Non-coherent Network Coding: An Arbitrarily Varying Channel Approach

Mahdi Jafari Siavoshani*, Shenghao Yang†, Raymond W. Yeung‡
*Ecole Polytechnique Fédérale de Lausanne
Email: mahdi.jafarisiavoshani@epfl.ch
†ITCS, Institute for Interdisciplinary Information Sciences, Tsinghua University
Email: shyang@tsinghua.edu.cn
‡Institute of Network Coding, The Chinese University of Hong Kong
Email: whyeung@ie.cuhk.edu.hk

***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" 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 vector is appended to learn the transfer matrix induced by the network.

A different approach, other 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 [4], the authors proposed communications via choosing subspaces and they introduced 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 to model the non-coherent randomized linear network coding channel, where these models enable one to define and characterize the capacity for such a channel. In all of these works, when there are no errors 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. 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 subspaces is sufficient to achieve the capacity. Moreover, they obtained the channel capacity as a solution of a convex optimization problem over $O(\min[M, N])$ variables and when the field size is greater than a threshold, they characterized 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 the coding over subspaces is sufficient to achieve the capacity. Yang *et al.* [10], [11] (see also [12], [13]) 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, following an approach similar to [8], they expressed the capacity as the solution of a convex optimization problem over $O(\min[M, N])$ variables. They also observed that in this case the subspace codes are sufficient to achieve 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.

978-1-4673-2579-0/12/$31.00 ©2012 IEEE

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 [16] (refer to [17] 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 randomized and deterministic codes (Theorem 1 and 2). We further show that the randomized and the deterministic code capacities of the non-coherent network coding model are the same (Theorem 3), and that subspace coding is sufficient to achieve the capacity (Corollary 4). This AVC approach to the non-coherent network coding provides a justification for the optimality of subspace coding in a more general setting.

II. Problem Setup and Notation

A. Notation

We use bold letters to denote vectors and matrices. For convenience of notation, we use $[i : j]$ to denote the set $\{i, i+1, \ldots, j-1, j\}$ where $i, j \in \mathbb{Z}$. Let $\mathsf{Uni}(\mathcal{M})$ denote the uniform distribution over the set $\mathcal{M}$. 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 all $m \times n$ matrices with rank r.

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 $\mathrm{rk}\,(\boldsymbol{H}[t])$ 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 [17] 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,

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

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 [17]).

Let the message set of a code be $\mathcal{M} = \{1, \ldots, K\}$. A length n block code is defined by a pair of mappings (ψ, ϕ), where $\psi : \mathcal{M} \to \mathcal{X}^n$ is the encoder mapping, and $\phi : \mathcal{Y}^n \to \mathcal{M} \cup \{0\}$ is the decoder mapping, where the output 0 indicates a decoding error. When this code is used on a PAVC and the state sequence is $\boldsymbol{s}$, the error probability for message i is

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

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

$$\bar{e}_d(\boldsymbol{s}) \triangleq \frac{1}{K}\sum_{i=1}^{K} e_d(i,\boldsymbol{s}). \tag{3}$$

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 a length n block code (ψ, ϕ) with*

$$\frac{1}{n}\log K > \mathfrak{R} - \delta, \qquad \textit{and}$$
$$\max_{P_{\boldsymbol{S}|\mathfrak{q}(\boldsymbol{S})}} \mathbb{E}\left[\bar{e}_d(\boldsymbol{S})\right] \triangleq$$
$$\max_{P_{\boldsymbol{S}|\mathfrak{q}(\boldsymbol{S})}} \sum_{\boldsymbol{s}} \bar{e}_d(\boldsymbol{s}) P_{\boldsymbol{S}|\mathfrak{q}(\boldsymbol{S})}\left(\boldsymbol{s}|\mathfrak{q}(\boldsymbol{s})\right) P_{Q^n}\left(\mathfrak{q}(\boldsymbol{s})\right) \le \epsilon,$$

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 $C_{\mathsf{pavc}}^{\mathsf{d,a}}$ (where superscript "a" denotes the average error probability criterion in (3) *and "d" denotes determinist code).*

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

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 receiver.

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 $\mathcal{M}$. Let us define

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

When this code is used on a PAVC and the state sequence is $\boldsymbol{s}$, the error probability for message i is

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

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

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

Similar to Definition 1, we define the capacity $C_{\mathsf{pavc}}^{\mathsf{r,a}}$ by replacing the function $\bar{e}_d(\boldsymbol{s})$ with $\bar{e}_r(\boldsymbol{s})$.

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 $U:\ \mathcal{X}\times\mathcal{Q}\to\mathcal{S}$, and for every x, x', and y we have*

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

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

The following two theorems, which are proved using techniques similar to those in [18], [19], characterize the capacity of the PAVC for the average error criterion. Due to space limit, we refer the reader to [20] for the proof of Theorem 1 and 2.

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

$$C_{\mathsf{pavc}}^{\mathsf{d,a}} = \max_{P_X}\min_{P_{S|\mathfrak{q}(S)}} I(P_X,\bar{W}_S) = \min_{P_{S|\mathfrak{q}(S)}}\max_{P_X} I(P_X,\bar{W}_S), \tag{4}$$

where

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

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

Theorem 2. *The* randomized *code capacity of the PAVC, denoted by $C_{\mathsf{pavc}}^{\mathsf{r,a}}$, is given by* (4).

