On the Capacity of Non-Coherent Network Coding

M. Jafari Siavoshani, S. Mohajer, C. Fragouli, S N. Diggavi

Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland
University of California, Los Angeles (UCLA), USA

Abstract

We consider the problem of multicasting information from a source to a set of receivers over a network where intermediate network nodes perform randomized network coding operations on the source packets. We propose a channel model for the non-coherent network coding introduced by Koetter and Kschischang in [6], that captures the essence of such a network operation, and calculate the capacity as a function of network parameters. We prove that use of subspace coding is optimal, and show that, in some cases, the capacity-achieving distribution uses subspaces of several dimensions, where the employed dimensions depend on the packet length. This model and the results also allow us to give guidelines on when subspace coding is beneficial for the proposed model and by how much, in comparison to a coding vector approach, from a capacity viewpoint. We extend our results to the case of multiple source multicast that creates a virtual multiple access channel[1].

Keywords

Network coding, non-coherent communication, subspace coding, channel capacity, multi-source multicast, randomized network coding.

I. Introduction

The network coding techniques for information transmission in networks introduced in [1] have attracted significant interest in the literature, both because of posing theoretically interesting questions,

S N. Diggavi was at EPFL and now is at the University of California, Los Angeles (UCLA). The work of M. Jafari Siavoshani and C. Fragouli was supported in part by the Swiss National Science Foundation through the grant # PP002-110483. The work of S. Mohajer and C. Fragouli was supported in part by the ERC Starting Investigator grant # 240317.

[1]Some parts of the work in this paper was presented at ISIT'08, ISIT'09, and ITW'09.

as well as because of potential impact in applications. The first fundamental result proved in network coding, and perhaps still the most useful from a practical point of view today, is that, using linear network coding [2], [3], one can achieve rates up to the common min-cut value when multicasting to $N_r \geq 1$ receivers. In general this may require operations over a field of size approximately $\sqrt{N_r}$, which translates to communication using packets of length $\frac{1}{2}\log N_r$ bits [4].

However, this result assumes that the receivers know perfectly the operations that the network nodes perform. In large dynamically changing networks, collecting network information comes at a cost, as it consumes bandwidth that could instead have been used for information transfer. In practical networks, where such deterministic knowledge is not sustainable, the most popular approach is to perform randomized network coding [5] and to append coding vectors at the headers of the packets to keep track of the linear combinations of the source packets they contain (see, *e.g.*, [12]). The coding vectors have an overhead of $h \log N_r$ bits, where h is the total number of packets to be linearly combined. This results in a loss of information rate that can be significant with respect to the min-cut value. In particular, in wireless networks such as sensor networks where communication is restricted to short packet lengths, the coding vector overhead can be a significant fraction of the overall packet length [27], [13].

Use of coding vectors is akin to use of training symbols to learn the transformation induced by a network. A different approach is to assume a non-coherent scenario for communication, as proposed in [6], 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 for creating end-to-end systems completely oblivious to the network state. Several natural questions arise considering this non-coherent framework: (i) what are the fundamental limits on the rates that can be achieved in a network where the intermediate node operations are unknown, (ii) how can they be achieved, and (iii) how do they compare to the coherent case.

In this work we address such questions for two different cases. First, we consider the scenario where a single source aims to transmit information to one or multiple receiver(s) over a network under the non-coherence assumption using fixed packet length. Because network nodes only perform linear operations, the overall network behavior from the source(s) to a receiver can be represented as a matrix multiplication of the sent source packets. We consider operation in time-slots, and assume that the channel transfer matrices are distributed uniformly at random and i.i.d. over different time-slots. Under this probabilistic model, we characterize the asymptotic capacity behavior of the introduced channel and show that using *subspace coding* we can achieve the optimal performance. We extend our model for the case of multiple sources and characterize the asymptotic behavior of the optimal rate region for the case of two sources. We

believe that this result can be extended to the case of more than two sources using the same method that is applied in §V. For the multi-source case we prove as well that encoding information using subspaces is sufficient to achieve the optimal rate region.

The idea of non-coherent modeling for randomized network coding was first proposed in the seminal work by Koetter and Kschischang in [6]. In that work, the authors focused on algebraic subspace code constructions over a Grassmannian. Independently and in parallel to our work in [9], Montanari *et al.* [14] introduced a different 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 subsequent time slots, but instead, allows the packets block length (in this paper terminology; packet length T) to increases to infinity, with the result that the overhead of coding vectors becomes negligible, very fast.

Silva *et al.* [16] independently and subsequent to our works in [9] and [10], also considered a probabilistic model for non-coherent network coding, which is an extension of the model introduced in [14] over multiple time-slots. In their model the transfer matrix is constrained to be square as well as full rank. This is in contrast to our model, where the transfer matrix can have arbitrary dimensions, and the elements of the transfer matrix are chosen uniformly at random, with the result that the transfer matrix itself may not have full rank (this becomes more pronounced for small matrices). Moreover, we extend our work to multiple source multicast, which corresponds to a virtual non-coherent multiple access channel. Our results coincide for the case of a single source, when the packet length and the finite field of operations are allowed to grow sufficiently large. Another difference is that the work in [16] focuses on additive error with constant dimensions; in contrast, we focus on packet erasures.

An interpretation of our results is that it is the finite field analog of the Grassmannian packing result for non-coherent MIMO channels as studied in the well known work in [19]. In particular, we show that for the non-coherent model over finite fields, the capacity critically depends on the relationship between the "coherence time" (or packet length T in our model) and the min-cut of the network. In fact the number of active subspace dimensions depend on this relationship; departing from the non-coherent MIMO analogy of [19].

The paper is organized as follows. We define our notation and channel model in §II; we state and discuss our main results in §III; we prove the capacity results for the single and multiple sources in sections §IV and §V respectively; and conclude the paper in §VI.

All the missing proofs for lemmas, theorems, and etc., are given in Appendix A unless otherwise stated.

II. Channel Model and Notation

A. Notation

We here introduce the notation and definitions we use in the following sections. Let $q \geq 2$ be a power of a prime. In this paper, all vectors and matrices have elements in a finite field $\mathbb{F}_q$. We use $\mathbb{F}_q^{n\times m}$ to denote the set of all $n \times m$ matrices over $\mathbb{F}_q$, and $\mathbb{F}_q^T$ to denote the set of all row vectors of length T. The set $\mathbb{F}_q^T$ forms a T-dimensional vector space over the field $\mathbb{F}_q$.

Throughout the paper, we use capital letters, *e.g.*, X, to denote random objects, including random variables, random matrices, or random subspaces, and corresponding lower-case letters, *e.g.*, x to denote their realizations. For example, we denote by $\amalg$ a "random subspace" which takes as values the subspaces in a vector space according to some distribution, and by π a specific realization. Also, bold capital letters, *e.g.*, $\mathbf{A}$, are reserved for deterministic matrices and bold lower-case letters, *e.g.*, $\mathbf{v}$, are used for deterministic vectors.

For subspaces π_1 and π_2, $\pi_1 \sqsubseteq \pi_2$ denotes that π_1 is a subspace of π_2. Recall that for two subspaces π_1 and π_2, $\pi_1 \cap \pi_2$ is the intersection of these subspaces which itself is a subspace. We use $\pi_1 + \pi_2$ to denote the smallest subspace that contains both π_1 and π_2, namely,

$$\pi_1 + \pi_2 = \left\{\mathbf{v}_1 + \mathbf{v}_2 | \mathbf{v}_1 \in \pi_1, \mathbf{v}_2 \in \pi_2\right\}.$$

It is well known that

$$\dim(\pi_1 + \pi_2) = \dim(\pi_1) + \dim(\pi_2) - \dim(\pi_1 \cap \pi_2).$$

For a set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ we denote their linear span by $\langle \mathbf{v}_1, \ldots, \mathbf{v}_k \rangle$. For a matrix $\mathbf{X}$, $\langle \mathbf{X} \rangle$ is the subspace spanned by the rows of $\mathbf{X}$ and $\langle \mathbf{X} \rangle_c$ is the subspace spanned by the columns of $\mathbf{X}$. We then have $\mathrm{rank}(\mathbf{X}) = \dim(\langle \mathbf{X} \rangle) = \dim(\langle \mathbf{X} \rangle_c)$.

We use the calligraphic symbols, *i.e.*, $\mathcal{X}$ or $\mathcal{Y}$ to denote a set of matrices. To denote a set of subspaces we use the same calligraphic symbols but with a "$\sim$", *i.e.*, $\widetilde{\mathcal{X}}$ or $\widetilde{\mathcal{Y}}$.

We use the symbols "$\succ$" and "$\prec$" to denote the element-wise inequality between vectors and matrices of the same size.

For two real valued functions $f(x)$ and $g(x)$ of x, we use $f(x) \doteq g(x)$ to denote that[2]

$$\lim_{x\to\infty} \frac{\log f(x)}{\log g(x)} \to 1.$$

[2]One has to specify the growing variable whenever "$\doteq$" is used for multi-variate functions. However, since in this work the growing variable is always q, the field size, we will not repeat it for sake of brevity.

Note that the definition of "$\doteq$" is different from the more standard definition which is $\lim_{x\to\infty} \frac{1}{x} \log \frac{f(x)}{g(x)} \to 0$. We also use a similar definition for $f \dot{\leq} g$ to denote that

$$\lim_{x\to\infty} \frac{\log f(x)}{\log g(x)} \to c \leq 1,$$

where c is a constant.

We use the big-O notation which is defined as follows. Let $f(x)$ and $g(x)$ be two functions defined on some subset of the real numbers. We write $f(x) = O\left(g(x)\right)$ as $x \to \infty$, if there exists a positive real number M and a real number x_0 such that $|f(x)| \leq M|g(x)|$ for all $x > x_0$. For the little o notation we use the following definition. We write $f(x) = o(g(x))$ as $x \to \infty$, if for all $\epsilon > 0$ there exists a real number x_0 such that $|f(x)| \leq \epsilon \cdot |g(x)|$ for all $x > x_0$. We use also the big-Ω notation which is defined as follows. We write $f(x) = \Omega\left(g(x)\right)$ as $x \to \infty$, if we have $g(x) = O\left(f(x)\right)$ as $x \to \infty$. Finally, we use the big-Θ notation to denote that a function is bounded both above and below by another function asymptotically. Formally, we write $f(x) = \Theta\left(g(x)\right)$ as $x \to \infty$, if and only if we have $f(x) = O\left(g(x)\right)$ and $f(x) = \Omega\left(g(x)\right)$ as $x \to \infty$.

***Definition** 1 (Grassmannian and Gaussian coefficient [22], [25]):* The Grassmannian $\mathrm{Gr}(T,d)_q$ is the set of all d-dimensional subspaces of the T-dimensional space over a finite field $\mathbb{F}_q$, namely,

$$\mathrm{Gr}(T,d)_q \triangleq \{\pi \sqsubseteq \mathbb{F}_q^T : \dim(\pi) = d\}.$$

The cardinality of $\mathrm{Gr}(T,d)_q$ is the Gaussian coefficient, namely,

$$\begin{bmatrix} T \\ d \end{bmatrix}_q \triangleq |\mathrm{Gr}(T,d)_q| = \frac{(q^T-1)\cdots(q^{T-d+1}-1)}{(q^d-1)\cdots(q-1)}. \tag{1}$$

***Definition** 2 (The set $\mathrm{Sp}(T,m)_q$):* We define $\mathrm{Sp}(T,m)_q$ to be the set (sphere) of all subspaces of dimension at most m in the T-dimensional space $\mathbb{F}_q^T$, namely

$$\mathrm{Sp}(T,m)_q \triangleq \bigcup_{d=0}^{\min[m,T]} \mathrm{Gr}(T,d)_q = \{\pi \sqsubseteq \mathbb{F}_q^T : \dim(\pi) \leq \min[m,T]\}.$$

The cardinality of $\mathrm{Sp}(T,m)_q$ equals

$$\mathcal{S}(T,m)_q \triangleq |\mathrm{Sp}(T,m)_q| = \sum_{d=0}^{\min[m,T]} |\mathrm{Gr}(T,d)_q|.$$

***Definition** 3 (The number $\psi(T,n,\pi_d)_q$):* We denote by $\psi(T,n,\pi_d)_q$ the number of different $n \times T$ matrices with elements from a field $\mathbb{F}_q$, such that their rows span a specific subspace $\pi_d \sqsubseteq \mathbb{F}_q^T$ of dimension $0 \leq d \leq \min[n,T]$.

