code is used on a PAVC and when the state sequence is given to be $\boldsymbol{s}\in\mathcal{S}^n$, equals

$$e_d(i,\boldsymbol{s}) \triangleq e(i,\boldsymbol{s},\psi,\phi), \tag{5}$$

and the average probability of error for a state sequence $\boldsymbol{s}$ is

$$\bar{e}_d(\boldsymbol{s}) \triangleq \frac{1}{K}\sum_{i=1}^{K} e_d(i,\boldsymbol{s}). \tag{6}$$

Definition 1. *A number $\mathfrak{R}>0$ is called an achievable rate for the given PAVC (for deterministic code and average error probability criterion) if for every $\epsilon>0$, $\delta>0$, and sufficiently large n, there exists length-n block code (ψ,ϕ) with*

$$\frac{1}{n}\log K > \mathfrak{R}-\delta, \tag{7}$$

and

$$\max_{P_{\boldsymbol{S}|\,\mathfrak{q}(\boldsymbol{S})}} \mathbb{E}\left[\bar{e}_d(\boldsymbol{S})\right] \triangleq \max_{P_{\boldsymbol{S}|\,\mathfrak{q}(\boldsymbol{S})}} \sum_{\boldsymbol{s}} \bar{e}_d(\boldsymbol{s}) P_{\boldsymbol{S}|\mathfrak{q}(\boldsymbol{S})}\left(\boldsymbol{s}|\mathfrak{q}(\boldsymbol{s})\right) P_{Q^n}\left(\mathfrak{q}(\boldsymbol{s})\right) \leq \epsilon, \tag{8}$$

where $P_{Q^n}(\boldsymbol{q}) \triangleq \prod_{t=1}^{n} P_Q(q_t)$. *The maximum achievable rate is called the capacity of the PAVC and is denoted by* $C_{\mathsf{pavc}}^{\mathsf{d},\mathsf{a}}$ *(where superscript "*a*" denotes for the average error probability criterion given by* (6) *and "*d*" denotes for the determinist code).*

Remark: Note that if there is no probability constraint on the state space in Definition 1 ($P_{\boldsymbol{S}}$ is unknown instead of $P_{\boldsymbol{S}|\mathsf{q}(\boldsymbol{S})}$), then by replacing the maximization over $P_{\boldsymbol{S}|\mathsf{q}(\boldsymbol{S})}$ with $P_{\boldsymbol{S}}$, we recover the average error criterion for an AVC, namely, $\max_{P_{\boldsymbol{S}}} \mathbb{E}\left[\bar{e}_d(\boldsymbol{S})\right] \leq \epsilon$ is equivalent to $\max_{\boldsymbol{s}} \bar{e}_d(\boldsymbol{s}) \leq \epsilon$.

In contrast to using deterministic codes, there exists another communication technique called *randomized coding* which can provide improvement in performance if a common source of randomness is available between the source and the destination.

Precisely, a randomized code (Ψ, Φ) is a random variable with values in the family of all length-n block codes (ψ, ϕ), defined earlier in this section, with the same message set $\mathcal{M}$. Then, the error probability for message i, when this code is used on a PAVC and when the state sequence is given to be $\boldsymbol{s} \in \mathcal{S}^n$, equals

$$e_r(i, \boldsymbol{s}) \triangleq \mathbb{E}_{\Psi,\Phi}\left[e(i, \boldsymbol{s}, \Psi, \Phi)\right], \tag{9}$$

and the average probability of error for a state sequence $\boldsymbol{s}$ is

$$\bar{e}_r(\boldsymbol{s}) \triangleq \frac{1}{K} \sum_{i=1}^{K} e_r(i, \boldsymbol{s}). \tag{10}$$

Similar to Definition 1, we define the capacity $C_{\mathsf{pavc}}^{\mathsf{r},\mathsf{a}}$ by replacing the function $\bar{e}_d(\boldsymbol{s})$ with $\bar{e}_r(\boldsymbol{s})$. Here, the superscript "r,a" denotes for *randomized codes* and *average error probability*.

Yet there is another communication scheme called *coding with stochastic encoder* which only allows randomization in the transmitter, *i.e.*, there is no shared randomness between the encoder and the decoder. More precisely, a code with stochastic encoder (Ψ, ϕ) is a random variable with values in the family of all length-n block codes (ψ, ϕ) with the same message set $\mathcal{M}$.

The error probability for message i, when this code is used on a PAVC and when the state sequence is given to be $\boldsymbol{s} \in \mathcal{S}^n$, equals

$$e_t(i, \boldsymbol{s}) \triangleq \mathbb{E}_{\Psi}\left[e(i, \boldsymbol{s}, \Psi, \phi)\right], \tag{11}$$

and the average probability of error for a state sequence $\boldsymbol{s}$ is

$$\bar{e}_t(\boldsymbol{s}) \triangleq \frac{1}{K} \sum_{i=1}^{K} e_t(i, \boldsymbol{s}). \tag{12}$$

Similar to Definition 1, we define the capacity $C_{\mathsf{pavc}}^{\mathsf{t},\mathsf{a}}$ by replacing the function $\bar{e}_d(\boldsymbol{s})$ with $\bar{e}_t(\boldsymbol{s})$. Here, the superscript "t,a" denotes for *codes with stochastic encoder* and *average error probability*.

III. Main Results

Our main goal is to characterize the capacity of the non-coherent network coding channel described in §II-B. Toward this end, we first determine the capacity of a general PAVC.

A. Capacity of a PAVC

Before stating the deterministic code capacity of a PAVC, we need the following definition.

Definition 2. *A PAVC is called* symmetrizable *if for some channel* $U: \mathcal{X} \times \mathcal{Q} \mapsto \mathcal{S}$*, and for every* x*,* x'*, and* y *we have*

$$\sum_s W(y|x;s)U\left(s|x',\mathfrak{q}(s)\right)P_Q\left(\mathfrak{q}(s)\right) = \sum_s W(y|x';s)U\left(s|x,\mathfrak{q}(s)\right)P_Q\left(\mathfrak{q}(s)\right). \tag{13}$$

Let $\mathcal{U}(\mathcal{X}\times\mathcal{Q}\to\mathcal{S})$ *be the set of all such channel. If* $\mathcal{U}(\mathcal{X}\times\mathcal{Q}\to\mathcal{S})=\emptyset$ *then the PAVC is called* non-symmetrizable.

Then, the following theorem characterizes the capacity of a PAVC for *deterministic* codes and average error criterion.

Theorem 1. *For the* deterministic *code capacity* $C_{\mathsf{pavc}}^{\mathsf{d},\mathsf{a}}$ *we have* $C_{\mathsf{pavc}}^{\mathsf{d},\mathsf{a}} > 0$ *if and only if the PAVC is non-symmetrizable. If* $C_{\mathsf{pavc}}^{\mathsf{d},\mathsf{a}} > 0$*, then we have*

$$C_{\mathsf{pavc}}^{\mathsf{d},\mathsf{a}} = \max_{P_X}\min_{P_{S|\mathfrak{q}(S)}} I(P_X,\bar{W}_S) = \min_{P_{S|\mathfrak{q}(S)}}\max_{P_X} I(P_X,\bar{W}_S), \tag{14}$$

where

$$\bar{W}_S(y|x) \triangleq \mathbb{E}\left[W(y|x;S)\right] = \sum_s W(y|x;s)P_{S|\mathfrak{q}(S)}\left(s|\mathfrak{q}(s)\right)P_Q\left(\mathfrak{q}(s)\right), \tag{15}$$

and $I(P_X,\bar{W}_S) \triangleq I(X;Y)$ *such that* Y *is connected to* X *through the channel* $\bar{W}_S$*.*

Proof: For the proof refer to Appendix A. ■

Theorem 2. *For a PAVC, the capacity of codes with stochastic encoder is equal to the deterministic code capacity,* i.e., $C_{\mathsf{pavc}}^{\mathsf{t},\mathsf{a}} = C_{\mathsf{pavc}}^{\mathsf{d},\mathsf{a}}$.

Proof: For the proof refer to Appendix B. ■

Remark: Theorem 2 shows that randomization at the encoder does not improve the deterministic code capacity of a PAVC.

The following theorem characterizes the capacity of a PAVC for *randomized* code.

Theorem 3. *The* randomized *code capacity of a PAVC is given by*

$$C_{\mathsf{pavc}}^{\mathsf{r},\mathsf{a}} = \max_{P_X}\min_{P_{S|\mathfrak{q}(S)}} I(P_X,\bar{W}_S) = \min_{P_{S|\mathfrak{q}(S)}}\max_{P_X} I(P_X,\bar{W}_S), \tag{16}$$

where $\bar{W}_S$ *is defined in* (15)*.*

Proof: For the proof refer to Appendix C. ■

Remark: Same as an AVC, the randomized code capacity of a PAVC for the maximum and the average error probability criteria are the same.

Remark: In a more general scenario, when $\mathfrak{q}(S_t)$, $t = 1, 2, \ldots$ are not i.i.d. but still for every time t the marginal probability $\mathbb{P}\left[\mathfrak{q}(S_t) = i\right] = P_Q(i)$, the adversary who controls the channel state has more power and hence the capacity in this case is less than or equal to the capacity of i.i.d. case.

B. Capacity of Non-coherent Network Coding

According to the definition of the PAVC in §II-C, the non-coherent network coding model defined by (1) is a PAVC for which the deterministic and stochastic code capacities are equal, as stated in Theorem 1 and Theorem 2, and can be characterized as follows.