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

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 whose deterministic code capacity, as stated in Theorem 1, can be characterized as follows.

Corollary 1. *The deterministic code capacity of the channel* (1) *is non-zero and is given by*

$$C = \max_{P_{\boldsymbol{X}}}\min_{P_{\boldsymbol{H}|\,\mathrm{rk}(\boldsymbol{H})}} I(\boldsymbol{X};\boldsymbol{Y}) = \min_{P_{\boldsymbol{H}|\,\mathrm{rk}(\boldsymbol{H})}}\max_{P_{\boldsymbol{X}}} I(\boldsymbol{X};\boldsymbol{Y}), \tag{5}$$

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 2, 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* (5).

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 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* (5). *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 [10, 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

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

Since $\mathbb{E}[\mathrm{rk}\,(\boldsymbol{H}[t])] = \mathbb{E}[R] > 0$, $\mathbb{P}[\mathrm{rk}\,(\boldsymbol{G}[t]\boldsymbol{H}[t]) = 0] < 1$, because otherwise $\boldsymbol{H}[t]$ is the zero matrix with probability one, a contradiction to the assumption that $\mathbb{E}[R] > 0$. 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*[1].

[1]The result in Lemma 2 was also claimed in [14] without any proof (see also [15, Theorem 5]).

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 $\boldsymbol{g}' = \boldsymbol{U}\boldsymbol{g}\boldsymbol{V}$ 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 $\boldsymbol{g}' = \boldsymbol{U}\boldsymbol{g}\boldsymbol{V}$.

Fix two decompositions $\boldsymbol{g} = \boldsymbol{b}\boldsymbol{c}$ and $\boldsymbol{g}' = \boldsymbol{b}'\boldsymbol{c}'$ with $\mathrm{rk}\,(\boldsymbol{b}) = \mathrm{rk}\,(\boldsymbol{b}') = \mathrm{rk}\,(\boldsymbol{g})$, 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 $\boldsymbol{U}\boldsymbol{b} = \boldsymbol{b}'$ and $\boldsymbol{c}\boldsymbol{V} = \boldsymbol{c}'$. Hence, $\boldsymbol{g}' = \boldsymbol{U}\boldsymbol{g}\boldsymbol{V}$. ∎

Lemma 3. *In the capacity expression* (5)*, 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 (5). 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 3. *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 3 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 result.

Theorem 4. *Subspace coding [4] 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 3 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. Indeed, to achieve the optimal communication rate, the u.g.r. distribution over the input symbols (input matrices $\boldsymbol{X}$) is sufficient.

C. An Alternative Proof of Theorem 3

Here, we present an alternative proof for Theorem 3 that is not based on the derived PAVC results[2].

We may consider the non-coherent network coding channel introduced in §II-B as a matrix channel where the transfer matrix is chosen arbitrarily by an adversary, in such a way that the constraint on the rank distribution is preserved. Let this channel is denoted by $\mathrm{Ch}_{\mathsf{adv}}$. Then we define $\mathrm{Ch}_{\mathsf{ugr}}$ to denote the discrete memory-less channel defined by $\boldsymbol{Y}[t]=\boldsymbol{X}[t]\boldsymbol{H}[t]$ with a u.g.r. transfer matrix [14].

Now we can state the following facts. First, for $\mathrm{Ch}_{\mathsf{adv}}$, the adversary can always emulate $\mathrm{Ch}_{\mathsf{ugr}}$ by choosing the transfer matrix according to the u.g.r. distribution of $\mathrm{Ch}_{\mathsf{ugr}}$. Secondly, by using the randomization described in Lemma 2, we may degrade $\mathrm{Ch}_{\mathsf{adv}}$ to $\mathrm{Ch}_{\mathsf{ugr}}$ for whatever strategy the adversary has been performing. Note that we need to apply a similar technique in Lemma 2 to show that the degradation is memoryless. Thus, from the above argument, for every definition of the capacity of $\mathrm{Ch}_{\mathsf{adv}}$, it must be equal to the capacity of $\mathrm{Ch}_{\mathsf{ugr}}$ with the same rank distribution.

Although the above argument provides a simpler proof for the capacity of $\mathrm{Ch}_{\mathsf{adv}}$, we believe that our AVC approach gives a different insight and enables us to study more general models without the knowledge of the full probabilistic characterization of the rank distribution.

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 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 a complex, dynamically changing network 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). The work of S. Yang was partially supported by the National Basic Research Program of China through Grant 2011CBA00300, 2011CBA00302.

[2]The idea of this proof is due to one of the anonymous reviewers where hereby we acknowledge him/her.

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, [Online]. Available: 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, 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*, 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, 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.

[11] 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.

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

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

[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] R. W. Nobrega, D. Silva, and B. F. Uchoa-Filho, "On the Capacity of Multiplicative Finite-Field Matrix Channels," 2011, [Online]. Available: http://arxiv.org/abs/1105.6115.

[16] 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.

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

[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.

[20] M. Jafari Siavoshani, S. Yang, and R. W. Yeung, "Non-coherent network coding: an arbitrarily varying channel approach," *EPFL Technical Report*. [Online]. Available: http://infoscience.epfl.ch/record/174714.