For simplicity, in the rest of the paper we will drop the subscript q in the previous definitions whenever it is obvious from the context.

B. Preliminary Lemmas

We here state some preliminary lemmas related to the definitions introduced in §II-A.

Existing bounds in the literature allow to approximate the Gaussian number, for example, we have from [6, Lemma 4] that [23, Section III]

$$q^{d(T-d)} < \begin{bmatrix} T \\ d \end{bmatrix} < \frac{q^{d(T-d)}}{\prod_{j=1}^{\infty}(1-q^{-j})} < 4q^{d(T-d)}, \quad \forall d : 0 < d < T. \tag{2}$$

Using Definition 1 and (2) we have Lemma 1.

***Lemma** 1:* For large q we can approximate the Gaussian number as follows

$$\begin{bmatrix} T \\ d \end{bmatrix} = q^{d(T-d)}(1 + O(q^{-1})) \doteq q^{d(T-d)}.$$

***Lemma** 2:* For $\psi(T, n, \pi_d)$ given in Definition 3, we have that [26]

$$\psi(T, n, \pi_d) = \prod_{i=0}^{d-1}(q^n - q^i) = q^{\binom{d}{2}} \prod_{i=0}^{d-1}(q^{n-i} - 1),$$

i.e., it does not depend on T.

Since $\psi(T, n, \pi_d)$ does not depend on T, and only depends on π_d through its dimension, as a shorthand notation we will also use $\psi(n, d)$ instead of $\psi(T, n, \pi_d)$, where $d = \dim(\pi_d)$.

Using Lemma 2 the following lower and upper bounds are straightforward

$$(1 - dq^{-n+d-1}) < \left(1 - \sum_{i=0}^{d-1} q^{-n+i}\right) < \frac{\psi(n, d)}{q^{nd}} < 1, \tag{3}$$

which imply Lemma 3 (see also [23]).

***Lemma** 3:* For large values of q the following approximation holds

$$\psi(n, d) = q^{nd}(1 + O(q^{-1})) \doteq q^{nd}.$$

It is also worthwhile to mention that $\psi(n, d)\begin{bmatrix} T \\ d \end{bmatrix}$ is the number of $n \times T$ matrices of rank d. We can count all the $n \times T$ matrices through the following Lemma 4, (also see [22], [25], and [26, Corollary 5]).

***Lemma** 4:* For every $n > 0$ and $T > 0$ we can write

$$\sum_{d=0}^{\min[n,T]} \psi(n, d) \begin{bmatrix} T \\ d \end{bmatrix} = q^{nT},$$

where $\psi(n, 0) = 1$.

C. The Non-Coherent Finite Field Channel Model

We consider a network where nodes perform random linear network coding over a finite field $\mathbb{F}_q$. We are interested in the maximum information rate at which a single (or multiple) source(s) can successfully communicate over such a network when neither the transmitter nor the receiver(s) have any channel state information (CSI). For simplicity, we will present the channel model and our analysis for the case of a single receiver; the extension to multiple receivers is straightforward, as we also discuss in the results section.

We assume that time is slotted and the channel is block time-varying. For the single source communication, at time slot t, the receiver observes

$$Y[t] = G[t]X[t], \tag{4}$$

where $X[t] \in \mathbb{F}_q^{m\times T}$, $G[t] \in \mathbb{F}_q^{n\times m}$, and $Y[t] \in \mathbb{F}_q^{n\times T}$. At each time-slot, the receiver receives n packets of length T (captured by the rows of matrix $Y[t]$) that are random linear combinations of the m packets injected by the source (captured by the rows of matrix $X[t]$). In our model, the packet length T can be interpreted as the coherence time of the channel, during which the transfer matrix remains constant. Each element of the transfer matrix $G[t]$ is chosen uniformly at random from $\mathbb{F}_q$, changes independently from time slot to time slot, and is unknown to both the source and the receiver. In other words, the channel transfer matrix is chosen uniformly at random from all possible matrices in $\mathbb{F}_q^{n\times m}$ and has i.i.d. distribution over different blocks. In general, the topology of the network may impose some constraints on the transfer matrix $G[t]$ (for example, some entries might be zero, see [3], [8], [20], [21]). However, we believe that this is a reasonable general model, especially for large-scale dynamically-changing networks where apart from random coefficients there exist many other sources of randomness. Formally, we define the non-coherent matrix channel as follows.

***Definition** 4 (Non-coherent matrix channel Ch_m):* This is defined to be the matrix channel Ch_m : $\mathcal{X} \to \mathcal{Y}$ described by (4) with the assumption that $G[t]$ is i.i.d. and uniformly distributed over all matrices $\mathbb{F}_q^{n\times m}$. It is a discrete memoryless channel with input alphabet $\mathcal{X} \triangleq \mathbb{F}_q^{m\times T}$ and output alphabet $\mathcal{Y} \triangleq \mathbb{F}_q^{n\times T}$.

The capacity of the channel Ch_m is given by

$$C_m = \max_{P_X(x)} I(X;Y), \tag{5}$$

where $P_X(x)$ is the input distribution. To achieve the capacity a coding scheme may employ the channel given in (4) multiple times, and a codeword is a sequence of input matrices from $\mathcal{X}$. For a coding strategy

that induces an input distribution $P_X(x)$, the achievable rate is

$$R = I(X;Y).$$

Now we define a non-coherent subspace channel Ch_s which takes as an input a subspace and outputs another subspace. Then, in Theorem 1 we will show that the two channels Ch_m and Ch_s are equivalent from the point of view of calculating the mutual information between their inputs and their outputs.

Definition *5 (Non-coherent subspace channel* Ch_s*):* This is defined to be the channel $\mathrm{Ch}_s : \widetilde{\mathcal{X}} \to \widetilde{\mathcal{Y}}$ with input alphabet $\widetilde{\mathcal{X}} = \mathrm{Sp}(T,m)$ and output alphabet $\widetilde{\mathcal{Y}} = \mathrm{Sp}(T,n)$ and transition probability

$$P_{\Pi_Y|\Pi_X}(\pi_y|\pi_x) \triangleq \begin{cases} \psi(T,n,\pi_y)q^{-n\dim(\pi_x)} & \pi_y \sqsubseteq \pi_x, \\ 0 & \text{otherwise}, \end{cases} \tag{6}$$

where Π_X and Π_Y are the input and output variables of the channel Ch_s.
The capacity of the channel Ch_s is given by

$$C_s = \max_{P_{\Pi_X}(\pi_x)} I(\Pi_X;\Pi_Y),$$

where $P_{\Pi_X}(\pi_x)$ is the input distribution defined over the set of subspaces $\widetilde{\mathcal{X}}$.

We next consider a multiple sources scenario, and the multiple access channel corresponding to (4). In this case, we have

$$Y[t] = \sum_{i=1}^{N_s} G_i[t]X_i[t], \tag{7}$$

where N_s is the number of sources, and each source i inserts m_i packets to the network. Thus, $X_i[t] \in \mathbb{F}_q^{m_i \times T}$, $G_i[t] \in \mathbb{F}_q^{n \times m_i}$ and $Y[t] \in \mathbb{F}_q^{n \times T}$. We can also collect all $G_i[t]$ in an $n \times \sum_{i=1}^{N_s} m_i$ matrix $G_{MAC}[t]$ and all $X_i[t]$ in an $\sum_{i=1}^{N_s} m_i \times T$ matrix $X_{MAC}[t]$ as following

$$X_{MAC}[t] = \begin{bmatrix} X_1[t] \\ \vdots \\ X_{N_s}[t] \end{bmatrix}, \quad \text{and} \quad G_{MAC}[t] = \begin{bmatrix} G_1[t] & \cdots & G_{N_s}[t] \end{bmatrix},$$

so we can rewrite (7) as

$$Y[t] = G_{MAC}[t]X_{MAC}[t].$$

Each source i then controls m_i rows of the matrix $X_{MAC}[t]$. Again we assume that each entry of the matrices $G_i[t]$ is chosen i.i.d. and uniformly at random from the field $\mathbb{F}_q$ for all source nodes and all time instances.

Definition *6 (The non-coherent multiple access matrix channel* $\mathrm{Ch}_{m\text{-}MAC}$*):* This is defined to be the channel $\mathrm{Ch}_{m\text{-}MAC} : \mathcal{X}_1 \times \cdots \times \mathcal{X}_{N_s} \to \mathcal{Y}$ described in (7), with the assumption that $G_i[t]$, $i = 1, \ldots, N_s$,

are i.i.d. and uniformly distributed over all matrices $\mathbb{F}_q^{n\times m_i}$, $i=1,\ldots,N_s$. It forms a discrete memoryless MAC with input alphabets $\mathcal{X}_i \triangleq \mathbb{F}_q^{m_i\times T}$, $i=1,\ldots,N_s$, and output alphabet $\mathcal{Y}\triangleq\mathbb{F}_q^{n\times T}$.
It is well known [15] that the rate region of any multiple access channel including $\mathrm{Ch}_{m\text{-}MAC}$ is given by the closure of the convex hull of the rate vectors satisfying

$$R_S \le I(X_S;Y|X_{S^c}) \quad \text{for all } S\subseteq\{1,\ldots,N_s\},$$

for some product distribution $P_{X_1}(x_1)\cdots P_{X_{N_s}}(x_{N_s})$. Note that $R_S=\sum_{i\in S}R_i$ where R_i is the transmission rate of the ith source, $X_S=\{X_i : i\in S\}$ and S^c is the complement set of S.

As before, we define a non-coherent subspace version[3] of the matrix multiple access channel and in Theorem 6 we show that from the point of view of rate region these two channels are equivalent.

***Definition** 7 (Non-coherent subspace multiple access channel $\mathrm{Ch}_{s\text{-}MAC}$):* This is defined to be the channel $\mathrm{Ch}_{s\text{-}MAC}:\widetilde{\mathcal{X}}_1\times\widetilde{\mathcal{X}}_2\to\widetilde{\mathcal{Y}}$ with input alphabets $\widetilde{\mathcal{X}}_i=\mathrm{Sp}(T,m_i)$, $i=1,2$, output alphabet $\widetilde{\mathcal{Y}}=\mathrm{Sp}(T,n)$ and transition probability

$$\Pr(\Pi_Y=\pi_y|\Pi_{X_1}=\pi_1,\Pi_{X_2}=\pi_2)=\begin{cases}\psi(T,n,\pi_y)q^{-n\dim(\pi_1+\pi_2)} & \pi_y\sqsubseteq\pi_1+\pi_2,\\ 0 & \text{otherwise},\end{cases} \tag{8}$$

where Π_{X_1} and Π_{X_2} are the input and Π_Y is the output variables of the channel $\mathrm{Ch}_{s\text{-}MAC}$.

III. Main Results

A. Single Source

Our main results, Theorem 2 and Theorem 3, characterize the capacity for non-coherent network coding for the model given in (4). We show that the capacity is achieved through subspace coding, where the information is communicated from the source to the receivers through the choice of subspaces. Formally, we have the following results.

***Theorem** 1:* The matrix channel $\mathrm{Ch}_m:\mathcal{X}\to\mathcal{Y}$ defined in Definition 4 and the subspace channel $\mathrm{Ch}_s:\widetilde{\mathcal{X}}\to\widetilde{\mathcal{Y}}$ defined in Definition 5 are equivalent in terms of evaluating the mutual information between the input and output. More precisely, for every input distribution for the channel Ch_s there is an input distribution for the channel Ch_m such that $I(X;Y)=I(\Pi_X;\Pi_Y)$ and vice versa. As a result, these channels have the same capacity $C_m=C_s$.