Corollary 1. *The deterministic and stochastic code capacities of the channel* (1) *are equal. They are non-zero and given by*

$$C = \max_{P_{\boldsymbol{X}}} \min_{P_{\boldsymbol{H}|\,\mathrm{rk}(\boldsymbol{H})}} I(\boldsymbol{X};\boldsymbol{Y}) = \min_{P_{\boldsymbol{H}|\,\mathrm{rk}(\boldsymbol{H})}} \max_{P_{\boldsymbol{X}}} I(\boldsymbol{X};\boldsymbol{Y}), \tag{17}$$

if and only if the channel is non-symmetrizable, i.e.*, if there is no stochastic matrix* $U : \mathcal{X} \times [0 : \min[M, N]] \mapsto \mathcal{H}$ *such that we have*

$$\sum_{r=0}^{\min[M,N]} \sum_{\boldsymbol{h}:\ \mathrm{rk}(\boldsymbol{h})=r} W_{\mathsf{m}}(\boldsymbol{y}|\boldsymbol{x};\boldsymbol{h})U(\boldsymbol{h}|\boldsymbol{x}',r)P_R(r) = \sum_{r=0}^{\min[M,N]} \sum_{\boldsymbol{h}:\ \mathrm{rk}(\boldsymbol{h})=r} W_{\mathsf{m}}(\boldsymbol{y}|\boldsymbol{x}';\boldsymbol{h})U(\boldsymbol{h}|\boldsymbol{x},r)P_R(r),$$

for all $\boldsymbol{x} \in \mathbb{F}_q^{T\times M}$*,* $\boldsymbol{x}' \in \mathbb{F}_q^{T\times M}$*, and* $\boldsymbol{y} \in \mathbb{F}_q^{T\times N}$*.*

Similarly, using Theorem 3, the randomized code capacity of the non-coherent network coding defined by (1) is stated in the following corollary.

Corollary 2. *The randomized code capacity of the channel defined by* (1) *is given by* (17)*.*

It is hard to show directly that the channel defined by (1) is non-symmetrizable. Instead, we prove this indirectly in the next lemma by showing the existence of a (stochastic) coding scheme that gives a non-zero transmission rate over the channel.

Lemma 1. *If* $\mathbb{E}[R] > 0$*, the channel defined by* (1) *is non-symmetrizable, and so by Corollary 1, its capacity is non-zero and is given by* (17)*. If* $\mathbb{E}[R] = 0$*, then the capacity is zero.*

Proof: The case for $\mathbb{E}[R] = 0$ follows because $\boldsymbol{H}[t]$ is the zero matrix with probability one. To show the non-symmetrizability of the channel defined by (1) when $\mathbb{E}[R] > 0$, we construct a deterministic coding scheme that can achieve a strictly positive rate. The idea is to degrade the channel defined by (1) to a binary memoryless Z-channel with a known cross-over probability.

For each time slot t, let $\boldsymbol{G}[t]$ be a random matrix over $\mathbb{F}_q^{1\times M}$ with uniform i.i.d. components. Define a binary-input binary-output channel as follows. Let $B[t]$ be the input of the channel at time t, which takes the value 0 or 1 in $\mathbb{F}_q$. The output of the channel at the time t is $Y[t] = \mathrm{rk}\,(B[t]\boldsymbol{G}[t]\boldsymbol{H}[t])$. Since the dimension of the matrix $B[t]\boldsymbol{G}[t]\boldsymbol{H}[t]$ is $1 \times N$, $Y[t]$ takes the integer value 0 or 1. Let us check the transition matrix of this channel. If $B[t] = 0$, then $Y[t] = 0$. If $B[t] = 1$, then $Y[t] = \mathrm{rk}\,(\boldsymbol{G}[t]\boldsymbol{H}[t])$. Note that $\mathrm{rk}\,(\boldsymbol{G}[t]\boldsymbol{H}[t])$ is a random variable whose distribution only depends on the distribution of $\mathrm{rk}\,(\boldsymbol{H}[t]) \sim R$ (see the computation in [12, Section IV]). Since $\mathrm{rk}\,(\boldsymbol{H}[t])$, $t = 1, 2, \ldots$ are independent, the channel is a binary memoryless Z channel.

What remains is to check the cross over probability of the Z channel given by

$$\Pr\{Y[t]=0|X[t]=1\}=\Pr\{\mathrm{rk}\,(\boldsymbol{G}[t]\boldsymbol{H}[t])=0\}.$$

Since $\mathbb{E}\left[\mathrm{rk}\,(\boldsymbol{H}[t])\right]=\mathbb{E}\left[R\right]>0$, $\Pr\{\mathrm{rk}\,(\boldsymbol{G}[t]\boldsymbol{H}[t])=0\}<1$, because otherwise $\boldsymbol{H}[t]$ is the zero matrix with probability one, a contradiction to the assumption that $\mathbb{E}\left[R\right]>0$. Hence, the channel has a positive capacity. ■

Definition 3 ([14])**.** *A random matrix is called* u.g.r. *(uniform given rank) if any two matrices with the same rank are equiprobable.*

Lemma 2. *For any $M\times N$ random matrix $\boldsymbol{H}$, $\boldsymbol{AHB}$ is u.g.r. with the same rank distribution as of $\boldsymbol{H}$, where $\boldsymbol{A}\sim\mathsf{Uni}(\mathbb{F}_q^{M\times M},M)$ and $\boldsymbol{B}\sim\mathsf{Uni}(\mathbb{F}_q^{N\times N},N)$ are uniform and full-rank random matrices, and $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{H}$ are independent.*

Proof: Let $\boldsymbol{G}=\boldsymbol{AHB}$. Then

$$P_{\boldsymbol{G}}(\boldsymbol{g})=\sum_{\substack{\boldsymbol{a}\in\mathbb{F}_q^{M\times M},\boldsymbol{b}\in\mathbb{F}_q^{N\times N},\\ \mathrm{rk}(\boldsymbol{a})=M,\mathrm{rk}(\boldsymbol{b})=N}}P_{\boldsymbol{A}}(\boldsymbol{a})P_{\boldsymbol{B}}(\boldsymbol{b})P_{\boldsymbol{H}}(\boldsymbol{a}^{-1}\boldsymbol{g}\boldsymbol{b}^{-1}),$$

where $P_{\boldsymbol{A}}(\boldsymbol{a})$ and $P_{\boldsymbol{B}}(\boldsymbol{b})$ respectively do not depend on $\boldsymbol{a}$ and $\boldsymbol{b}$. Now, for another instance $\boldsymbol{g}'$ of $\boldsymbol{G}$ with $\boldsymbol{g}'=\boldsymbol{UgV}$ for some full rank matrices $\boldsymbol{U}$ and $\boldsymbol{V}$, we can see that $P_{\boldsymbol{G}}(\boldsymbol{g})=P_{\boldsymbol{G}}(\boldsymbol{g}')$. In the following we show that if $\mathrm{rk}\,(\boldsymbol{g})=\mathrm{rk}\,(\boldsymbol{g}')$, then there exist full rank matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ such that $\boldsymbol{g}'=\boldsymbol{UgV}$.

Fix two decompositions $\boldsymbol{g}=\boldsymbol{bc}$ and $\boldsymbol{g}'=\boldsymbol{b}'\boldsymbol{c}'$ with $\mathrm{rk}\,(\boldsymbol{b})=\mathrm{rk}\left(\boldsymbol{b}'\right)=\mathrm{rk}\,(\boldsymbol{g})$, which implies $\mathrm{rk}\,(\boldsymbol{c})=\mathrm{rk}\,(\boldsymbol{c}')=\mathrm{rk}\,(\boldsymbol{g})$. Then there exist full rank square matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ such that $\boldsymbol{Ub}=\boldsymbol{b}'$ and $\boldsymbol{cV}=\boldsymbol{c}'$. Hence, $\boldsymbol{g}'=\boldsymbol{UgV}$. ■

Lemma 3. *In the capacity expression* (17)*, the u.g.r. distribution for $P_{\boldsymbol{H}|\,\mathrm{rk}(\boldsymbol{H})}$ is a minimizer for the expression.*

Proof: Let $P^*_{\boldsymbol{H}|\,\mathrm{rk}(\boldsymbol{H})}$ be the distribution that minimizes (17). Now consider a new channel defined by $\boldsymbol{AHB}$ where $\boldsymbol{A}\sim\mathsf{Uni}(\mathbb{F}_q^{M\times M},M)$ and $\boldsymbol{B}\sim\mathsf{Uni}(\mathbb{F}_q^{N\times N},N)$ are uniform full rank random matrices (note that $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{H}$ are independent). Then by Lemma 2, the rank distribution of $\boldsymbol{AHB}$ is the same as that of $\boldsymbol{H}$, but $\boldsymbol{AHB}$ has a u.g.r. distribution.

By the data processing inequality, the mutual information between the input and output of the new channel is less than or equal to the original channel. So if $P^*_{\boldsymbol{H}|\,\mathrm{rk}(\boldsymbol{H})}$ is a minimizer, then the u.g.r. distribution with the same rank distribution is also a minimizer. ■

From Corollary 1, Corollary 2, Lemma 1, and Lemma 3 we obtain the following theorem.

Theorem 4. *The randomized and deterministic code capacities of the non-coherent network coding model,* i.e.*, the matrix channel defined by* (1)*, are the same and are equal to the capacity of the matrix channel $\boldsymbol{Y}=\bar{\boldsymbol{H}}\boldsymbol{X}$ where $\bar{\boldsymbol{H}}$ has the same rank distribution as $\boldsymbol{H}$ but has uniform distribution among matrices having the same rank,* i.e.*,*

$$C=\max_{P_{\boldsymbol{X}}}\min_{P_{\boldsymbol{H}|\,\mathrm{rk}(\boldsymbol{H})}}I(\boldsymbol{X};\boldsymbol{Y})=\max_{P_{\boldsymbol{X}}}I(\boldsymbol{X};\bar{\boldsymbol{H}}\boldsymbol{X}).$$

Theorem 4 shows that, if only the knowledge of the rank distribution of the transfer matrix is available, the maximum rate that we can communicate over the channel defined by (1) is equal to the communication rate over a channel which has the same rank distribution but the channel transfer matrix is u.g.r.

Now, it is shown in [14, Theorem 16] that for a matrix multiplicative channel with u.g.r. distribution over the transfer matrix, the subspace coding is sufficient to achieve the capacity. So we have the following corollary.

Corollary 3. *Subspace coding is sufficient to achieve the capacity (randomized and deterministic) of the non-coherent network coding channel discussed in §II-B.*

Although determining the exact value of the capacity in Theorem 4 is still open, as shown in [14], the capacity can be expressed as the solution of a convex optimization problem with only $O\left(\min[M,N]\right)$ parameters which is computationally tractable.

CONCLUSION

In this work, we proposed an arbitrarily varying channel (AVC) approach to model the non-coherent network coding by a matrix channel where the channel statistics is known only up to a rank distribution over the transfer matrix.

The previous works investigate the capacity of non-coherent network coding (modeled by the matrix channel) for certain probability distributions. In contrast, we relax the problem model by considering that only the rank distribution of the transfer matrix is known and apart from that the transfer matrix can be changed arbitrarily from time-slot to time-slot. We believe that this AVC approach better fits to model complex networks where relay nodes perform randomized network coding.

In order to characterize the capacity of such a channel, we defined a new class of channels, called partially AVC (PAVC), with a partial probabilistic constraint over the state space. By extending the previous result on AVC to PAVC, we proved that the subspace coding is optimal to achieve the capacity of non-coherent network coding.

ACKNOWLEDGMENT

The authors would like to thank Ning Cai and Emre Telatar for many useful discussions. The work of M. Jafari Siavoshani was supported by the Swiss National Science Foundation through Grant PP00P2-128639. The work of S. Yang and R. W. Yeung was partially supported by a grant from the University Grants Committee of the Hong Kong Special Administrative Region, China (Project No. AoE/E-02/08).

APPENDIX A
PROOF OF THEOREM 1

In this section, we prove Theorem 1. The proof goes along similar steps as it goes in [19]. However, for completeness, we will be going to write the whole steps here.

Let us start with some definitions. For $\eta \geq 0$, let us define a family of joint distribution P_{XSY} of random variables X, S, and Y with values from the sets $\mathcal{X}$, $\mathcal{S}$, and $\mathcal{Y}$, respectively, by

$$\mathscr{D}_\eta \triangleq \left\{P_{XSY} :\ D\left(P_{XSY}||P_X \times P_S \times W\right) \leq \eta \text{ where } P_S(s) = P_Q(\mathfrak{q}(s)) \times P_{S|\mathfrak{q}(S)}(s|\mathfrak{q}(s))\right\}, \tag{18}$$

where $D(\cdot||\cdot)$ denotes Kullback-Leibler information divergence and $P_X \times P_Q \times P_{S|\mathfrak{q}(S)} \times W$ denotes a joint distribution on $\mathcal{X} \times \mathcal{S} \times \mathcal{Y}$ with probability mass function $P_X(x)P_Q(\mathfrak{q}(s))P_{S|\mathfrak{q}(S)}(s|\mathfrak{q}(s))W(y|x;s)$. Note that in the above definitions, P_Q is known and fix for a particular PAVC. We also define, for any distribution P on $\mathcal{X}$, the quantity

$$I(P) \triangleq \min_{\substack{P_{S|\mathfrak{q}(S)}:\\ P_{XSY}\in\mathscr{D}_0,\ P_X=P}} I(X;Y), \tag{19}$$

where $\mathscr{D}_0$ denotes $\mathscr{D}_\eta$ for $\eta = 0$.

From [17], we define the *type* of a sequence $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathcal{X}^n$ to be the distribution $P_{\boldsymbol{x}}$ on $\mathcal{X}$ where $P_{\boldsymbol{x}}(a)$ is the relative frequency of $a \in \mathcal{X}$ in $\boldsymbol{x}$. Similarly, *joint types* are distributions on product spaces. Joint types of length-n sequences will be represented by joint distributions of dummy random variables. For example, if X, S, Y represents a joint type, *i.e.,* $P_{XSY} = P_{\boldsymbol{x},\boldsymbol{s},\boldsymbol{y}}$ for some $\boldsymbol{x} \in \mathcal{X}^n$, $\boldsymbol{s} \in \mathcal{S}^n$, and $\boldsymbol{y} \in \mathcal{Y}^n$, we write

$$\begin{aligned}
\mathsf{T}_X &\triangleq \left\{\boldsymbol{x} :\ \boldsymbol{x} \in \mathcal{X}^n, P_{\boldsymbol{x}} = P_X\right\},\\
\mathsf{T}_{XY} &\triangleq \left\{(\boldsymbol{x},\boldsymbol{y}) :\ \boldsymbol{x} \in \mathcal{X}^n, \boldsymbol{y} \in \mathcal{Y}^n, P_{\boldsymbol{x},\boldsymbol{y}} = P_{XY}\right\},\\
\mathsf{T}_{XSY} &\triangleq \left\{(\boldsymbol{x},\boldsymbol{s},\boldsymbol{y}) :\ \boldsymbol{x} \in \mathcal{X}^n, \boldsymbol{s} \in \mathcal{S}^n, \boldsymbol{y} \in \mathcal{Y}^n, P_{\boldsymbol{x},\boldsymbol{s},\boldsymbol{y}} = P_{XSY}\right\}.
\end{aligned} \tag{20}$$

Similarly, we use notation for sections of T_{XY}, T_{XSY}, etc.; for example

$$\begin{aligned}
\mathsf{T}_{Y|X}(\boldsymbol{x}) &\triangleq \left\{\boldsymbol{y} :\ (\boldsymbol{x},\boldsymbol{y}) \in \mathsf{T}_{XY}\right\},\\
\mathsf{T}_{Y|XS}(\boldsymbol{x},\boldsymbol{s}) &\triangleq \left\{\boldsymbol{y} :\ (\boldsymbol{x},\boldsymbol{s},\boldsymbol{y}) \in \mathsf{T}_{XSY}\right\}.
\end{aligned} \tag{21}$$

Lemma 4. *If the PAVC is non-symmetrizable (see Definition 2), then $I(P)$ defined by* (19) *is positive for every P satisfying $P(x) > 0$ for all $x \in \mathcal{X}$.*

Proof: In fact, if $I(P)$ were zero for such a P, then (19) implies the existence of random variable S such that for $P_{XSY} = P_X P_Q P_{S|\mathfrak{q}(S)} W$, X and Y are independent. Thus, we have

$$\sum_{s\in\mathcal{S}} W(y|x;s)P_{S|\mathfrak{q}(S)}(s|\mathfrak{q}(s))P_Q(\mathfrak{q}(s)) = P_Y(y),$$

which does not depend on x. This implies the symmetrizability of the channel in a trivial manner, with $U(s|x,q) = P_{S|\mathfrak{q}(S)}(s|q)$, which leads to a contradiction. ■

Now, the proof of Theorem 1 proceeds as follows.

Proof of Theorem 1: First, note that by [18, Lemma 3.1] we have

$$\max_{P_X} \min_{P_{S|\mathfrak{q}(S)}} I(P_X, \bar{W}_S) = \min_{P_{S|\mathfrak{q}(S)}} \max_{P_X} I(P_X, \bar{W}_S). \tag{22}$$

The converse part of this theorem follows by applying Lemma 5 and Lemma 6.

By Lemma 4, non-symmetrizability implies that $I(P) > 0$ for every strictly positive P. In order to prove that for a non-symmetrizable PAVC, $\max_P I(P)$ is an achievable rate, we use the continuity of $I(P)$ as a function of P and by applying Lemma 12, we conclude the achievability part of Theorem 1. ■

The following lemma, Lemma 5, is similar to [19, Lemma 1] and describes the converse part of the proof when the channel is symmetrizable.

Lemma 5. *For a symmetrizable PAVC, any deterministic code of block length n with $K \geq 2$ codewords, each of type P has*

$$\mathbb{E}\left[\bar{e}_d(\boldsymbol{S})\right] = \max_{P_{\boldsymbol{S}|\mathfrak{q}(\boldsymbol{S})}} \sum_{\boldsymbol{s}\in\mathcal{S}^n} \bar{e}_d(\boldsymbol{s}) P_{\boldsymbol{S}|\mathfrak{q}(\boldsymbol{S})}(\boldsymbol{s}|\mathfrak{q}(\boldsymbol{s})) P_{Q^n}(\mathfrak{q}(\boldsymbol{s})) \geq \frac{1}{4}. \tag{23}$$

Proof: Consider an arbitrary code with codeword set $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_K\}$ and decoder ϕ, where $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{in})$ for $i \in [1:K]$. For some $U \in \mathcal{U}(\mathcal{X} \times \mathcal{Q} \to \mathcal{S})$ satisfying (13) consider K random sequences $\boldsymbol{S}_j = (S_{j1}, \ldots, S_{jn})$ where $\boldsymbol{S}_j \in \mathcal{S}^n$, with statistically independent components, where

$$\mathbb{P}\left[S_{jk} = s\right] = U(s|x_{jk}, \mathfrak{q}(s)) P_Q(\mathfrak{q}(s)). \tag{24}$$

Then for each pair (i,j) and every $\boldsymbol{y} = (y_1, \ldots, y_n) \in \mathcal{Y}^n$ we can write

$$\begin{aligned}\mathbb{E}\left[W^n(\boldsymbol{y}|\boldsymbol{x}_i, \boldsymbol{S}_j)\right] &= \prod_{k=1}^{n} \mathbb{E}\left[W(y_k|x_{ik}, S_{jk})\right] \\ &= \prod_{k=1}^{n} \sum_{s\in\mathcal{S}} W(y_k|x_{ik}, s) U(s|x_{jk}, \mathfrak{q}(s)) P_Q(\mathfrak{q}(s)).\end{aligned} \tag{25}$$

So, by using (13), it follows that

$$\mathbb{E}\left[W^n(\boldsymbol{y}|\boldsymbol{x}_i, \boldsymbol{S}_j)\right] = \mathbb{E}\left[W^n(\boldsymbol{y}|\boldsymbol{x}_j, \boldsymbol{S}_i)\right], \tag{26}$$

and hence for $i \neq j$ we have

$$\begin{aligned}\mathbb{E}\left[e_d(i, \boldsymbol{S}_j)\right] + \mathbb{E}\left[e_d(j, \boldsymbol{S}_i)\right] &= \sum_{\boldsymbol{y}:\ \phi(\boldsymbol{y})\neq i} \mathbb{E}\left[W^n(\boldsymbol{y}|\boldsymbol{x}_i; \boldsymbol{S}_j)\right] + \sum_{\boldsymbol{y}:\ \phi(\boldsymbol{y})\neq j} \mathbb{E}\left[W^n(\boldsymbol{y}|\boldsymbol{x}_j; \boldsymbol{S}_i)\right] \\ &\geq \sum_{\boldsymbol{y}\in\mathcal{Y}^n} \mathbb{E}\left[W^n(\boldsymbol{y}|\boldsymbol{x}_i; \boldsymbol{S}_j)\right] \\ &= 1.\end{aligned} \tag{27}$$