For the proof of Theorem 1 refer to Appendix A and for more discussion refer to §IV-A.

[3]For simplicity, we restrict this definition to only two source nodes. However, generalization to N_s sources is straightforward.

***Theorem** 2:* For the channel $\mathrm{Ch}_m : \mathcal{X} \to \mathcal{Y}$ defined in Definition 4, the capacity for large q is as follows

$$C_m = i^*(T - i^*) \log_2 q + o(1), \tag{9}$$

where $i^* = \min\left[m, n, \lfloor T/2 \rfloor\right]$.

Theorem 2 is proved in §IV-B. The result of Theorem 2 is for the large alphabet regime[4]. The following result, Theorem 3, is valid for a finite field size, and therefore is a non-asymptotic result.

***Theorem** 3:* Consider the channel $\mathrm{Ch}_m : \mathcal{X} \to \mathcal{Y}$ defined in Definition 4. There exists a finite number q_0 such that for $q > q_0$ the optimal input distribution is nonzero only for matrices of rank in the set

$$\mathcal{A} = \left\{\min\left[(T-n)^+, m, n, T\right], \ldots, \min\left[m, n, T\right]\right\}. \tag{10}$$

Moreover, for $q > q_0$ the optimal input distribution is uniform over all matrices X of the same rank, and the probability of transmitting a matrix X of rank i equals

$$\alpha_i^* = 2^{-C_m} q^{i(T-i)} \left[1 + o(1)\right], \quad \forall i \in \mathcal{A}, \tag{11}$$

where $\mathbb{P}\left[\mathrm{rank}(X) = i\right] = \alpha_i^*$.

The proof of Theorem 3 is presented in §IV-C and §IV-D, and uses standard techniques from convex optimization, as well as large field size approximations. Note that, for receivers with the same channel parameters (*i.e.*, values of n, m and T) the same coding scheme at the source simultaneously achieves the capacity for all of them. That is, each receiver is able to successfully decode.

The result of Theorem 3 for the active set of input dimensions is not asymptotic in q. However, it is not easy to find analytically the minimum value of q_0 such that the theorem statement holds for all $q > q_0$. Theorem 4 demonstrates how we can analytically characterize q_0 given in Theorem 3 for the case $T > n + \min[m, n]$. The proof of Theorem 4 is presented in §IV-E.

***Theorem** 4:* If we have $T > n + \min[m, n]$ then the capacity of Ch_m for $q \geq q_0$ is given by

$$\begin{aligned} C_m &= \sum_{l=0}^{i^*} \psi(n, l) \begin{bmatrix} i^* \\ l \end{bmatrix} q^{-ni^*} \log_2 \left(\frac{\begin{bmatrix} T \\ l \end{bmatrix}}{\begin{bmatrix} i^* \\ l \end{bmatrix}} \right) \\ &= i^*(T - i^*) \log_2 q - 1_{\{n \leq m\}}(T - i^*) \frac{\log_2 q}{q} + q^{-1} + o(q^{-1}), \end{aligned} \tag{12}$$

[4]We gratefully acknowledge the contribution of an anonymous reviewer who gave an alternate proof, which focused on the asymptotic q regime. We have included that proof in §IV-B. Our original proof was based partially on the proof now given for Theorem 3.

where q_0 is the minimum field size that satisfies the set of inequalities

$$\frac{\epsilon_{q_0}(l) - \epsilon_{q_0}(i^*)}{(T-n-i^*)(i^*-l)} \leq \log_2 q_0, \quad \forall l : 0 \leq l \leq (i^*-1),$$

and

$$\frac{\epsilon_{q_0}(l) - \epsilon_{q_0}(i^*)}{i^*(l-i^*)} \leq \log_2 q_0, \quad \forall l : (i^*+1) \leq l \leq m,$$

where $i^* = \min[m,n]$ and

$$\epsilon_q(l) \triangleq \sum_{d_y=0}^{\min[n,l]} \psi(n,d_y) \begin{bmatrix} l \\ d_y \end{bmatrix} q^{-nl} \log_2 \left(\frac{\begin{bmatrix} T \\ d_y \end{bmatrix}}{\begin{bmatrix} i^* \\ d_y \end{bmatrix}} \right) - \min[n,l](T-i^*).$$

The capacity is achieved by sending matrices X such that their rows span different i^*-dimensional subspaces.

Moreover, asymptotically in T, we can show that $q_0^{n-m+1} \geq 5(i^*)^2$ is sufficient for the case $m \leq n$ and $q_0 \geq i^*T$ is sufficient if $m > n$.

Theorems 2 and 3 state that the capacity behaves as $i^*(T-i^*)\log_2 q$, for sufficiently large q. However, numerical simulations indicate a very fast convergence to this value as q increases. Fig. 1 depicts the capacity for small values of q, calculated using the Differential Evolution toolbox for matlab [11]. This shows that the result is relevant at much lower field size than dictated by the formalism of the statement of Theorems 2 and 3.

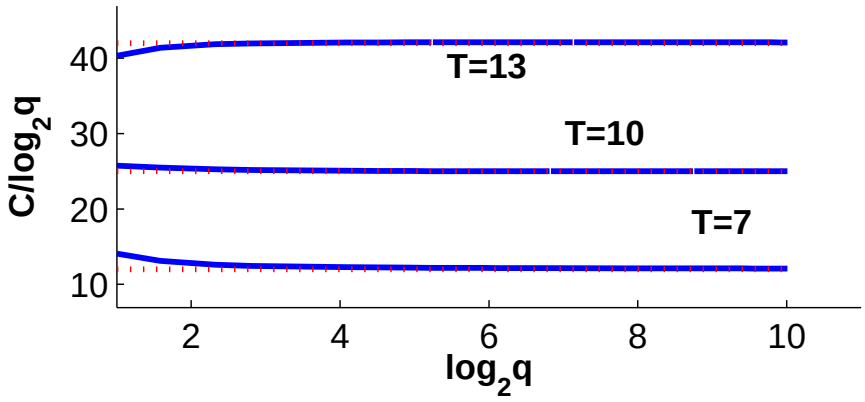

Fig. 1. Numerical calculation of the capacity for small values of q and $m = 11$, $n = 7$. The dotted line depicts $i^*(T-i^*)$.

From Theorem 3, we can derive the following guidelines for non-coherent network code design.

1) Choice of subspaces: The optimal input distribution uses subspaces of a single dimension equal to $\min[m,n]$ for $T \geq \min[m,n]+n$. As T reduces, the set of used subspaces gradually increases, by activating one by one smaller and smaller dimensional subspaces, until, for $T \leq n$, all subspaces are used with equal probability. Fig. 2 pictorially depicts this gradual inclusion of subspaces.

This behavior is different from the result of [16] where the subspaces up to dimension equal to the min-cut appeared in the optimal input distribution. This difference is due to the different channel model used in our work and in [16].

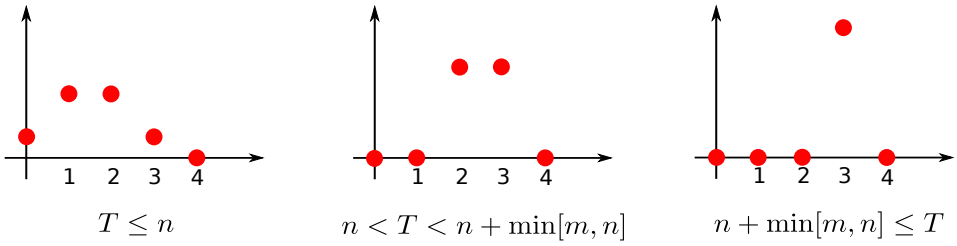

Fig. 2. Probability mass function of the active subspace dimensions for channel parameters $m = 4$, $n = 3$. As it is shown in Theorem 3 there exists three different regimes.

2) Values of m and n: For a given and fixed packet length T, the optimal value of m and n equals $m = n = \lfloor T/2 \rfloor$ (optimality is in the sense of minimum required to achieve the maximum information transfer for this T). For fixed T and m, the optimal value of n equals $n = \min[m, \lfloor T/2 \rfloor]$. For fixed T and n, the optimal value of m equals $m = \min[n, \lfloor T/2 \rfloor]$.

TABLE I

INFORMATION LOSS FROM USING CODING VECTORS WHEN $n = m$.

	$T \leq 2m$	$T > 2m$
$C_m - R_{cv}$	$o(1)$	$o(1) = (i^* - 1)(T - i^*)\frac{\log_2 q}{q} + O(q^{-1})$

3) Subspace coding vs. coding vectors: One of the aims of this work was to find the regimes in which the using of coding vectors [12] is far from optimal. Table I summarizes this difference. As we see from the Table I subspace coding does not offer benefits as compared to the coding vectors approach for large

field size[5].

Table I is calculated as follows. The achievable rate R_{cv} using coding vectors equals

$$R_{cv} \triangleq \mathbb{P}\left[\text{rank}(G_k) = k\right] k(T-k)\log_2 q,$$

where $0 < k \leq m$ is the number of packets in each generation, *i.e.*, each packet includes a coding vector of length k and $T-k$ information symbols. Equivalently, we assume that we use k of the m possible input packets. The matrix G_k is the $k \times k$ sub-matrix of G that is applied over the input packets. To calculate R_{cv}, we know that $\mathbb{P}\left[\text{rank}(G_k) = k\right] = \prod_{i=0}^{k-1}(1-q^{-k+i}) = 1 - q^{-1} + O(q^{-2})$. Assume we choose $k = i^*$ we have $R_{cv} = i^*(T-i^*)\log_2 q - i^*(T-i^*)\frac{\log_2 q}{q}$, where $i^* = \min\left[m, n, \lfloor T/2 \rfloor\right]$. For the capacity C_m we use the large q-regime as considered in Theorem 2 for the case $T \leq 2m$ and the finite q-regime of Theorem 4 for the case $T > 2m$.

B. Extension to the packet erasure networks

After the error free single source scenario, we consider packet erasure networks, and calculate an upper and lower bound on the capacity for this case. The work in [16], which is closest to ours, did not consider erasures but instead constant-dimension additive errors. In practice, which model is more suitable depends on the application: for example, if network coding is deployed at an application layer, then, unless there exist malicious attackers, packet erasures are typically used to abstract both the underlying physical channel errors, as well as packet dropped at queues or lost due to expired timers.

We model the erasures in the network as an end-to-end phenomenon which randomly erases packets according to some probability distribution. Formally, we rewrite the channel defined in (4) as

$$Y[t] = E[t]G[t]X[t], \tag{13}$$

where $E \in \mathbb{F}_q^{m \times m}$ is a diagonal random matrix whose elements on its diagonal are either 1 or 0. We also assume that q is large, and as a result the transfer matrix is full rank with high probability. Moreover, we consider the case where $m > \frac{T}{2}$, i.e. the matrix X is a fat matrix. Recall that we can think of the rows of this matrix as packets send by the source, and the rows of the Y matrix as packets received at the destination.

[5]In the algebraic framework of [6], the lifting construction used coding vectors, and they showed that this construction achieved almost the same rates as optimal algebraic subspace codes. However, we demonstrated in this paper that this phenomenon occurs for longer packet lengths using an information-theoretic framework.

Note that in equation (13) all of the erasure events are captured by the erasure matrix $E[t]$. Moreover, the erasure pattern is important only up to determining the number of packets that the destination receives, since the transfer matrix $G[t]$ is unknown and distributed uniformly at random over all full rank matrices. Thus, we let the number N of received packets (number of non-zero elements on the diagonal of $E[t]$), with $0 \leq N \leq m$, be a random variable with some distribution that depends on the packet erasures in the network. In this case the capacity is

$$C_e = \max_{P_X} I(X; Y, N).$$

We can then use our previous result, Theorem 2, to find an upper and lower bound for the capacity C_e when we have packet erasure in the network, as the following Theorem 5 describes.

***Theorem** 5:* Let the number of received packets at the destination be a random variable N defined over the set of integers $0 \leq N \leq m$. Also, assume that $m \leq \lfloor T/2 \rfloor$. Then for large q, we have the following upper and lower bound for the capacity C_e,

$$\mu_1(T-m)\log_2 q \leq C_e \leq \mu_1 \left(T - \frac{\mu_2}{\mu_1}\right) \log_2 q,$$

where $\mu_1 \triangleq \mathbb{E}_N[N]$ and $\mu_2 \triangleq \mathbb{E}_N\left[N^2\right]$.

Proof: For the proof and more discussion refer to Appendix B. ■

Note that because we do not necessarily employ full-rank matrices X, it is possible that although some packets are erased at the destination, the received packets still span a matrix of the same rank as X; thus erasing packets is not equivalent to erasing dimensions.

C. Multiple Sources

In several practical applications, such as sensor networks, data sources are not necessarily co-located. We thus extend our work to the case where multiple not co-located sources transmit information to a common receiver. In particular, we consider the non-coherent MAC channel introduced in Definition 6, and characterize the capacity region of this network for the case of two sources and packet length $\frac{T}{2} > m_1 + m_2$. We believe that this technique can be extended to more than two sources.

To find the rate region of the matrix multiple access channel $\mathrm{Ch}_{\textit{m-MAC}}$, we first show that the two channels $\mathrm{Ch}_{\textit{m-MAC}}$ and $\mathrm{Ch}_{\textit{s-MAC}}$ are equivalent, as stated in Theorem 6. We then find the rate region of the subspace multiple access channel $\mathrm{Ch}_{\textit{s-MAC}}$ which is stated in Theorem 7. To avoid repetition, we state Theorem 6 without a proof because its proof is very similar to that of Theorem 1.

***Theorem** 6:* The matrix MAC channel $\mathrm{Ch}_{\textit{m-MAC}}$ defined in Definition 6 is equivalent to the subspace MAC channel $\mathrm{Ch}_{\textit{s-MAC}}$ defined in Definition 7 in the sense that the optimal rate region for these two channels is the same.

***Theorem** 7:* The asymptotic (in the field size q) rate region of the MAC channel $\mathrm{Ch}_{\textit{m-MAC}}$ given in Definition 6 for $\frac{T}{2} > m_1 + m_2$ can be calculated as

$$\mathcal{R}^* \triangleq \text{convex hull} \bigcup_{(d_1,d_2)\in\mathcal{D}^*} \mathcal{R}(d_1, d_2),$$

where

$$\mathcal{R}(d_1, d_2) \triangleq \{(R_1, R_2) : R_i \leq R_i(d_1, d_2),\ i = 1, 2\}, \tag{14}$$

$$R_i(d_1, d_2) \triangleq d_i(T - d_1 - d_2) \log_2 q, \qquad i = 1, 2,$$

and

$$\mathcal{D}^* \triangleq \{(d_1, d_2) :\ 0 \leq d_i \leq \min[n, m_i], 0 \leq d_1 + d_2 \leq \min[n, m_1 + m_2]\}.$$

We note that the rate region forms a polytopes that has the following number of corner points (see Corollary 1 in §V)

$$\min\left[m_1, (n - m_2)^+\right] + \min\left[m_2, (n - m_1)^+\right] + 2 - 1_{[n\geq m_1+m_2]}.$$

The rate region $\mathcal{R}^*$ is shown in Fig. 3 for a particular choice of parameters.

The proof of this theorem is provided in §V. We first derive an outer bound by deriving two other bounds: a cooperative bound and a coloring bound. For the coloring bound, we utilize a combinatorial approach to bound the number of *distinguishable* symbol pairs that can be transmitted from the sources to the destination. We then show that a simple scheme that uses coding vectors achieves the outer bound. We thus conclude that, for the case of two sources when $\frac{T}{2} > m_1 + m_2$, use of coding vectors is (asymptotically) optimal.

IV. The Channel Capacity: Single Source Scenario

In this section we will prove Theorem 2, Theorem 3, and Theorem 4.

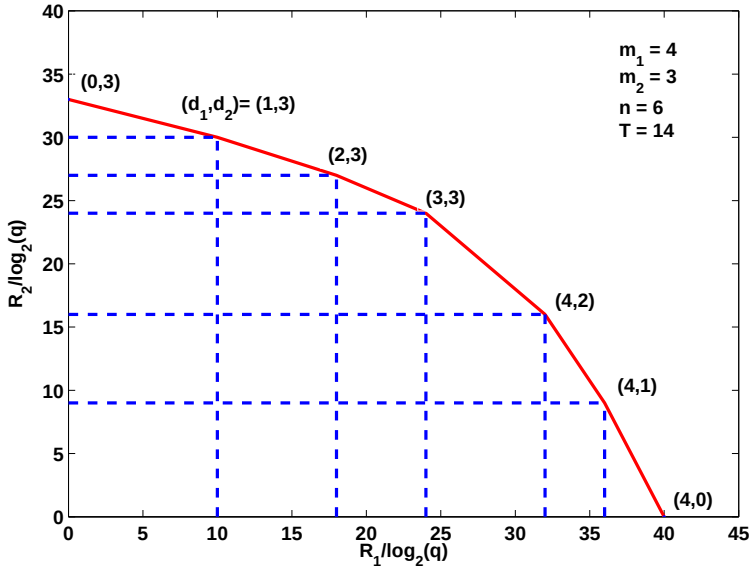

Fig. 3. The MAC region $\mathcal{R}^*$ for parameters $m_1 = 4$, $m_2 = 3$, $n = 3$, $T = 14$.

A. *Equivalence of the Matrix Channel* Ch_m *and the Subspace Channel* Ch_s

For convenience let us rewrite the channel (4) again[6]

$$Y = GX.$$

To find the capacity of the above channel we need to maximize the mutual information between the input and the output of the channel with respect to the input distribution $P_X(x)$. Since the rows of G are chosen independently of each other, assuming that a matrix $X = x$ has been transmitted, we can think of the rows of the received matrix Y as chosen independently from each other, among all the possible vectors in the row span of x. The independence of rows of Y allows us to write the conditional probability of Y given X, referred to as the channel transition probability, as follows

$$P_{Y|X}(y|x) = \begin{cases} q^{-n\dim(\langle x\rangle)} & \langle y\rangle \sqsubseteq \langle x\rangle\,, \\ 0 & \text{otherwise}, \end{cases} \tag{15}$$

where $x \in \mathcal{X} = \mathbb{F}_q^{m\times T}$, and $y \in \mathcal{Y} = \mathbb{F}_q^{n\times T}$.

[6]In the rest of the paper we will omit for convenience the time index t.

The mutual information $I(X;Y)$ between X and Y is a function of $P_X(x)$ and $P_{Y|X}(y|x)$ that can be expressed as

$$I(X;Y) = \sum_{\substack{x\in\mathcal{X},\\ y\in\mathcal{Y}}} P_X(x)P_{Y|X}(y|x)\log_2\left(\frac{P_{Y|X}(y|x)}{P_Y(y)}\right). \tag{16}$$

It is clear from (15) that $P_{Y|X}(y|x_1) = P_{Y|X}(y|x_2)$ for all $x_1, x_2 \in \mathcal{X}$ such that $\langle x_1\rangle = \langle x_2\rangle$ which reveals symmetry for the channel Ch_m. We exploit this symmetry to show that $C_m = C_s$ as it is stated in Theorem 1 and proved in Appendix A.

The proof of Theorem 1 determines how we can map an input distribution of Ch_s to an input distribution for Ch_m that achieves the same mutual information. The input distribution $P_X(x)$ should be chosen such that we have $\sum_{x\in\mathcal{X}:\langle x\rangle=\pi_x} P_X(x) = P_{\Pi_X}(\pi_x)$. One simple way to do this is to put all the probability mass of π_x on one matrix x such that $\langle x\rangle = \pi_x$.

B. Upper and Lower bound for the Capacity of Ch_m

Here, we state the proof of Theorem 2 by giving upper and lower bounds for the capacity that differ in $o(1)$ bits as $q \to \infty$.

Let $C_m(n,m,T)$ denote the capacity of the channel Ch_m. Let $C_{f\text{-}m}(n,m,T)$ denote the capacity of the channel $Y = AX$ where $A \in \mathbb{F}_q^{n\times m}$ is a full-rank matrix chosen uniformly at random among all the full-rank matrices in $\mathbb{F}_q^{n\times m}$. Then, we have the following lemma.

***Lemma** 5:* We can bound $C_m(n,m,T)$ from above and below as follows

$$C_m(h,h,T) \le C_m(n,m,T) \le C_{f\text{-}m}(n,m,T) \le C_{f\text{-}m}(h,h,T),$$

where $h = \min[m,n]$ and $i^* = \min[h, \lfloor T/2\rfloor = \min[m,n,\lfloor T/2\rfloor]$.

Proof: Let $U_{n\times m} \in \mathbb{F}_q^{n\times m}$ denote a generic random matrix chosen uniformly at random and independently from any other variables. Similarly, let $A_{n\times m} \in \mathbb{F}_q^{n\times m}$ denote a generic full-rank matrix chosen uniformly at random among all such full-rank matrices and independent from any other variables. (Note that each new instance of such a matrix in the same equation denotes a different random variable which is independent from the other random variables.)

Since the channel $Y = A_{n\times m}X$ is statistically equivalent to the channel $Y = A_{n\times n}A_{n\times m}A_{m\times m}X$, we have by the data processing inequality that $C_{f\text{-}m}(n,m,T) \le C_{f\text{-}m}(h,h,T)$.

Using the same argument, since the channel $Y = U_{n\times m}X$ is equivalent to the channel $Y = U_{n\times n}A_{n\times m}X$ if $n \ge m$, and is equivalent to the channel $Y = A_{n\times m}U_{m\times m}X$ if $n \le m$ we have $C_m(n,m,T) \le C_{f\text{-}m}(n,m,T)$.

To obtain the lower bound we proceed as follows. Let us choose $X = \left[\begin{smallmatrix} I_h \\ 0 \end{smallmatrix}\right] \overline{X}$ and $\overline{Y} = [I_h \;\; 0]$, where $Y = U_{n\times m}X$. Then we can write

$$\overline{Y} = [I_h \;\; 0]\, U_{n\times m} \begin{bmatrix} I_h \\ 0 \end{bmatrix} \overline{X} = U_{h\times h}\overline{X}.$$

Thus we have $C_m(h,h,T) \leq C_m(n,m,T)$. ■

***Lemma** 6:* For $C_m(n,m,T)$ we have

$$C_m(n,m,T) \leq i^*(T-i^*)\log_2 q + o(1).$$

Proof: By Lemma 5 we have

$$\begin{aligned} C_m(n,m,T) &\leq C_{f\text{-}m}(h,h,T) \\ &\overset{(a)}{=} \log_2 \left(\sum_{i=0}^{h} \begin{bmatrix} T \\ i \end{bmatrix} \right) \\ &\overset{(b)}{=} i^*(T-i^*)\log_2 q + o(1), \end{aligned}$$

where (a) follows from [16, Corollary 2] and (b) follows from Lemma 1. ■

***Lemma** 7:* For $C_m(n,m,T)$ we have

$$C_m(n,m,T) \geq i^*(T-i^*)\log_2 q - o(1).$$

Proof: For every subspace $\Pi \in \mathrm{Gr}(T,i^*)$, let $\mathrm{RREF}(\Pi) \in \mathbb{F}_q^{i^*\times T}$ be a matrix in reduced row echelon form such that $\Pi = \langle \mathrm{RREF}(\Pi) \rangle$. Choose $X = \left[\begin{smallmatrix} I_{i^*} \\ 0 \end{smallmatrix}\right] \times \mathrm{RREF}(\Pi_X) \in \mathbb{F}_q^{m\times T}$, where Π_X is chosen uniformly at random from $\mathrm{Gr}(T,i^*)$. Define the random variable $Q = 1_{\{\mathrm{rank}(Y)=i^*\}}$. Note that $\Pi_Y = \Pi_X$ when $Q=1$. Thus, we have $H(\Pi_Y|\Pi_X, Q=1) = 0$ and $H(\Pi_Y|Q=1) = H(\Pi_X) = \log_2 \left[\begin{smallmatrix} T \\ i^* \end{smallmatrix}\right] \geq i^*(T-i^*)\log_2 q$. Then, it follows that

$$\begin{aligned} C_m(n,m,T) &\overset{(a)}{\geq} I(\Pi_X;\Pi_Y) \\ &= I(\Pi_X;\Pi_Y,Q) \\ &= I(\Pi_X;Q) + I(\Pi_X;\Pi_Y|Q) \\ &\geq \mathbb{P}\left[Q=1\right] I(\Pi_X;\Pi_Y|Q=1) \\ &\geq \mathbb{P}\left[Q=1\right] i^*(T-i^*)\log_2 q, \end{aligned}$$

where (a) follows from Theorem 1. Now, note that we can write

$$\begin{aligned}\mathbb{P}\left[Q=1\right] &= \mathbb{P}\left[\mathrm{rank}(U_{h\times h}X)=i^*\right]\\ &= \mathbb{P}\left[\mathrm{rank}\left(U_{h\times h}\begin{bmatrix} I_{i^*}\\ 0\end{bmatrix}\right)=i^*\right]\\ &= \mathbb{P}\left[\mathrm{rank}(U_{h\times i^*})=i^*\right]\\ &\geq 1-\frac{i^*}{q^{k-i^*+1}}\\ &\geq 1-\frac{i^*}{q},\end{aligned}$$

and thus we obtain the desired result. ■

Combining Lemma 6 and Lemma 7 recovers Theorem 2.