Now, using this fact we can write

$$\begin{aligned}\frac{1}{K}\sum_{j=1}^{K}\mathbb{E}\left[\bar{e}_d(\boldsymbol{S}_j)\right] &= \frac{1}{K^2}\sum_{i=1}^{K}\sum_{j=1}^{K}\mathbb{E}\left[e_d(i,\boldsymbol{S}_j)\right]\\ &\geq \frac{1}{K^2}\cdot\frac{K(K-1)}{2}\\ &= \frac{K-1}{2K},\end{aligned} \tag{28}$$

so it follows that for some $j \in [1:K]$ we have

$$\mathbb{E}\left[\bar{e}_d(\boldsymbol{S}_j)\right] \geq \frac{K-1}{2K} \geq \frac{1}{4}. \tag{29}$$

This leads to the desired result because $\mathbb{E}\left[\bar{e}_d(\boldsymbol{S})\right] \geq 1/4$ for some distribution over $\boldsymbol{S}$ such that the kth element of the random sequence $\boldsymbol{S}$ is distributed independently according to the distribution of the form $P_{S|\mathfrak{q}(S)}P_Q$ where $P_{S|\mathfrak{q}(S)}(s|q) = U(s|x_{jk},q)$. So in general we have $\max_{P_{\boldsymbol{S}|\mathfrak{q}(\boldsymbol{S})}}\mathbb{E}\left[\bar{e}_d(\boldsymbol{S})\right] \geq 1/4$. ■

The following lemma, Lemma 6, is similar to [19, Lemma 2] and describes the converse part of the proof when the rate is greater than $I(P)$.

Lemma 6. *For any $\delta > 0$ and $\epsilon < 1$, there exists n_0 such that for any code of block length $n \geq n_0$ with codewords, each of type P, $\frac{1}{n}\log K \geq I(P) + \delta$ implies*

$$\mathbb{E}\left[\bar{e}_d(\boldsymbol{S})\right] = \max_{P_{\boldsymbol{S}|\mathfrak{q}(\boldsymbol{S})}}\sum_{\boldsymbol{s}\in\mathcal{S}^n}\bar{e}_d(\boldsymbol{s})P_{\boldsymbol{S}|\mathfrak{q}(\boldsymbol{S})}(\boldsymbol{s}|\mathfrak{q}(\boldsymbol{s}))P_{Q^n}(\mathfrak{q}(\boldsymbol{s})) > \epsilon.$$

Proof: Suppose that $P^*_{S|\mathfrak{q}(S)}$ achieves the minimum in (19). So for

$$P_{XSY}(x,s,y) = P(x)P_Q(\mathfrak{q}(s))P^*_{S|\mathfrak{q}(S)}(s|\mathfrak{q}(s))W(y|x;s) \tag{30}$$

we have $I(X;Y) = I(P)$.

Now consider any code with codewords $\{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_K\}$ and decoder ϕ, and let $\boldsymbol{S} = (S_1,\ldots,S_n)$ be n independent realization of S according to the distribution $P^*_{S|\mathfrak{q}(S)}P_Q$. Then we can write

$$\begin{aligned}\mathbb{E}\left[\bar{e}_d(\boldsymbol{S})\right] &= \frac{1}{K}\sum_{i=1}^{K}\mathbb{E}\left[e_d(i,\boldsymbol{S})\right]\\ &= \frac{1}{K}\sum_{i=1}^{K}\sum_{\boldsymbol{y}:\phi(\boldsymbol{y})\neq i}\mathbb{E}\left[W^n(\boldsymbol{y}|\boldsymbol{x}_i;\boldsymbol{S})\right]\\ &= \frac{1}{K}\sum_{i=1}^{K}\sum_{\boldsymbol{y}:\phi(\boldsymbol{y})\neq i}\prod_{j=1}^{n}\mathbb{E}\left[W(y_j|x_{ij};S_j)\right].\end{aligned} \tag{31}$$

If we introduce a new discrete memory-less channel (DMC) $\bar{W}_S$ defined by

$$\bar{W}_S(y|x) = \mathbb{E}\left[W(y|x;S)\right] = \sum_{s\in\mathcal{S}}W(y|x;s)P_{S|\mathfrak{q}(S)}(s|\mathfrak{q}(s))P_Q(\mathfrak{q}(s)),$$

then we have $\mathbb{E}\left[\bar{e}_d(\boldsymbol{S})\right] = \bar{e}_{(\bar{W}_S)}$, where $\bar{e}_{(\bar{W}_S)}$ is the average probability of error when the given code is used on the DMC $\bar{W}_S$.

Now, notice that (30) means that Y is connected to X by the channel $\bar{W}_S$. As mentioned before, we have $I(X;Y) = I(P)$ so by the strong converse to the coding theorem for a DMC with codewords of type P (see [17, Corollary 1.4, p.104]), $\bar{e}_{(\bar{W}_S)}$ is arbitrary close to 1 if $\frac{1}{n}\log K \geq I(P) + \delta$ and n is large enough. This completes the proof of Lemma 6. ■

In order to prove the achievability part of Theorem 1, we need to define a suitable decoder ϕ. Here, we will use the same decoder as introduced in [19, Definition 3].

Definition 4 ([19, Definition 3])**.** *Given the codewords $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_K\}$, let $\phi(\boldsymbol{y}) = i$ if and only if an $\boldsymbol{s} \in \mathcal{S}^n$ exists such that*

1) *the joint type $P_{\boldsymbol{x}_i, \boldsymbol{s}, \boldsymbol{y}}$ belongs to $\mathscr{D}_\eta$;*
2) *for each competitor $j \neq i$, such that $P_{\boldsymbol{x}_j, \boldsymbol{s}', \boldsymbol{y}} \in \mathscr{D}_\eta$ for some $\boldsymbol{s}' \in \mathcal{S}^n$, we have $I(XY; X'|S) \leq \eta$, where X, X', S, Y denote dummy random variables such that $P_{XX'SY} = P_{\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{s}, \boldsymbol{y}}$.*

If no such i exists, we set $\phi(\boldsymbol{y}) = 0$, i.e.*, declare an error.*

Before proceeding further, let us state the following lemmas (Lemma 7-Lemma 9) which are some basic bounds on types (*e.g.*, see [17, Chapter 1]).

Lemma 7. *The number of possible joint types of sequences of length n is a polynomial in n.*

Lemma 8. *If $\mathsf{T}_X \neq \emptyset$, we have*

$$(n+1)^{-|\mathcal{X}|} \exp\{nH(X)\} \leq |\mathsf{T}_X| \leq \exp\{nH(X)\},$$

and if $\mathsf{T}_{Y|X}(\boldsymbol{x}) \neq \emptyset$, we have

$$(n+1)^{-|\mathcal{X}||\mathcal{Y}|} \exp\{nH(Y|X)\} \leq |\mathsf{T}_{Y|X}(\boldsymbol{x})| \leq \exp\{nH(Y|X)\}.$$

Lemma 9. *For any channel $V : \mathcal{X} \mapsto \mathcal{Y}$, we have*

$$\sum_{\boldsymbol{y} \in \mathsf{T}_{Y|X}(\boldsymbol{x})} V^n(\boldsymbol{y}|\boldsymbol{x}) \leq \exp\{-nD(P_{XY}||P_X \times V)\},$$

where $P_X \times V$ denotes the distribution on $\mathcal{X} \times \mathcal{Y}$ with pmf $P_X(x)V(y|x)$ and $V^n(\boldsymbol{y}|\boldsymbol{x}) \triangleq \prod_{t=1}^n V(y_t|x_t)$.

The set of codewords $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_K\}$ used in proving the achievability result is any set with the properties stated in Lemma 10. It is shown in [19, Appendix] that a randomly chosen codeword set have these properties with probability arbitrarily close to 1.

Lemma 10 ([19, Lemma 3])**.** *For any $\epsilon > 0$, $n \geq n_0(\epsilon)$, $K \geq \exp(n\epsilon)$, and type P, there exist codewords $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_K\}$ in $\mathcal{X}^n$, each of type P, such that for every $\boldsymbol{x} \in \mathcal{X}^n$, $\boldsymbol{s} \in \mathcal{S}^n$, and every joint type $P_{XX'S}$, by setting $R = \frac{1}{n}\log K$, we have*

$$|\{j : (\boldsymbol{x}, \boldsymbol{x}_j, \boldsymbol{s}) \in \mathsf{T}_{XX'S}\}| \leq \exp\left\{n\left(|R - I(X'; XS)|^+ + \epsilon\right)\right\}, \tag{32}$$

$$\frac{1}{K}\left|\{i : (\boldsymbol{x}_i, \boldsymbol{s}) \in \mathsf{T}_{XS}\}\right| \leq \exp(-n\epsilon/2), \qquad \textit{if } I(X;S) > \epsilon, \tag{33}$$

and

$$\frac{1}{K}\left|\{i : (\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{s}) \in \mathsf{T}_{XX'S} \textit{ for some } j \neq i\}\right| \leq \exp(-n\epsilon/2)$$
$$\textit{if } I(X;X'S) - |R - I(X';S)|^+ > \epsilon. \tag{34}$$

In addition to Lemma 10, we need Lemma 11 (which is similar to [19, Lemma 4]), in order to establish the inambiguity of the decoding rule given in Definition 4.

Lemma 11. *If the PAVC is non-symmetrizable and $\beta > 0$, then for a sufficiently small η, no set of random variables X, X', S, S', Y can simultaneously satisfy*

$$P_X = P_{X'} = P \qquad \text{with} \qquad \min_{x \in \mathcal{X}} P(x) \geq \beta, \tag{35}$$

$$P_{XSY} \in \mathscr{D}_\eta, \qquad P_{X'S'Y} \in \mathscr{D}_\eta, \tag{36}$$

and

$$I(XY;X'|S) \leq \eta, \qquad I(X'Y;X|S') \leq \eta. \tag{37}$$

Proof: The proof technique is very similar to the proof of [19, Lemma 4]. ■

So assuming that the decoder ϕ is being used as defined in Definition 4, lemma 11 proves that this decoder is unambiguously defined if η is chosen sufficiently small. In fact, if for some $\boldsymbol{y} \in \mathcal{Y}^n$ and some $i \neq j$, both $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ satisfied conditions (1) and (2) in Definition 4, then some $\boldsymbol{s}$ and $\boldsymbol{s}'$ would exist, with the joint types of $(\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{s}, \boldsymbol{s}', \boldsymbol{y})$ represented by the dummy random variables X, X', S, S', Y (*i.e.,* $(\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{s}, \boldsymbol{s}', \boldsymbol{y}) \in \mathsf{T}_{XX'SS'Y}$) that satisfy conditions stated in Lemma 11. This is in contradiction with Lemma 11.

The following lemma, Lemma 12, provides the error analysis for the decoder given in Definition 4.

Lemma 12. *Given any non-symmetrizable PAVC and arbitrary $\beta > 0$, $\delta > 0$, for any block length $n \geq n_0$ and any type P with $\min_x P(x) > \beta$, there exists a code with codewords $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_K\}$, each of type P, such that*

$$\frac{1}{n}\log K > I(P) - \delta, \tag{38}$$

and

$$\max_{P_{\boldsymbol{S}|\mathfrak{q}(\boldsymbol{S})}} \mathbb{E}\left[\bar{e}_d(\boldsymbol{S})\right] = \max_{P_{\boldsymbol{S}|\mathfrak{q}(\boldsymbol{S})}} \sum_{\boldsymbol{s} \in \mathcal{S}^n} \bar{e}_d(\boldsymbol{s}) P_{\boldsymbol{S}|\mathfrak{q}(\boldsymbol{S})}(\boldsymbol{s}|\mathfrak{q}(\boldsymbol{s})) P_{Q^n}(\mathfrak{q}(\boldsymbol{s})) < \exp(-n\gamma). \tag{39}$$

Here, n_0 and $\gamma > 0$ depend only on the given PAVC, and on β and δ.

Proof: Let $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_K\}$ be as in Lemma 10, with $R = 1/n \log K$ satisfying

$$I(P) - \delta < R < I(P) - \frac{2}{3}\delta, \tag{40}$$

and with ϵ (from Lemma 10) to be specified later. Let the decoder ϕ be as defined in Definition 4. Lemma 11 proves that this decoder ϕ is unambiguously defined if η is chosen sufficiently small.

To bound the decoding error, let us fix $P_{\boldsymbol{S}|\mathfrak{q}(\boldsymbol{S})}$ and write

$$\begin{aligned}\mathbb{E}\left[\bar{e}_d(\boldsymbol{S})\right] &= \sum_{\boldsymbol{s}\in\mathcal{S}^n} \bar{e}_d(\boldsymbol{s}) P_{\boldsymbol{S}|\mathfrak{q}(\boldsymbol{S})}(\boldsymbol{s}|\mathfrak{q}(\boldsymbol{s})) P_{Q^n}(\mathfrak{q}(\boldsymbol{s})) \\ &= \sum_{\boldsymbol{s}\in\mathcal{S}^n} \frac{1}{K}\sum_{i=1}^{K} \sum_{\boldsymbol{y}:\ \phi(\boldsymbol{y})\neq i} W^n(\boldsymbol{y}|\boldsymbol{x}_i;\boldsymbol{s}) P_{\boldsymbol{S}|\mathfrak{q}(\boldsymbol{S})}(\boldsymbol{s}|\mathfrak{q}(\boldsymbol{s})) P_{Q^n}(\mathfrak{q}(\boldsymbol{s})) \\ &= \sum_{\mathsf{T}_{\hat{S}}} \sum_{\boldsymbol{s}\in\mathsf{T}_{\hat{S}}} P_{\boldsymbol{S}|\mathfrak{q}(\boldsymbol{S})}(\boldsymbol{s}|\mathfrak{q}(\boldsymbol{s})) P_{Q^n}(\mathfrak{q}(\boldsymbol{s})) \underbrace{\left(\frac{1}{K}\sum_{i=1}^{K} \sum_{\boldsymbol{y}:\ \phi(\boldsymbol{y})\neq i} W^n(\boldsymbol{y}|\boldsymbol{x}_i;\boldsymbol{s})\right)}_{\leq 1}. \end{aligned} \tag{41}$$

For $\eta \geq 0$, let us define a family of distribution P_S of random variables S with values from the set $\mathcal{S}$ by

$$\mathscr{S}_\eta \triangleq \left\{P_S:\ D\left(P_S||P_Q \times P_{S|\mathfrak{q}(S)}\right) \leq \eta\right\}, \tag{42}$$

where $P_{S|\mathfrak{q}(S)}$ is arbitrary and P_Q is the pmf over the channel classes of the PAVC, *i.e.*, it is known and fixed. Then, by [17, Lemma 2.6, p.32], we may bound summation over $P_{\boldsymbol{S}|\mathfrak{q}(\boldsymbol{S})}(\boldsymbol{s}|\mathfrak{q}(\boldsymbol{s}))P_{Q^n}(\mathfrak{q}(\boldsymbol{s}))$ as follows

$$\begin{aligned}\sum_{\boldsymbol{s}\in\mathsf{T}_{\hat{S}}} P_{\boldsymbol{S}|\mathfrak{q}(\boldsymbol{S})}(\boldsymbol{s}|\mathfrak{q}(\boldsymbol{s}))P_{Q^n}(\mathfrak{q}(\boldsymbol{s})) &\leq \sum_{\boldsymbol{s}\in\mathsf{T}_{\hat{S}}} P_{Q^n}(\mathfrak{q}(\boldsymbol{s})) \\ &= P_{Q^n}(\mathsf{T}_{\hat{Q}}) \\ &\leq \exp\left\{-nD(P_{\hat{Q}}||P_Q)\right\}, \end{aligned} \tag{43}$$

where $P_{\hat{Q}}$ is the distribution on $\mathfrak{q}(\hat{S})$ which is implied by $P_{\hat{S}}$. Now by Lemma 7, we have

$$\mathbb{E}\left[\bar{e}_d(\boldsymbol{S})\right] \leq \sum_{\substack{\mathsf{T}_{\hat{S}}:\\ P_{\hat{S}}\in\mathscr{S}_\eta}} \sum_{\boldsymbol{s}\in\mathsf{T}_{\hat{S}}} P_{\boldsymbol{S}|\mathfrak{q}(\boldsymbol{S})}(\boldsymbol{s}|\mathfrak{q}(\boldsymbol{s}))P_{Q^n}(\mathfrak{q}(\boldsymbol{s})) \underbrace{\left(\frac{1}{K}\sum_{i=1}^{K} \sum_{\boldsymbol{y}:\ \phi(\boldsymbol{y})\neq i} W^n(\boldsymbol{y}|\boldsymbol{x}_i;\boldsymbol{s})\right)}_{\bar{e}_d(\boldsymbol{s})} + \exp\left(-n\frac{\eta}{2}\right). \tag{44}$$

The rest of the proof is similar to that of [19, Lemma 5]. By fixing $\boldsymbol{s}$ such that $P_{\boldsymbol{s}} \in \mathscr{S}_\eta$ and following similar steps stated in [19, Lemma 5], we may bound the inner term in front of summation in the above expression and show that it is exponentially vanishing as $n \to \infty$. This in fact completes the proof of Lemma 12.

However, for completeness, we will state the rest of the proof as well. As we mentioned before, let us fix $\boldsymbol{s}$ such that $P_{\boldsymbol{s}} \in \mathscr{S}_\eta$ and observe that by (33) and Lemma 7 we have

$$\begin{aligned}\frac{1}{K}\left|\left\{i:(\boldsymbol{x}_i,\boldsymbol{s}) \in \bigcup_{I(X;S)>\epsilon} \mathsf{T}_{XS}\right\}\right| &\leq (\text{number of joint types})\cdot\exp(-n\epsilon/2) \\ &\leq \exp(-n\epsilon/3), \end{aligned} \tag{45}$$

for n larger than a suitable threshold n_0, that depends on ϵ.

So, in order to obtain an exponentially decreasing upper bound on $\bar{e}_d(\boldsymbol{s})$ (for those $\boldsymbol{s}$ such that $P_{\boldsymbol{s}} \in \mathscr{S}_\eta$), it is sufficient to consider only those codewords $\boldsymbol{x}_i$ for which $(\boldsymbol{x}_i,\boldsymbol{s}) \in \mathsf{T}_{XS}$ with $I(X;S) \leq \epsilon$. Then, for $P_{XSY} \notin \mathscr{D}_\eta$

(see (18)), we have

$$D(P_{XSY}||P_{XS}\times W) = D(P_{XSY}||P_X\times P_Q\times P_{S|\mathfrak{q}(S)}\times W) - I(X;S)$$
$$> \eta-\epsilon, \tag{46}$$

and thus by Lemma 9, we can write

$$\sum_{\boldsymbol{y}\in\mathsf{T}_{Y|XS}(\boldsymbol{x}_i,\boldsymbol{s})} W^n(\boldsymbol{y}|\boldsymbol{x}_i;\boldsymbol{s}) \le \exp\left\{-D(P_{XSY}||P_{XS}\times W)\right\}$$
$$\le \exp\{-n(\eta-\epsilon)\}.$$

Hence by Lemma 7, we have

$$\sum_{\boldsymbol{y}:\ P_{\boldsymbol{x}_i,\boldsymbol{s},\boldsymbol{y}}\notin\mathscr{D}_\eta} W^n(\boldsymbol{y}|\boldsymbol{x}_i;\boldsymbol{s}) \le \exp\left\{-n(\eta-2\epsilon)\right\}. \tag{47}$$