C. The Optimal Solution: General Approach

Generally, we are interested in finding the capacity and input distribution of Ch_m exactly. It is shown in Theorem 1 that instead of the channel Ch_m we can focus on the channel Ch_s. Thus, we are interested in optimizing the following quantity

$$I(\Pi_X;\Pi_Y)=\sum_{\substack{\pi_x\in\widetilde{\mathcal{X}},\\ \pi_y\in\widetilde{\mathcal{Y}}}} P_{\Pi_X}(\pi_x)P_{\Pi_Y|\Pi_X}(\pi_y|\pi_x)\log_2\left(\frac{P_{\Pi_Y|\Pi_X}(\pi_y|\pi_x)}{P_{\Pi_Y}(\pi_y)}\right). \tag{17}$$

Remember that $\widetilde{\mathcal{X}}=\mathrm{Sp}(T,m)$ and $\widetilde{\mathcal{Y}}=\mathrm{Sp}(T,n)$.

The following lemma states that the optimal solution for the channel Ch_s should be uniform over all subspaces with the same dimension, as it is intuitively expected from the symmetry of the channel.

***Lemma** 8:* The input distribution that maximizes $I(\Pi_X;\Pi_Y)$ for Ch_s is the one which is uniform over all subspaces having the same dimension.

Lemma 8 shows that the optimal input distribution can be expressed as

$$\mathbb{P}\left[\Pi_X=\pi_x\right]=\frac{\alpha_{d_x}}{\begin{bmatrix}T\\ d_x\end{bmatrix}}, \tag{18}$$

where $d_x=\dim(\pi_x)$, $\alpha_{d_x}=\mathbb{P}\left[\dim(\Pi_X)=d_x\right]$, and we have $\sum_{d_x=0}^{\min[m,T]}\alpha_{d_x}=1$. We can then simplify $I(\Pi_X;\Pi_Y)$ as stated in the following lemma.

***Lemma** 9:* Assuming an optimal input probability distribution of the form in (18), the mutual information $I(\Pi_X;\Pi_Y)$ can be simplified to

$$\begin{aligned} I(\Pi_X;\Pi_Y) = &- \sum_{d_x=0}^{\min[m,T]} \alpha_{d_x} n d_x \log_2 q \\ &- \sum_{d_x=0}^{\min[m,T]} \alpha_{d_x} q^{-nd_x} \sum_{d_y=0}^{\min[n,d_x]} \psi(n,d_y) \begin{bmatrix} d_x \\ d_y \end{bmatrix} \log_2(f(d_y)), \end{aligned} \tag{19}$$

where

$$f(d_y) \triangleq \frac{P_{\Pi_Y}(\pi_y)}{\psi(n,d_y)} = \frac{1}{\begin{bmatrix} T \\ d_y \end{bmatrix}} \sum_{d_x=d_y}^{\min[m,T]} \begin{bmatrix} d_x \\ d_y \end{bmatrix} q^{-nd_x} \alpha_{d_x}. \tag{20}$$

Lemmas 8 and 9 show that the problem of finding the optimal input distribution for the channel Ch_s is reduced to finding the optimal choice for $\alpha_i,\ i = 0,\ldots,\min[m,T]$. We know that the mutual information is a concave function with respect to $P_{\Pi_X}(\pi_x)$'s. Observation 1 implies that because (18) is a linear transformation from $P_{\Pi_X}(\pi_x)$'s to α_i's, as a result the mutual information $I(\Pi_X;\Pi_Y)$ is also concave with respect to α_i's [18].

***Observation** 1:* Let $g(\mathbf{x})$ be a concave function and let $\mathbf{x} = h(\mathbf{z})$ be a linear transform from $\mathbf{z}$ to $\mathbf{x}$. Then $g(h(\mathbf{z}))$ is also a concave function.

Using Observation 1, we know that the mutual information is a concave function with respect to α_i's. This allows us to use the Kuhn-Tucker theorem [18] to solve the convex optimization problem. According to this theorem, the set of probabilities α_i^*, $0 \le i \le \min[m,T]$, maximize the mutual information if and only if there exists some constant λ such that

$$\begin{cases} \left.\frac{\partial I(\Pi_X;\Pi_Y)}{\partial \alpha_k}\right|_{\boldsymbol{\alpha}^*} = \lambda \quad \forall k:\ \alpha_k^* > 0, \\ \\ \left.\frac{\partial I(\Pi_X;\Pi_Y)}{\partial \alpha_k}\right|_{\boldsymbol{\alpha}^*} \le \lambda \quad \forall k:\ \alpha_k^* = 0, \end{cases} \tag{21}$$

where $\sum_{i=0}^{\min[m,T]} \alpha_i^* = 1$, $0 \le k \le \min[m,T]$, and $\boldsymbol{\alpha}^*$ is the vector of the optimum input probabilities of choosing subspaces of certain dimension,

$$\boldsymbol{\alpha}^* = \begin{bmatrix} \alpha_0^* & \cdots & \alpha_{\min[m,T]}^* \end{bmatrix}^{\mathrm{T}}.$$

***Lemma** 10:* By taking the partial derivative of the mutual information given in (19) with respect to α_k, we have

$$I_k' \triangleq \frac{\partial I(\Pi_X;\Pi_Y)}{\partial \alpha_k} = -nk\log_2 q - \sum_{d_y=0}^{\min[n,k]} \psi(n,d_y) \begin{bmatrix} k \\ d_y \end{bmatrix} q^{-nk} \log_2\left(f(d_y)\right) - \log_2 e. \tag{22}$$

Multiplying both sides of (22) by α_k and summing over k we get

$$I - \log_2 e = \sum_{k=0}^{\min[m,T]} \alpha_k I'_k.$$

By choosing the optimal values $\alpha_k = \alpha_k^*$ for $0 \leq k \leq \min[m,T]$, the RHS becomes λ, and the mutual information increases to C_s. So we may write $\lambda = C_s - \log_2 e$.

D. Solution for Large Field Size

In this subsection, we focus on large size fields, $q \gg 1$. This assumption allows us to use some approximations to simplify the conditions in (21). Assuming large q we can rewrite (22) as follows

$$I'_k = -nk \log_2 q - \log_2 e - \sum_{d_y=0}^{\min[n,k]} \left(1 + O(q^{-1})\right) q^{-(n-d_y)(k-d_y)} \log_2 \left(f(d_y)\right), \tag{23}$$

where we have used Lemma 1 and Lemma 3. Using similar approximations, $\log_2 f(d_y)$ defined in (20) can be approximated as

$$\log_2 \left(f(d_y)\right) = -\, d_y T \log_2 q + O(q^{-1}) + \log_2 \left(\sum_{d_x=d_y}^{\min[m,T]} q^{-(n-d_y)d_x} \alpha_{d_x} \right). \tag{24}$$

Then we have the following result, Lemma 11.

***Lemma** 11:* The dominating term in the summation in (23) is the one obtained for $d_y = \min[n,k]$.

From the proof of Lemma 11 written in Appendix A, we can also see that the remaining terms in the summation of (23) are of order $o(1)$, so we can write

$$I'_k = [T \min[n,k] - nk] \log_2 q + \underbrace{o(1)}_{\epsilon_q(k)} - \log_2 e - \log_2 \left(\sum_{d_x=\min[n,k]}^{\min[m,T]} q^{-[n-\min[n,k]]d_x} \alpha_{d_x} \right). \tag{25}$$

Assuming that the expression inside the $\log(\cdot)$ function in (25) is not zero for every $0 \leq k \leq \min[m,T]$, we can rewrite the Kuhn-Tucker conditions as

$$\sum_{d_x=\min[n,k]}^{\min[m,T]} q^{-[n-\min[n,k]]d_x} \alpha_{d_x} \geq 2^{-C_s + o(1)} q^{[T \min[n,k] - nk]},$$

where the inequality holds with equality for all k with $\alpha_k^* > 0$.

Let $\delta \triangleq \min[m,T]$ and define the $(\delta+1) \times (\delta+1)$ matrix $\mathbf{A}$ with elements

$$\mathbf{A}_{ij} \triangleq \begin{cases} q^{-[n-\min[n,i]]j} & \min[n,i] \leq j \leq \delta, \\ 0 & \text{otherwise.} \end{cases}$$

We also define the column vector $\mathbf{b}$ with elements $\mathbf{b}_i \triangleq q^{[T\min[n,i]-ni]}$ for $0 \leq i \leq \delta$. Note that for convenience the indices of matrix $\mathbf{A}$ and vector $\mathbf{b}$ start from 0. Using these definitions, we are able to rewrite the Kuhn-Tucker conditions in the matrix form as

$$\mathbf{A}\boldsymbol{\alpha}^* \succeq 2^{-C_s+o(1)}\mathbf{b}. \tag{26}$$

In the following, we consider two cases for $\delta \leq n$ and $\delta > n$, and find $\boldsymbol{\alpha}^*$ for each of them, separately. **First case:** $\delta \leq n$**.** In this case we can explicitly write the matrix $\mathbf{A}$ and vector $\mathbf{b}$ as

$$\mathbf{A} = \begin{bmatrix} 1 & q^{-n} & \cdots & q^{-(\delta-1)n} & q^{-\delta n} \\ 0 & q^{-(n-1)} & \cdots & q^{-(\delta-1)(n-1)} & q^{-\delta(n-1)} \\ 0 & 0 & \cdots & q^{-(\delta-1)(n-2)} & q^{-\delta(n-2)} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & q^{-(\delta-1)(n-\delta+1)} & q^{-\delta(n-\delta+1)} \\ 0 & 0 & \cdots & 0 & q^{-\delta(n-\delta)} \end{bmatrix},$$

and

$$\mathbf{b} = \begin{bmatrix} 1 & q^{(T-n)} & \cdots & q^{\delta(T-n)} \end{bmatrix}^{\mathrm{T}}.$$

The fact that the expression inside the $\log_2(\cdot)$ function in (25) is non-zero for $k = \delta$, forces α^*_δ to be positive. Thus the last row of the matrix inequality in (26) should be satisfied as an equality. Therefore,

$$\alpha^*_\delta = \frac{q^{\delta(T-n)}}{q^{-\delta(n-\delta)}} 2^{-C_s+o(1)} = q^{\delta(T-\delta)} 2^{-C_s+o(1)}.$$

Now we use induction to show that the optimal solution has the form

$$\alpha^*_i = \begin{cases} q^{i(T-i)} 2^{-C_s+o(1)} & : \quad \kappa \leq i \leq \delta, \\ 0 & : \quad 0 \leq i < \kappa, \end{cases} \tag{27}$$

where we will determine κ later.

Let us fix l and assume that $\alpha^*_i = q^{i(T-i)} 2^{-C_s+o(1)}$ for $0 \leq l < i \leq \delta$. Then for α^*_l we can write

$$A_{ll}\alpha^*_l + \sum_{j=l+1}^{\delta} q^{-(n-l)j}\alpha^*_j \geq q^{l(T-n)} 2^{-C_s+o(1)},$$

or equivalently

$$\begin{aligned} A_{ll}\alpha^*_l \geq & q^{l(T-n)} 2^{-C_s+o(1)} - \sum_{j=l+1}^{\delta} q^{-(n-l)j}\alpha^*_j \\ = & q^{l(T-n)} 2^{-C_s+o(1)} \left[1 - \sum_{j=l+1}^{\delta} q^{(T-n-j)(j-l)} \right]. \end{aligned} \tag{28}$$

We can use induction for one step more to show that α_l^* is of the desired form (27) if the previous expression is satisfied with equality. This is true if we have $1-\sum_{j=l+1}^{\delta} q^{(T-n-j)(j-l)} \geq 0$, or equivalently (assuming large q) if we have $(T-n-j)|_{j=l+1} < 0$. So we can conclude that we should have $(T-n)^+ \leq l \leq \delta$. It can be easily verified that for $i < (T-n)^+$ the Kuhn-Tucker equation for α_i^* satisfies the strict inequality so $\alpha_i^* = 0$ for $i < \min[(T-n)^+, \delta]$. The above argument results in a solution of the following form for the case $\delta \leq n$

$$\alpha_i^* = \begin{cases} q^{i(T-i)} 2^{-C_s + o(1)} & : \quad \min\left[(T-n)^+, \delta\right] \leq i \leq \delta, \\ 0 & : \quad 0 \leq i < \min\left[(T-n)^+, \delta\right]. \end{cases} \tag{29}$$

Second case: $\delta > n$**.** We now write matrix $\mathbf{A}$ and vector $\mathbf{b}$ as

$$\mathbf{A} = \left[\begin{array}{ccccccc} 1 & q^{-n} & \cdots & \cdots & \cdots & \cdots & q^{-\delta n} \\ 0 & q^{-(n-1)} & \cdots & \cdots & \cdots & \cdots & q^{-\delta(n-1)} \\ \vdots & \ddots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & q^{-(n-1)} & q^{-n} & \cdots & q^{-\delta} \\ \hline 0 & \cdots & 0 & 0 & 1 & \cdots & 1 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 1 & \cdots & 1 \end{array}\right],$$

and

$$\mathbf{b} = \left[1 \quad q^{(T-n)} \quad \ldots \quad q^{(n-1)(T-n)} \quad q^{n(T-n)} \quad q^{n(T-n-1)} \quad \ldots \quad q^{n(T-\delta)}\right]^{\mathrm{T}}.$$

The last $\delta - n + 1$ rows of $\mathbf{A}$ are the same while b_i is decreasing with i for $i \geq n$. Thus, the last $\delta - n$ inequalities are strict and therefore,

$$\alpha_{n+1}^* = \cdots = \alpha_\delta^* = 0. \tag{30}$$

The remaining equations can simply be reduced to the first case. Define

$$\tilde{\mathbf{A}} = \left[\begin{array}{ccccc} 1 & q^{-n} & \cdots & q^{-(n-1)n} & q^{-n^2} \\ 0 & q^{-(n-1)} & \cdots & q^{-(n-1)(n-1)} & q^{-n(n-1)} \\ 0 & 0 & \cdots & q^{-(n-1)(n-2)} & q^{-n(n-2)} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & q^{-(n-1)} & q^{-n} \\ 0 & 0 & \cdots & 0 & 1 \end{array}\right],$$

and

$$\tilde{\mathbf{b}} = \left[\begin{array}{cccc} 1 & q^{(T-n)} & \ldots & q^{n(T-n)} \end{array}\right]^{\mathrm{T}}.$$

The remaining conditions in this case can be written as

$$\tilde{\mathbf{A}}\boldsymbol{\alpha}^* \succeq 2^{-C_s+o(1)}\tilde{\mathbf{b}},$$

which is exactly similar to (26), for $\delta = n$. Therefore, the optimal solution for the first case will also satisfy these conditions, *i.e.*,

$$\alpha_i^* = \begin{cases} q^{i(T-i)}2^{-C_s+o(1)} & \kappa \leq i \leq n, \\ 0 & 0 \leq i < \kappa, \end{cases} \tag{31}$$

with $\kappa = \min[(T-n)^+, n]$. Summarizing (30) and (31), we can obtain the optimal solution for this regime, as

$$\alpha_i^* = \begin{cases} 0 & n < i \leq \delta, \\ q^{i(T-i)}2^{-C_s+o(1)} & \kappa \leq i \leq n, \\ 0 & 0 \leq i < \kappa, \end{cases} \tag{32}$$

where $\kappa = \min[(T-n)^+, n]$. This completes the proof of Theorem 3. By normalizing α_i^* to 1 we can also obtain an alternative proof to Theorem 2.

Discussion: To characterize the exact value of q_0 one have to consider the exact form of the set of equations given in (28) (for each l) which are as follows,

$$A_{ll}\alpha_l^* \geq q^{l(T-n)}2^{-C_s+\epsilon_q(l)}\left[1 - \sum_{j=l+1}^{\delta} q^{(T-n-j)(j-l)}2^{[\epsilon_q(j)-\epsilon_q(l)]}\right].$$

Although it is hard to find q_0 exactly, it is possible to show that there exists finite q_0 such that result of Theorem 3 holds for. This can be done by solving above equations assuming that $\epsilon_q(k)$ is zero for every k (assuming $q \gg 1$). Then, it can be observed that the RHS of (28) are either greater or less than zero. Now by assuming finite but large enough q and considering the exact form of (28) we have some small perturbations that cannot change the sign of RHS of (28) so we are done.

E. Proof of Theorem 4

Let $\epsilon_q(k)$ denotes the error term in (25). We can easily write the exact expression for $\epsilon_q(k)$ which is as follows

$$\begin{aligned}\epsilon_q(k) = &- \sum_{d_y=0}^{r_k} \psi(n,d_y)\begin{bmatrix} k \\ d_y \end{bmatrix} q^{-nk}\log_2\left(\sum_{d_x=d_y}^{\min[m,T]} \alpha_{d_x}\frac{\begin{bmatrix} d_x \\ d_y \end{bmatrix}}{\begin{bmatrix} T \\ d_y \end{bmatrix}} q^{-nd_x}\right) \\ &+ \log_2\left(\sum_{d_x=r_k}^{\min[m,T]} q^{r_k(d_x-r_k)-nd_x}\alpha_{d_x}\right) - r_k(T-r_k)\log_2 q,\end{aligned}$$

where $r_k = \min[n,k]$.

We consider the case where $T > n + \min[m,n]$ so Theorem 3 implies that for the optimal input distribution we have $\alpha_{i^*} = 1$ where $i^* = \min[m,n]$ and $q > q_0$. Then we can simplify $\epsilon_q(k)$ more and write

$$\epsilon_q(k) = \sum_{d_y=0}^{r_k} \psi(n,d_y) \begin{bmatrix} k \\ d_y \end{bmatrix} q^{-nk} \log_2 \left(\frac{\begin{bmatrix} T \\ d_y \end{bmatrix}}{\begin{bmatrix} i^* \\ d_y \end{bmatrix}} \right) - r_k(T - i^*), \tag{33}$$

where we also use Lemma 4 in the above simplification.

To find q_0, the minimum value of q that the result of Theorem 4 is valid for, we should consider the exact form of (28) and check that the RHS of (28) is less than or equal to zero for $0 \le l \le (i^* - 1)$. So from (28) for every $0 \le l \le (i^* - 1)$ we may write

$$\left[1 - q^{(T-n-i^*)(i^*-l)} 2^{[\epsilon_q(i^*) - \epsilon_q(l)]} \right] \le 0,$$

or equivalently

$$\frac{\epsilon_{q_0}(l) - \epsilon_{q_0}(i^*)}{(T-n-i^*)(i^*-l)} \le \log_2 q_0, \quad \forall l : 0 \le l \le (i^* - 1). \tag{34}$$

Using a similar argument we should have also

$$\frac{\epsilon_{q_0}(l) - \epsilon_{q_0}(i^*)}{i^*(l-i^*)} \le \log_2 q_0, \quad \forall l : (i^* + 1) \le l \le m. \tag{35}$$

From (32) for the capacity C_s we have $C_s = i^*(T - i^*) \log_2 q + \epsilon_q(i^*)$. Evaluating (33) at $k = i^*$ we have

$$\epsilon_q(i^*) = \sum_{d_y=0}^{i^*} \psi(n,d_y) \begin{bmatrix} i^* \\ d_y \end{bmatrix} q^{-ni^*} \log_2 \left(\frac{\begin{bmatrix} T \\ d_y \end{bmatrix}}{\begin{bmatrix} i^* \\ d_y \end{bmatrix}} \right) - i^*(T - i^*) \log_2 q,$$

which results in the capacity stated in the assertion of Theorem 4.

Discussion: We derive a sufficient condition on the minimum size of q to satisfy the set of conditions stated in (34) and (35). Using this sufficient condition we explore the behavior of q_0 as T increases.

For $k \ne i^*$ we can write

$$\begin{aligned} \epsilon_q(k) &\overset{(a)}{\le} 4 \sum_{d_y=0}^{r_k} q^{-(n-d_y)(k-d_y)} \log_2 \left(4q^{d_y(T-i^*)} \right) - r_k(T-i^*) \log_2 q \\ &\le 8 + 4r_k q^{-(\max[n,k] - \min[n,k] + 1)} \left(2 + (r_k - 1)(T - i^*) \log_2 q \right) \\ &\overset{(b)}{\le} (8 + 8r_k) + \left(4r_k(r_k - 1)(T - i^*) \frac{\log_2 q}{q^{(\max[n,k] - \min[n,k] + 1)}} \right), \end{aligned} \tag{36}$$

where (a) follows from (2) and (3), and in (b) we use the fact that $k \ne i^*$.

Then for $k = i^*$ we can write

$$\begin{aligned}\epsilon_q(i^*) &\geq \psi(n, i^*) q^{-ni^*} \log_2 \begin{bmatrix} T \\ i^* \end{bmatrix} - i^*(T - i^*) \log_2 q \\ &\overset{(a)}{\geq} -(i^*)^2 (T - i^*) \frac{\log_2 q}{q^{n-i^*+1}}, \end{aligned} \tag{37}$$

where (a) follows from (2) and (3).

Let us consider two cases. First, we assume that $m \leq n$ so $i^* = m$. To find a sufficient condition for q_0 we have to only consider conditions given in (34). Using (36) and (37) and assuming that $T \to \infty$ we should have $\log_2 q_0 \geq 5m^2 q_0^{-n+m-1} \log_2 q_0$, or equivalently $q_0^{n-m+1} \geq 5(i^*)^2$.

For the second case we have $m > n$ which means $i^* = n$. Here, using a similar argument to the one given above for the first case we can show that conditions (34) give some constant q_0 as $T \to \infty$. However, the conditions (35) give a sufficient condition for q_0 which grows as $T \to \infty$. Now, using (35), (36), and (37) and assuming that $T \to \infty$, a sufficient condition for q_0 would be $\log_2 q_0 \geq 4nTq_0^{-2} \log_2 q_0 + nTq_0^{-1} \log_2 q_0$. For large T for the sufficient condition we have $q_0 \geq i^* T$.