Next, note that if $P_{\boldsymbol{x}_i,\boldsymbol{s},\boldsymbol{y}}\in\mathscr{D}_\eta$ and $\phi(\boldsymbol{y})\neq i$, then condition (2) of Definition 4 must be violated. So let us denote by $\mathscr{E}_\eta$ the set of all joint distributions $P_{XX'SY}$ such that (i) $P_{XSY}\in\mathscr{D}_\eta$; (ii) $P_{X'S'Y}\in\mathscr{D}_\eta$ for some S'; and (iii) $I(XY;X'|S)>\eta$. Then, it follows that

$$\sum_{\substack{\boldsymbol{y}:\ P_{\boldsymbol{x}_i,\boldsymbol{s},\boldsymbol{y}}\in\mathscr{D}_\eta\\ \phi(\boldsymbol{y})\neq i}} W^n(\boldsymbol{y}|\boldsymbol{x}_i;\boldsymbol{s}) \le \sum_{P_{XX'SY}\in\mathscr{E}_\eta} e_{XX'SY}(i,\boldsymbol{s}), \tag{48}$$

where

$$e_{XX'SY}(i,\boldsymbol{s}) \triangleq \sum_{\substack{\boldsymbol{y}:\ (\boldsymbol{x}_i,\boldsymbol{x}_j,\boldsymbol{s},\boldsymbol{y})\in\mathsf{T}_{XX'SY}\\ \text{for some } j\neq i}} W^n(\boldsymbol{y}|\boldsymbol{x}_i;\boldsymbol{s}), \tag{49}$$

and the summation (48) extends to all joint types $P_{XX'SY}\in\mathscr{E}_\eta$ (of course, $e_{XX'SY}(i,\boldsymbol{s})=0$ unless $P_{X'}=P_X=P$ and $P_{XS}=P_{\boldsymbol{x}_i,\boldsymbol{s}}$).

Combining (45)-(48), for those $\boldsymbol{s}$ such that $P_{\boldsymbol{s}}\in\mathscr{S}_\eta$, we obtain that

$$\bar{e}_d(\boldsymbol{s}) \le \exp\{-n\epsilon/3\} + \exp\{-n(\eta-2\epsilon)\} + \frac{1}{K}\sum_{i=1}^{K}\sum_{P_{XX'SY}\in\mathscr{E}_\eta} e_{XX'SY}(i,\boldsymbol{s}). \tag{50}$$

Before finding an upper bound for $e_{XX'SY}(i,\boldsymbol{s})$, note that it is sufficient to do so only when $P_{XX'SY}\in\mathscr{E}_\eta$ satisfies

$$I(X;X'S) \le \left|R-I(X';S)\right|^+ + \epsilon, \tag{51}$$

otherwise, by (34), we have

$$\frac{1}{K}\left|\{i:(\boldsymbol{x}_i,\boldsymbol{x}_j,\boldsymbol{s})\in\mathsf{T}_{XX'S}\ \text{ for some } j\neq i\}\right| < \exp\{-n\epsilon/2\}. \tag{52}$$

Since $(\boldsymbol{x}_i,\boldsymbol{x}_j,\boldsymbol{s})\in\mathsf{T}_{XX'S}$ for some $j\neq i$ is a necessary condition for $e_{XX'SY}(i,\boldsymbol{s})>0$ (see (49)), it follows from Lemma 7 that the contribution to the double summation in (50) of the terms with $P_{XX'SY}\in\mathscr{E}_\eta$ not satisfying (51) is less than $\exp\{-n\epsilon/3\}$.

Now, from (49), we can write

$$e_{XX'SY}(i,\boldsymbol{s}) \le \sum_{j:(\boldsymbol{x}_i,\boldsymbol{x}_j,\boldsymbol{s})\in\mathsf{T}_{XX'S}}\ \sum_{\boldsymbol{y}\in\mathsf{T}_{Y|XX'S}(\boldsymbol{x}_i,\boldsymbol{x}_j,\boldsymbol{s})} W^n(\boldsymbol{y}|\boldsymbol{x}_i;\boldsymbol{s}). \tag{53}$$

Because $W^n(\boldsymbol{y}|\boldsymbol{x}_i;\boldsymbol{s})$ is constant for $\boldsymbol{y} \in \mathsf{T}_{Y|XS}(\boldsymbol{x}_i,\boldsymbol{s})$ and this constant is less than or equal to $(|\mathsf{T}_{Y|XS}(\boldsymbol{x}_i,\boldsymbol{s})|)^{-1}$, the inner sum in (53) is bounded above by

$$\left|\mathsf{T}_{Y|XX'S}(\boldsymbol{x}_i,\boldsymbol{x}_j,\boldsymbol{s})\right| \cdot \left(\left|\mathsf{T}_{Y|XS}(\boldsymbol{x}_i,\boldsymbol{s})\right|\right)^{-1},$$

which in turn, by Lemma 8, is less than or equal to $\exp\{-n[I(Y;X'|XS)-\epsilon]\}$. Now by using (32), it follows from (53) that

$$e_{XX'SY}(i,\boldsymbol{s}) \leq \exp\left\{-n\left[I(Y;X'|XS) - |R - I(X';XS)|^+ - 2\epsilon\right]\right\}. \tag{54}$$

In order to further bound $e_{XX'SY}(i,\boldsymbol{s})$ when (51) holds, we distinguish between two cases: a) $R \leq I(X';S)$, and b) $R > I(X';S)$.

For the case a), from (51) we have

$$I(X;X'|S) \leq I(X;X'S) \leq \epsilon,$$

and hence by condition (iii) in the definition of $\mathscr{E}_\eta$, we can write

$$I(Y;X'|XS) = I(XY;X'|S) - I(X;X'|S) \geq \eta - \epsilon.$$

Since for this case we have $R \leq I(X';S) \leq I(X';XS)$, it follows from (54) that

$$e_{XX'SY}(i,\boldsymbol{s}) \leq \exp\{-n(\eta - 3\epsilon)\}. \tag{55}$$

In case b), from (51) we have

$$\begin{aligned} R &> I(X;X'S) + I(X';S) - \epsilon \\ &= I(X';XS) + I(X;S) - \epsilon \\ &\geq I(X';XS) - \epsilon, \end{aligned}$$

and hence

$$|R - I(X';XS)|^+ \leq R - I(X';XS) + \epsilon.$$

Substituting this into (54) it follows that

$$\begin{aligned} e_{XX'SY}(i,\boldsymbol{s}) &\leq \exp\left\{-n\left[I(X';XSY) - R - 3\epsilon\right]\right\} \\ &\leq \exp\left\{-n\left[I(X';Y) - R - 3\epsilon\right]\right\}. \end{aligned} \tag{56}$$

Note that $P_{XX'SY} \in \mathscr{E}_\eta$ implies that $P_{X'S'Y} \in \mathscr{D}_\eta$ for some S'. So by definition of $\mathscr{D}_\eta$ given in (18), $P_{X'S'Y}$ is arbitrary close to $P_{X''S''Y''} \in \mathscr{D}_0$ defined by $P_{X''S''Y''} = P \times P_Q \times P_{S'|\mathfrak{q}(S')} \times W$. Now if η is sufficiently small, then $I(X';Y)$ is arbitrarily close to $I(X'';Y'')$, say, $I(X';Y) \geq I(X'';Y'') - \delta/3$. Using the definition of $I(P)$ given in (19) and the assumption (40), we can write

$$I(X';Y) - R \geq I(X'';Y'') - \delta/3 - R \geq I(P) - \delta/3 - R \geq \delta/3,$$

if η is sufficiently small and depends only on δ. Fixing η accordingly and also small enough for the decoding rule to be unambiguous, (56) yields for case b) that

$$e_{XX'SY}(i,\boldsymbol{s}) \leq \exp\left\{-n\left[\frac{\delta}{3}-3\epsilon\right]\right\}. \tag{57}$$

Now, from (50), by using (55) and (57) and Lemma 7, we obtain that

$$\bar{e}_d(\boldsymbol{s}) \leq \exp(-n\epsilon/4),$$

if, for instance, $\epsilon \leq \min[\eta/4, \delta/10]$ and n is sufficiently large. Because the bound holds uniformly for those $\boldsymbol{s}$ such that $P_{\boldsymbol{s}} \in \mathscr{S}_\eta$, then by substituting it into (44) and using Lemma 7, the proof of Lemma 12 becomes complete. ■

APPENDIX B
PROOF OF THEOREM 2

Proof of Theorem 2: Because deterministic codes are special cases of codes with stochastic encoder, the achievability part of this theorem directly follows from that of Theorem 1.

The converse part of the theorem follows from similar steps that have been used in the proof of Theorem 1, *i.e.*, Lemma 5 and Lemma 6.

When the rate is greater than $I(P)$, defined in (19), the converse proof follows from the converse proof of randomized codes, *i.e.*, Lemma 14, by choosing the random decoder Φ to be a fixed decoder ϕ (this does not change any part of the proof). When the channel is symmetrizable, the converse follows from Lemma 13 and this completes the proof. ■

Lemma 13. *For a symmetrizable PAVC, any stochastic code of block length n with $K \geq 2$ codewords, each of type P has*

$$\mathbb{E}\left[\bar{e}_d(\boldsymbol{S})\right] = \max_{P_{\boldsymbol{S}|\mathfrak{q}(\boldsymbol{S})}} \sum_{\boldsymbol{s}\in\mathcal{S}^n} \bar{e}_d(\boldsymbol{s}) P_{\boldsymbol{S}|\mathfrak{q}(\boldsymbol{S})}(\boldsymbol{s}|\mathfrak{q}(\boldsymbol{s})) P_{Q^n}(\mathfrak{q}(\boldsymbol{s})) \geq \frac{1}{4}. \tag{58}$$

Proof: Consider an arbitrary stochastic code (Ψ, ϕ) which is defined over the message set $\mathcal{M} = \{1, \ldots, K\}$. Let the random variable Ψ be defined over a set of L encoders $\left\{\psi^{(1)}, \ldots, \psi^{(L)}\right\}$ with a pmf P_Ψ where $P_\Psi(l)$ is the probability of choosing the lth encoder $\psi^{(l)}$.