V. Multiple Sources Scenario: The Rate Region

The goal of this section is to characterize $\mathcal{R}$, the set of all achievable rate pairs (R_1, R_2) for two user communication over the multiple access channel $\mathcal{C}_{\textit{m-MAC}}$ described in Definition 6. More precisely, we will show that $\mathcal{R} = \mathcal{R}^*$. In order to do this, we first formulate a mathematical model for this channel. Then, we present an achievability scheme, to show that $\mathcal{R}^*$ is achievable, *i.e.,* $\mathcal{R}^* \subseteq \mathcal{R}$. In the next subsection we prove the optimality of this scheme and show that $\mathcal{R} \subseteq \mathcal{R}^*$.

The proof of the converse part of the theorem is based on two outer bounds, namely, a cooperative bound and a coloring bound. For the coloring bound, we utilize a combinatorial argument to bound the number of *distinguishable* symbol pairs that can be transmitted from the two sources to the destination. This bound allows us to restrict the *effective* input alphabets of the sources to subsets of the original alphabets, with significantly smaller size. We can then easily bound the capacity region of the network using the restricted input alphabet.

The transition probability of the channel given by Definition 6, $P_{Y|X_1X_2}$, can be written as [9]

$$P_{Y|X_1X_2}(y|x_1, x_2) = \begin{cases} q^{-n \dim(\langle x_1 \rangle + \langle x_2 \rangle)} & \langle y \rangle \sqsubseteq \langle x_1 \rangle + \langle x_2 \rangle, \\ 0 & \text{otherwise.} \end{cases} \tag{38}$$

Our first result, stated in Theorem 6, is that the multiple access matrix channel described in Definition 6 is equivalent to the "subspace" channel $\mathrm{Ch}_{\textit{s-MAC}}$ described in Definition 7, that has subspaces as inputs

and outputs. So to characterize the optimal rate region of $\mathrm{Ch}_{m\text{-}MAC}$, we can focus on finding the optimal rate region of $\mathrm{Ch}_{s\text{-}MAC}$. We will use this equivalence in the rest of this section.

We know from [15] that the rate region of the multiple access channel $\mathrm{Ch}_{s\text{-}MAC}$ is given by the closure of the convex hull of the rate vectors satisfying

$$R_S \leq I(\Pi_{X_S}; \Pi_Y | \Pi_{X_{S^c}}) \quad \text{for all } S \subseteq \{1, \ldots, N_s\},$$

for some product distribution $P_{\Pi_{X_1}}(\pi_1) \cdots P_{\Pi_{X_{N_s}}}(\pi_{N_s})$. Note that $R_S = \sum_{i \in S} R_i$, where R_i is the transmission rate of the ith source, $\Pi_{X_S} = \{\Pi_{X_i} : i \in S\}$ and S^c is the complement set of S.

A. Achievability Scheme

In this subsection we illustrate a simple achievability scheme for the corner points of the rate region defined in Theorem 7. The remaining points in the rate region can be achieved using time-sharing.

For given $(d_1, d_2) \in \mathcal{D}^*$, define the following subspace code-books

$$\widetilde{\mathcal{C}}_1 \triangleq \left\{ \langle X_1 \rangle : X_1 = \left[\begin{array}{c|c|c} \mathbf{I}_{d_1 \times d_1} & \mathbf{0}_{d_1 \times d_2} & \mathbf{U}_1 \\ \hline \mathbf{0}_{(m_1-d_1)\times d_1} & \mathbf{0}_{(m_1-d_1)\times d_2} & \mathbf{0}_{(m_1-d_1)\times(T-d_1-d_2)} \end{array}\right], \mathbf{U}_1 \in \mathbb{F}_q^{d_1 \times (T-d_1-d_2)} \right\}$$

and

$$\widetilde{\mathcal{C}}_2 \triangleq \left\{ \langle X_2 \rangle : X_2 = \left[\begin{array}{c|c|c} \mathbf{0}_{d_2 \times d_1} & \mathbf{I}_{d_2 \times d_2} & \mathbf{U}_2 \\ \hline \mathbf{0}_{(m_2-d_2)\times d_1} & \mathbf{0}_{(m_2-d_2)\times d_2} & \mathbf{0}_{(m_2-d_2)\times(T-d_1-d_2)} \end{array}\right], \mathbf{U}_2 \in \mathbb{F}_q^{d_2 \times (T-d_1-d_2)} \right\}.$$

If we transmit messages from these code-books, we have

$$\begin{aligned} Y &= H_1 X_1 + H_2 X_2 \\ &= \left[\begin{array}{c|c|c} \hat{H}_1 & \hat{H}_2 & \hat{H}_1 \mathbf{U}_1 + \hat{H}_2 \mathbf{U}_2 \end{array}\right], \end{aligned}$$

where $\hat{H}_i$ captures the first d_i columns of H_i. Therefore, decoding at the receiver would be just recovering of $\mathbf{U}_1$ and $\mathbf{U}_2$ given $\hat{H}_1 \mathbf{U}_1 + \hat{H}_2 \mathbf{U}_2$, $\hat{H}_1$, and $\hat{H}_2$. Since $d_1 + d_2 \leq n$, the matrix $[\hat{H}_1 \ \hat{H}_2]$ is full-rank with high probability, and therefore the decoder is able to decode $\mathbf{U}_1$ and $\mathbf{U}_2$.

Note that the achievability scheme uses effectively the coding vectors approach [12]. This indicates that for $\frac{T}{2} > \max[m_1 + m_2, n]$ and q large enough, the subspace coding and the coding vectors approach achieve the same rate.

B. Outer bound on the Admissible Rate Region

In the following we will present an outer bound for $\mathcal{R}$, the admissible rate region of the non-coherent two-user multiple access channel $\mathrm{Ch}_{m\text{-}MAC}$. Recall that by Theorem 6 we can focus on the subspace channel $\mathrm{Ch}_{s\text{-}MAC}$. We first show in Proposition 1 that $\mathcal{R} \subseteq \mathcal{R}_{\mathrm{coop}}$, a cooperative outer-bound. Then Proposition 2 demonstrates that $\mathcal{R} \subseteq \mathcal{R}_{\mathrm{col}}$, a coloring outer-bound. Finally we show that $\mathcal{R}_{\mathrm{col}} \cap \mathcal{R}_{\mathrm{coop}} \subseteq \mathcal{R}$, yielding the desired outer-bound $\mathcal{R} \subseteq \mathcal{R}^*$ which matches the achievability of §V-A.

The first outer bound, called cooperating outer bound, is simply obtained by letting the two transmitters cooperate to transmit their messages to the receiver, i.e. we assume they form a super-source. Applying Theorem 2 for the non-coherent scenario for the single super-source, the one who controls the packets of both transmitters, we have the following proposition.

***Proposition** 1:* Let $\frac{T}{2} \geq m_1 + m_2$. We have $\mathcal{R} \subseteq \mathcal{R}_{\mathrm{coop}}$ where

$$\mathcal{R}_{\mathrm{coop}} \triangleq \{(R_1, R_2) : \ R_1 + R_2 \leq k(T-k)\log_2 q\},$$

and $k = \min[m_1 + m_2, n]$.

The rest of this section is dedicated to deriving the second outer bound which is denoted by $\mathcal{R}_{\mathrm{col}}$. This bound is based on an argument on the number of messages per channel use that each user can reliably communicate over the multiple access channel.

Let $(R_1, R_2) \in \mathcal{R}$ be an achievable rate pair for which there exists an encoding and decoding scheme with block length N and small error probability. One can follow the usual converse proof of the multiple access channel from [15] to show that

$$R_1 \leq I(\Pi_{X_1}^N; \Pi_Y^N | \Pi_{X_2}^N) \leq \frac{1}{N}\sum_{t=1}^{N} I(\Pi_{X_1 t}; \Pi_{Y t} | \Pi_{X_2 t}),$$

$$R_2 \leq I(\Pi_{X_2}^N; \Pi_Y^N | \Pi_{X_1}^N) \leq \frac{1}{N}\sum_{t=1}^{N} I(\Pi_{X_2 t}; \Pi_{Y t} | \Pi_{X_1 t}),$$

$$R_1 + R_2 \leq I(\Pi_{X_1}^N, \Pi_{X_2}^N; \Pi_Y^N) \leq \frac{1}{N}\sum_{t=1}^{N} I(\Pi_{X_1 t}, \Pi_{X_2 t}; \Pi_{Y t}).$$

For each time instance t, denote by $\widetilde{\mathcal{C}}_{i,t}$, the projection of the code-book used by user i to its t-th element. For a single source scenario, we have shown in §IV that we can use the set $\mathrm{Sp}(T, m)$ as our input alphabet for all time slots, and have the receiver successfully decode the sent messages, and hence, the user can communicate $\mathcal{S}(T, m)$ distinct messages. For the multi-source case, $\widetilde{\mathcal{C}}_{i,t}$ is more restricted. The main reason for this is that the transition probability of the multiple access channel $P_{\Pi_Y | \Pi_{X_1} \Pi_{X_2}}$ is of the form

$P_{\Pi_Y|\Pi_{X_1}+\Pi_{X_2}}$. That is, if $(\pi_1,\pi_2)\in\widetilde{\mathcal{X}}_1\times\widetilde{\mathcal{X}}_2$ and $(\pi_1',\pi_2')\in\widetilde{\mathcal{X}}_1\times\widetilde{\mathcal{X}}_2$ satisfy $\pi_1+\pi_2=\pi_1'+\pi_2'$, then $P(\Pi_Y|\pi_1,\pi_2)=P(\Pi_Y|\pi_1',\pi_2')$, and hence the receiver cannot distinguish between the two pairs.

In the following we will discuss this indistinguishability in detail, and derive the maximum number of distinguishable pairs which can be conveyed through the channel. In order to do so, we start with some useful definitions and lemmas.

***Definition** 8:* For a fixed $\pi_1\in\mathrm{Gr}(T,d_1)$, we denote by $\mathcal{N}(\pi_1,d_2,d_{12})$ the set of subspaces of dimension d_2 that intersect with π_1 at d_{12} dimensions, *i.e.*,

$$\mathcal{N}(\pi_1,d_2,d_{12})\triangleq\{\pi_2\in\mathrm{Gr}(T,d_2):\dim(\pi_1\cap\pi_2)=d_{12}\}. \tag{39}$$

It turns out that the cardinality of the set $\mathcal{N}(\pi_1,d_2,d_{12})$ depends on π_1 only through its dimension, $d_1=\dim(\pi_1)$. Therefore, we denote this number by $n(d_1,d_2,d_{12})$, which is characterized in the following lemma.

***Lemma** 12:* The cardinality of the set $N(\pi_1,d_2,d_{12})$ is given by

$$n(d_1,d_2,d_{12})=|N(\pi_1,d_2,d_{12})|\doteq q^{d_{12}(d_1-d_{12})+(d_2-d_{12})(T-d_2)}. \tag{40}$$

***Definition** 9:* For a fixed $\pi_1\in\mathrm{Gr}(T,d_1)$ and $\pi_2\in\mathrm{Gr}(T,d_2)$, we define

$$A(\pi_1,\pi_2)\triangleq\{\pi_2'\in\mathrm{Gr}(T,d_2):\pi_1+\pi_2'=\pi_1+\pi_2\}. \tag{41}$$

***Lemma** 13:* The cardinality of the set $A(\pi_1,\pi_2)$ only depends on the dimensions of the two subspaces and their intersection, $d_1=\dim(\pi_1)$, $d_2=\dim(\pi_2)$, and $d_{12}=\dim(\pi_1\cap\pi_2)$. Moreover, it can be asymptotically characterized by

$$a(d_1,d_2,d_{12})=|A(\pi_1,\pi_2)|\doteq q^{d_2(d_1-d_{12})}. \tag{42}$$

.