For some $U \in \mathcal{U}(\mathcal{X} \times \mathcal{Q} \to \mathcal{S})$ satisfying (13) consider K random sequences $\boldsymbol{S}_j = (S_{j1}, \ldots, S_{jn})$ where $\boldsymbol{S}_j \in \mathcal{S}^n$, $j \in [1:K]$, is chosen according to the following distribution

$$\begin{aligned}
\mathbb{P}\left[\boldsymbol{S}_j = \boldsymbol{s}\right] &= \sum_{l=1}^{L}\left[\prod_{k=1}^{n} U(s_k|\psi^{(l)}(j)_k, \mathfrak{q}(s_k)) P_Q(\mathfrak{q}(s_k))\right] P_\Psi(l) \\
&= \left[\sum_{l=1}^{L}\prod_{k=1}^{n} U(s_k|\psi^{(l)}(j)_k, \mathfrak{q}(s_k)) P_\Psi(l)\right]\left[\prod_{k'=1}^{n} P_Q(\mathfrak{q}(s_{k'}))\right] \\
&= \underbrace{\left[\sum_{l=1}^{L}\prod_{k=1}^{n} U(s_k|\psi^{(l)}(j)_k, \mathfrak{q}(s_k)) P_\Psi(l)\right]}_{P_{\boldsymbol{S}|\mathfrak{q}(\boldsymbol{S})}} P_{Q^n}(\mathfrak{q}(\boldsymbol{s})).
\end{aligned} \tag{59}$$

Then for each pair (i,j) and every $\boldsymbol{y}=(y_1,\ldots,y_n)\in\mathcal{Y}^n$ we can write

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{S}_j}\left[\mathbb{E}_{\Psi}\left[W^n(\boldsymbol{y}|\Psi(i);\boldsymbol{S}_j)\right]\right]\\
&\quad=\mathbb{E}_{\Psi}\left[\sum_{\boldsymbol{s}\in\mathcal{S}^n}\left[\prod_{k=1}^{n}W(y_k|\Psi(i)_k;s_k)\right]\mathbb{P}\left[\boldsymbol{S}_j=\boldsymbol{s}\right]\right]\\
&\quad=\mathbb{E}_{\Psi}\left[\sum_{l=1}^{L}\left[\sum_{\boldsymbol{s}\in\mathcal{S}^n}\prod_{k=1}^{n}W(y_k|\Psi(i)_k;s_k)U(s_k|\psi^{(l)}(j)_k,\mathfrak{q}(s_k))P_Q(\mathfrak{q}(s_k))\right]P_{\Psi}(l)\right]\\
&\quad=\sum_{l'=1}^{L}\sum_{l=1}^{L}\left[\sum_{\boldsymbol{s}\in\mathcal{S}^n}\prod_{k=1}^{n}W(y_k|\psi^{(l')}(i)_k;s_k)U(s_k|\psi^{(l)}(j)_k,\mathfrak{q}(s_k))P_Q(\mathfrak{q}(s_k))\right]P_{\Psi}(l)P_{\Psi}(l')\\
&\quad=\sum_{l'=1}^{L}\sum_{l=1}^{L}\left[\prod_{k=1}^{n}\sum_{s\in S}W(y_k|\psi^{(l')}(i)_k;s)U(s|\psi^{(l)}(j)_k,\mathfrak{q}(s))P_Q(\mathfrak{q}(s))\right]P_{\Psi}(l)P_{\Psi}(l').
\end{aligned}\tag{60}$$

So, by using (13), it follows that

$$\mathbb{E}_{\boldsymbol{S}_j}\left[\mathbb{E}_{\Psi}\left[W^n(\boldsymbol{y}|\Psi(i);\boldsymbol{S}_j)\right]\right]=\mathbb{E}_{\boldsymbol{S}_i}\left[\mathbb{E}_{\Psi}\left[W^n(\boldsymbol{y}|\Psi(j);\boldsymbol{S}_i)\right]\right],\tag{61}$$

and hence for $i\neq j$ we have

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{S}_j}\left[e_t(i,\boldsymbol{S}_j)\right]+\mathbb{E}_{\boldsymbol{S}_i}\left[e_t(j,\boldsymbol{S}_i)\right]&=\sum_{\boldsymbol{y}:\ \phi(\boldsymbol{y})\neq i}\mathbb{E}_{\boldsymbol{S}_j}\left[\mathbb{E}_{\Psi}\left[W^n(\boldsymbol{y}|\Psi(i);\boldsymbol{S}_j)\right]\right]+\sum_{\boldsymbol{y}:\ \phi(\boldsymbol{y})\neq j}\mathbb{E}_{\boldsymbol{S}_i}\left[\mathbb{E}_{\Psi}\left[W^n(\boldsymbol{y}|\Psi(j);\boldsymbol{S}_i)\right]\right]\\
&\geq\sum_{\boldsymbol{y}\in\mathcal{Y}^n}\mathbb{E}_{\boldsymbol{S}_j}\left[\mathbb{E}_{\Psi}\left[W^n(\boldsymbol{y}|\Psi(i);\boldsymbol{S}_j)\right]\right]\\
&=1.
\end{aligned}\tag{62}$$

Now, from here on the proof is very similar to that of Lemma 5. Using the above fact we can write

$$\begin{aligned}
\frac{1}{K}\sum_{j=1}^{K}\mathbb{E}_{\boldsymbol{S}_j}\left[\bar{e}_t(\boldsymbol{S}_j)\right]&=\frac{1}{K^2}\sum_{i=1}^{K}\sum_{j=1}^{K}\mathbb{E}_{\boldsymbol{S}_j}\left[e_t(i,\boldsymbol{S}_j)\right]\\
&\geq\frac{1}{K^2}\cdot\frac{K(K-1)}{2}\\
&=\frac{K-1}{2K},
\end{aligned}\tag{63}$$

so it follows that for some $j\in[1:K]$ we have

$$\mathbb{E}_{\boldsymbol{S}_j}\left[\bar{e}_t(\boldsymbol{S}_j)\right]\geq\frac{K-1}{2K}\geq\frac{1}{4}.\tag{64}$$

This leads to the desired result because $\mathbb{E}\left[\bar{e}_t(\boldsymbol{S})\right]\geq 1/4$ for some distribution over $\boldsymbol{S}$ of the form $P_{\boldsymbol{S}|\mathfrak{q}(\boldsymbol{S})}P_{Q^n}$ where $P_{\boldsymbol{S}|\mathfrak{q}(\boldsymbol{S})}$ is given in (59). So in general we have $\max_{P_{\boldsymbol{S}|\mathfrak{q}(\boldsymbol{S})}}\mathbb{E}\left[\bar{e}_d(\boldsymbol{S})\right]\geq 1/4$ and we are done. ■

APPENDIX C
PROOF OF THEOREM 3

Suppose that there are k non-negative-valued functions $l_1,\ldots,l_k$ on $\mathcal{S}$ where for simplicity we assume that $\min_{s\in\mathcal{S}}l_i(s)=0$. Given $\Lambda_1,\ldots,\Lambda_k$, we say that $\boldsymbol{s}\in\mathcal{S}^n$ satisfies state constraints $\Lambda_1,\ldots,\Lambda_k$, if $l_i(\boldsymbol{s})\leq\Lambda_i$ for

all i, where

$$l(\boldsymbol{s}) = \frac{1}{n}\sum_{t=1}^{n} l(s_t), \qquad \boldsymbol{s} \in \mathcal{S}^n.$$

By applying the same method of [18], the result of [18, Theorem 3.1] can be extended to multiple state constraints as stated in the following result.

Theorem 5. *The* randomized *code capacity of the AVC* (2) *under state constraint* $\Lambda_1, \ldots, \Lambda_k$*, denoted by* $C_{\mathsf{avc}}^{\mathsf{r}}(\Lambda)$*, is determined in [18], and is given by*

$$C_{\mathsf{avc}}^{\mathsf{r}}(\Lambda_1, \ldots, \Lambda_k) = \max_{P_X} \min_{P_S:\forall i\ \mathbb{E}[l_i(S)]\le\Lambda_i} I(P_X, \bar{W}_S) = \min_{P_S:\forall i\ \mathbb{E}[l_i(S)]\le\Lambda_i} \max_{P_X} I(P_X, \bar{W}_S).$$

Proof of Theorem 3: The converse part, using a similar argument to [18, Lemma 3.2 and Theorem 3.1], follows from Lemma 14. In the following we prove the achievability part.

Define an AVC with the following convergent state constraints. For each $i \in \mathcal{Q}$, define a non-negative-valued function l_i on $\boldsymbol{s} \in \mathcal{S}^n$ as

$$l_i(\boldsymbol{s}) \triangleq \frac{1}{n}\sum_{t=1}^{n} \mathbb{1}_{\mathfrak{q}(s_t)=i}.$$

For any $\epsilon > 0$, consider the state constraints

$$|l_i(\boldsymbol{s}) - P_Q(i)| \le \epsilon, \forall i \in \mathcal{Q}. \tag{65}$$

By Theorem 5, the capacity of the AVC under the state constraints (65) is

$$C_{\mathsf{avc}}^{\mathsf{r}}(P_Q, \epsilon) \triangleq \max_{P_X} \min_{\substack{P_S:\\ \forall i\in\mathcal{Q},\ |\mathbb{P}[\mathfrak{q}(S)=i]-P_Q(i)|\le\epsilon}} I(P_X, \bar{W}_S) = \min_{\substack{P_S:\\ \forall i\in\mathcal{Q},\ |\mathbb{P}[\mathfrak{q}(S)=i]-P_Q(i)|\le\epsilon}} \max_{P_X} I(P_X, \bar{W}_S),$$

where we use $\mathbb{E}\left[l_i(S)\right] = \mathbb{P}\left[\mathfrak{q}(S) = i\right]$. By the monotonicity and the continuity of $C_{\mathsf{avc}}^{\mathsf{r}}(P_Q, \epsilon)$ as a function of ϵ,

$$C_{\mathsf{pavc}}^{\mathsf{r}} = \sup_{\epsilon>0} C_{\mathsf{avc}}^{\mathsf{r}}(P_Q, \epsilon). \tag{66}$$

Then we show that any rate $\mathfrak{R} < C_{\mathsf{pavc}}^{\mathsf{r}} = \sup_{\epsilon>0} C_{\mathsf{avc}}^{\mathsf{r}}(P_Q, \epsilon)$ is achivable for PAVC.

Pick an ϵ_0 such that $\mathfrak{R} < C_{\mathsf{avc}}^{\mathsf{r}}(P_Q, \epsilon_0)$, which is possible by (66). Fix any $\varepsilon > 0$ and $\delta > 0$. Choose ε' with $0 < \varepsilon' < \varepsilon$. Since $\mathfrak{R}$ is achievable for the AVC with the state constraints (65), with ε' in place of ϵ and for sufficiently large n, there exists a random code (Ψ, Φ) of blocklength n, rate larger than $\mathfrak{R} - \delta$ and

$$\bar{e}_r(\boldsymbol{s}) \le \varepsilon'$$

for all state sequences satisfying (65) with ε' in place of ϵ. For a random sequence $\boldsymbol{S}$ of PAVC, by Hoeffding's inequality,

$$\mathbb{P}\left[|l_i(\boldsymbol{S}) - P_Q(i)| \le \epsilon_0, \forall i \in \mathcal{Q}\right] \ge 1 - 2\exp(-2\epsilon_0^2 n).$$

For random code (Ψ, Φ) with sufficiently large n such that $2\exp(-2\epsilon_0^2 n) < \varepsilon - \varepsilon'$, we have

$$\begin{aligned}\mathbb{E}\left[\bar{e}_r(\boldsymbol{S})\right] &\le \mathbb{E}\left[\bar{e}_r(\boldsymbol{S})|\ |l_i(\boldsymbol{S}) - P_Q(i)| \le \epsilon_0, \forall i \in \mathcal{Q}\right] + \mathbb{P}\left[|l_i(\boldsymbol{S}) - P_Q(i)| > \epsilon_0, \text{ for some } i \in \mathcal{Q}\right] \\ &< \varepsilon' + \varepsilon - \varepsilon'.\end{aligned}$$

Thus for sufficiently large n, there exists blocklength n random code for PAVC with rate larger than $\mathfrak{R}-\delta$ and $\mathbb{E}\left[\bar{e}_r(S)\right]<\varepsilon$. Therefore, $\mathfrak{R}$ is achievable for PAVC. This completes the proof of the theorem. ■

Lemma 14. *For any $\delta>0$ and $\epsilon<1$, there exists n_0 such that for any randomized code (Ψ,Φ) of block length $n\geq n_0$, having $\frac{1}{n}\log K\geq \min_{P_{S|\mathfrak{q}(S)}}\max_{P_X} I(P_X,\bar{W}_S)+\delta$ implies*

$$\mathbb{E}\left[\bar{e}_r(\boldsymbol{S})\right]=\max_{P_{\boldsymbol{S}|\mathfrak{q}(\boldsymbol{S})}}\sum_{\boldsymbol{s}\in\mathcal{S}^n}\bar{e}_r(\boldsymbol{s})P_{\boldsymbol{S}|\mathfrak{q}(\boldsymbol{S})}(\boldsymbol{s}|\mathfrak{q}(\boldsymbol{s}))P_{Q^n}(\mathfrak{q}(\boldsymbol{s}))>\epsilon.$$

Proof: Let us fix $P_{S|\mathfrak{q}(S)}$ and assume that $P_X=P^*$ achieves the maximum of $I(P_X,\bar{W}_S)$ for this choice. Now, let $\boldsymbol{S}=(S_1,\ldots,S_n)$ be n independent realization of S according to the distribution $P_{S|\mathfrak{q}(S)}P_Q$. Then we can write

$$\begin{aligned}
\mathbb{E}\left[\bar{e}_r(\boldsymbol{S})\right]&=\frac{1}{K}\sum_{i=1}^{K}\mathbb{E}\left[e_r(i,\boldsymbol{S})\right]\\
&=\frac{1}{K}\sum_{i=1}^{K}\mathbb{E}_{\boldsymbol{S}}\left[\mathbb{E}_{\Psi,\Phi}\left[e(i,\boldsymbol{S},\Psi,\Phi)\right]\right]\\
&=\frac{1}{K}\sum_{i=1}^{K}\mathbb{E}_{\Psi,\Phi}\left[\sum_{\boldsymbol{y}:\Phi(\boldsymbol{y})\neq i}\mathbb{E}_{\boldsymbol{S}}\left[W^n(\boldsymbol{y}|\Psi(\boldsymbol{x});\boldsymbol{S})\right]\right]\\
&=\mathbb{E}_{\Psi,\Phi}\left[\frac{1}{K}\sum_{i=1}^{K}\sum_{\boldsymbol{y}:\Phi(\boldsymbol{y})\neq i}\prod_{j=1}^{n}\mathbb{E}_{\boldsymbol{S_j}}\left[W(y_j|\Psi(\boldsymbol{x})_j;S_j)\right]\right].
\end{aligned}\tag{67}$$

All of the random variables S_j are i.i.d., so if we introduce a new discrete memory-less channel (DMC) $\bar{W}_S$ defined by

$$\bar{W}_S(y|x)=\mathbb{E}\left[W(y|x;S)\right],$$

then we have

$$\begin{aligned}
\mathbb{E}\left[\bar{e}_r(\boldsymbol{S})\right]&=\mathbb{E}_{\Psi,\Phi}\left[\frac{1}{K}\sum_{i=1}^{K}\sum_{\boldsymbol{y}:\Phi(\boldsymbol{y})\neq i}\prod_{j=1}^{n}\bar{W}_S(y_j|\Psi(\boldsymbol{x})_j)\right],\\
&=\mathbb{E}_{\Psi,\Phi}\left[\bar{e}_{(\bar{W}_S)}(\Psi,\Phi)\right],
\end{aligned}\tag{68}$$

where $\bar{e}_{(\bar{W}_S)}(\psi,\phi)$ is the average probability of error when a code (ψ,ϕ) is used on the DMC $\bar{W}_S$. Now, by using the strong converse to the coding theorem for the DMC $\bar{W}_S$, every code (ψ,ϕ) of rate $R\geq\max_{P_X} I(P_X,\bar{W}_S)+\delta$ has an average error probability $\bar{e}_{(\bar{W}_S)}(\psi,\phi)$ arbitrary close to 1 if n is large enough. So as a result, for every $\epsilon<1$ we have $\mathbb{E}\left[\bar{e}_r(\boldsymbol{S})\right]>\epsilon$ and this completes the proof. ■

REFERENCES

[1] T. Ho, M. Medard, R. Koetter, D. R. Karger, M. Effros, J. Shi, and B. Leong, "A random linear network coding approach to multicast," *IEEE Trans. Inform. Theory*, vol. 52, no. 10, pp. 4413–4430, Oct. 2006.

[2] R. Ahlswede, N. Cai, S-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inform. Theory*, vol. 46, no. 4, pp. 1204–1216, Jul. 2000.

[3] S.-Y. R. Li, N. Cai, and R. W. Yeung, "Linear network coding," *IEEE Trans. Inform. Theory*, vol. 49, no. 2, pp. 371–381, Feb. 2003.

[4] R. Koetter and F. R. Kschischang, "Coding for errors and erasures in random network coding," *IEEE Trans. Inform. Theory*, vol. 54, no. 8, pp. 3579–3591,Aug. 2008.

[5] A. Montanari and R. Urbanke, "Coding for network coding," Dec. 2007, available online : http://arxiv.org/abs/0711.3935/.

[6] M. Jafari Siavoshani, C. Fragouli, and S. Diggavi, "Non-coherent multisource network coding," *IEEE International Symposium on Information Theory*, pp. 817–821, Canada, Toronto, Jul. 2008.

[7] M. Jafari Siavoshani, S. Mohajer, C. Fragouli, and S. Diggavi, "On the capacity of non-coherent network coding", *IEEE International Symposium on Information Theory*, Seoul, Korea, pp. 273–277, Jun. 2009.

[8] M. Jafari Siavoshani, S. Mohajer, C. Fragouli, and S. N. Diggavi, "On the capacity of noncoherent network coding," *IEEE Trans. Inform. Theory*, vol. 57, no. 2, pp. 1046–1066, Feb. 2011.

[9] D. Silva, F. R. Kschischang, and R. Koetter, "Communication over finite-field matrix channels," *IEEE Trans. Inform. Theory*, vol. 56, no. 3, pp. 1296–1305, Mar. 2010.

[10] S. Yang, J. Meng, and E. hui Yang, "Coding for linear operator channels over finite fields," *IEEE International Symposium on Information Theory*, Jun. 2010.

[11] S. Yang, S.-W. Ho, J. Meng, and E.-h. Yang, "Optimality of subspace coding for linear operator channels over nite elds," *IEEE Information Theory Workshop*, pp. 400–404, Jan. 2010.

[12] S. Yang, S.-W. Ho, J. Meng, E. hui Yang, and R. W. Yeung, "On Linear operator channels over finite fields," 2010. [Online]. Available: http://arxiv.org/abs/1002.2293v2

[13] S. Yang, S.-W. Ho, J. Meng, and E.-hui Yang, "Symmetric properties and subspace degradations of linear operator channels over finite fields," 2011. [Online]. Available: http://arxiv.org/abs/1108.4257.

[14] R. W. Nobrega, B. F. Uchoa-Filho, D. Silva, "On the capacity of multiplicative finite-field matrix channels," *IEEE International Symposium on Information Theory*, pp. 341–345, 2011.

[15] D. Blackwell, L. Breiman, and A. J. Thomasian, "The Capacities of Certain Channel Classes Under Random Coding," *The Annals of Mathematical Statistics*, vol. 31, no. 2, pp. 558–567, Sep. 1960.

[16] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *IEEE Trans. Inform. Theory* , vol. 44, no. 6, pp. 2148–2177, Oct. 1998.

[17] I. Csiszar and J. Korner, Information Theory: Coding Theorems for Discrete Memoryless Systems. New York: Academic, 1981.

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

[19] I. Csiszar and P. Narayan, "The capacity of the arbitrarily varying channel revisited: Positivity, constraints," *IEEE Trans. Inform. Theory*, vol. 34, pp. 181–193, Jan. 1988.