***Definition** 10:* For an arbitrary set $\widetilde{\mathcal{C}}\subseteq\mathrm{Sp}(T,m)$, we denote the projection of $\widetilde{\mathcal{C}}$ onto the set of d-dimensional Grassmannian $\widetilde{\mathcal{C}}(d)$. Formally,

$$\widetilde{\mathcal{C}}(d)\triangleq\widetilde{\mathcal{C}}\cap\mathrm{Gr}(T,d)=\{\pi\in\widetilde{\mathcal{C}}:\dim(\pi)=d\}.$$

For a fixed time instance t, and corresponding subsets $\widetilde{\mathcal{C}}_{1,t}$ and $\widetilde{\mathcal{C}}_{2,t}$, we can construct a table with $|\widetilde{\mathcal{C}}_{1,t}|$ rows and $|\widetilde{\mathcal{C}}_{2,t}|$ columns, each row (column) corresponding to one subspace π_1 (π_2) in $\widetilde{\mathcal{C}}_{1,t}$ ($\widetilde{\mathcal{C}}_{2,t}$). In the following, we define an equivalence relation for the cells of this table.

***Definition** 11:* A *coloring* for a table constructed as above is an assignment of colors to the cells of the table using a function col : $\widetilde{\mathcal{C}}_{1,t}\times\widetilde{\mathcal{C}}_{2,t}\to\mathbb{N}$ such that $\mathrm{col}(\pi_1,\pi_2)=\mathrm{col}(\pi_1',\pi_2')$ if and only if $\pi_1+\pi_2=\pi_1'+\pi_2'$.

It is clear that the coloring definition above exactly matches with that of indistinguishability we discussed before. More precisely, two pairs of subspaces (π_1, π_2) and (π_1', π_2') are distinguishable if and only if their corresponding cells in the table have different colors. The following theorem upper bounds the cardinality of the subspace sets based on this fact.

***Theorem** 8:* For each pair of uniquely distinguishable sets $(\widetilde{\mathcal{C}}_{1,t}, \widetilde{\mathcal{C}}_{2,t})$ defined on the input alphabet $\widetilde{\mathcal{X}}_1 \times \widetilde{\mathcal{X}}_2$ for the multiple access channel $\mathrm{Ch}_{\textit{s-MAC}}$, there exist integer numbers $0 \leq \delta_i(t) \leq m_i$ such that

$$|\widetilde{\mathcal{C}}_{i,t}| \doteq q^{\delta_i(t)(T-\delta_1(t)-\delta_2(t))}, \qquad i = 1, 2. \tag{43}$$

Proof: We may drop the time index t in this proof for brevity. For a fixed t, let δ_i be the *dominating* dimension in the set $\widetilde{\mathcal{C}}_i$, *i.e.*,

$$\delta_i \triangleq \arg\max_d |\widetilde{\mathcal{C}}_i(d)|,$$

where $\widetilde{\mathcal{C}}_i(d)$ is as defined in Definition 10. It is clear that

$$|\widetilde{\mathcal{C}}_i| = \sum_d |\widetilde{\mathcal{C}}_i(d)| \leq m_i |\widetilde{\mathcal{C}}_i(\delta_i)| \doteq |\widetilde{\mathcal{C}}_i(\delta_i)|, \tag{44}$$

where the last asymptotic equality follows from the fact that m_i is a constant with respect to the underlying field size q. This means that we may lose only a constant factor in the code-book size by removing all subspaces from $\widetilde{\mathcal{C}}_1$ ($\widetilde{\mathcal{C}}_2$), except the ones that have dimension δ_1 (δ_2) . Therefore the loss in the rate values would be negligible as q grows. Consider the table constructed for $\widetilde{\mathcal{C}}_1(\delta_1)$ and $\widetilde{\mathcal{C}}_2(\delta_2)$. Let $\pi_1 \in \widetilde{\mathcal{C}}_1(\delta_1)$ be a δ_1-dimensional subspace, and consider the corresponding row of the table. We further partition the columns of the table with respect to π_1 into $\bigcup_{d_{12}=0}^{\min[\delta_1,\delta_2]} \widetilde{\mathcal{C}}_2(\pi_1, \delta_2, d_{12})$, where

$$\widetilde{\mathcal{C}}_2(\pi_1, \delta_2, d_{12}) \triangleq \{\pi_2 \in \widetilde{\mathcal{C}}_2(\delta_2) : \dim(\pi_1 \cap \pi_2) = d_{12}\}. \tag{45}$$

We use $K(\pi_1, \delta_2)$ and $K(\pi_1, \delta_2, d_{12})$ to denote the number of different colors in the row that corresponds to π_1 and its intersection with $\widetilde{\mathcal{C}}_2(\pi_1, \delta_2, d_{12})$, respectively.

Note that $\widetilde{\mathcal{C}}_2(\pi_1, \delta_2, d_{12}) \subseteq \mathcal{N}(\pi_1, \delta_2, d_{12})$, and therefore the number of different colors that appear in this partition of the row, cannot exceed the number of colors that could potentially appear if $\mathcal{N}(\pi_1, \delta_2, d_{12}) \subseteq \widetilde{\mathcal{C}}_2$. Recall that $\mathcal{N}(\pi_1, \delta_2, d_{12})$ has $n(\delta_1, \delta_2, d_{12})$ elements, which are split into subsets of size $a(\delta_1, \delta_2, d_{12})$ of the same color. Therefore, for a large field size, the number of different colors in this partition of the row corresponding to π_1, can be upper bounded as

$$K(\pi_1, \delta_2, d_{12}) \leq \frac{n(\delta_1, \delta_2, d_{12})}{a(\delta_1, \delta_2, d_{12})} \doteq q^{(\delta_2 - d_{12})(T-\delta_1-\delta_2+d_{12})}. \tag{46}$$

Hence,

$$\begin{aligned} K(\pi_1,\delta_2) &= \sum_{d_{12}=0}^{\min[\delta_1,\delta_2]} K(\pi_1,\delta_2,d_{12}) \\ &\dot{\leq} \sum_{d_{12}=0}^{\min[\delta_1,\delta_2]} q^{(\delta_2-d_{12})(T-\delta_1-\delta_2+d_{12})} \\ &\doteq q^{\max_{0\leq d_{12}\leq \min[\delta_1,\delta_2]}(\delta_2-d_{12})(T-\delta_1-\delta_2+d_{12})} \\ &= q^{\delta_2(T-\delta_1-\delta_2)} \end{aligned} \tag{47}$$

where the asymptotic inequality and equality hold for large q. Moreover, the last equality is based on the assumption $T \geq 2(m_1+m_2) \geq 2(\delta_1+\delta_2)$ and the fact that the exponent is a decreasing function of d_{12} for $0 \leq d_{12} \leq \min[\delta_1,\delta_2]$.

It is worth mentioning that this argument holds for each choice of $\pi_1 \in \widetilde{\mathcal{C}}_1(\delta_1)$. This means if the first user transmits a δ_1-dimensional subspace, the receiver cannot distinguish more that $q^{\delta_2(T-\delta_1-\delta_2)}$ different symbols. The same argument holds for a fixed column $\pi_2 \in \widetilde{\mathcal{C}}_2$ which yields an upper bound to the number of distinguishable messages as $q^{\delta_1(T-\delta_1-\delta_2)}$. ■

Theorem 8 essentially upper bounds the single letter mutual information $I(\Pi_{X_1t};\Pi_{Yt}|\Pi_{X_2t})$ for any time instance t. The following proposition summarizes this discussion.

***Proposition** 2:* We have $\mathcal{R} \subseteq \mathcal{R}_{\text{col}}$ where

$$\mathcal{R}_{\text{col}} \triangleq \text{convex hull} \bigcup_{(d_1,d_2)\in\mathcal{D}_{\text{col}}} \mathcal{R}(d_1,d_2),$$

in which $\mathcal{R}(d_1,d_2)$ is as defined in (14), and

$$\mathcal{D}_{\text{col}} \triangleq \{(d_1,d_2):\ 0 \leq d_i \leq m_i\}.$$

Proof: Using Theorem 8, we can upper bound the number of distinguishable pairs for each time instance. For a fixed t, let $\delta_1(t)$ and $\delta_2(t)$ denote the dominating dimensions. Therefore, we have

$$\begin{aligned} R_1 &\leq \frac{1}{N}\sum_{t=1}^{N} I(\Pi_{X_1t};\Pi_{Yt}|\Pi_{X_2t}), \\ &\dot{\leq} \frac{1}{N}\sum_{t=1}^{N} \log_2 q^{[\delta_1(t)(T-\delta_1(t)-\delta_2(t))]} \\ &= \frac{1}{N}\sum_{t=1}^{N} \delta_1(t)(T-\delta_1(t)-\delta_2(t))\log_2 q, \end{aligned}$$

where $0 \leq \delta_i(t) \leq m_i$ for $t = 1, \ldots, N$, and $i = 1, 2$. Similarly, we have

$$R_2 \leq \frac{1}{N} \sum_{t=1}^{N} \delta_2(t)(T - \delta_1(t) - \delta_2(t)) \log_2 q.$$

Therefore,

$$(R_1, R_2) \leq \frac{1}{N} \sum_{t=1}^{N} \left(\delta_1(t)(T - \delta_1(t) - \delta_2(t)) \log_2 q, \delta_2(t)(T - \delta_1(t) - \delta_2(t)) \log_2 q\right). \tag{48}$$

It is clear that the RHS of (48) is a convex linear combination of the points

$$\{\delta_1(t)(T - \delta_1(t) - \delta_2(t)) \log_2 q, \delta_1(t)(T - \delta_1(t) - \delta_2(t)) \log_2 q\}_{t=1}^{N}$$

which are in the region $\mathcal{R}(\delta_1(t), \delta_2(t))$. This completes the proof. ■

Summarizing Proposition 1 and Proposition 2, we have $\mathcal{R} \subseteq \mathcal{R}_{\text{coop}} \cap \mathcal{R}_{\text{col}}$. So, it only remains to prove the following theorem in order to show that $\mathcal{R}^*$ is an outer bound for the admissible rate region.

***Theorem** 9:* We have $\mathcal{R}_{\text{coop}} \cap \mathcal{R}_{\text{col}} \subseteq \mathcal{R}^*$.

Before presenting the proof of the theorem, we give the following two lemmas, which help us to characterize the corner points of the region of our interest.

***Lemma** 14:* The set of corner points of $\mathcal{R}_{\text{col}}$ is the set of all rate pairs of the form

$$(R_1, R_2) = (R_1(d_1, d_2), R_2(d_1, d_2)),$$

for some $(d_1, d_2) \in \widetilde{\mathcal{D}}$, where

$$\widetilde{\mathcal{D}} = \{(0, m_2), (1, m_2), \ldots, (m_1, m_2), (m_1, m_2 - 1), \ldots, (m_1, 1), (m_1, 0)\}.$$

***Lemma** 15:* If $\mathcal{R}_{\text{col}} \nsubseteq \mathcal{R}_{\text{coop}}$, then any intersecting point of $R_1 + R_2 = k(T - k) \log_2 q$ with the boundary of $\mathcal{R}_{\text{col}}$ is a point of the form $(R_1(d_1, d_2), R_2(d_1, d_2))$, where

$$(d_1, d_2) \in \widetilde{\mathcal{D}} \cup \{(m_1 - 1, 0), \ldots, (0, 0), (0, 1), \ldots, (0, m_2 - 1)\}.$$

That is, the boundaries of $\mathcal{R}_{\text{col}}$ and $\mathcal{R}_{\text{coop}}$ can only intersect on either the corner points of $\mathcal{R}_{\text{col}}$ or the $R_1 - R_2$ axes.

Proof of Theorem 9: Note that $\mathcal{R}_{\text{coop}} \cap \mathcal{R}_{\text{col}}$ is a convex polytope, formed as intersection of a polytope and the convex hull of a finite number of polytopes. Therefore, it suffices to prove the theorem only for its corner points. Let $(R_1, R_2) \in \mathcal{R}_{\text{coop}} \cap \mathcal{R}_{\text{col}}$ be a corner point. It is clear that one of the followings occurs.

(i) (R_1, R_2) is a corner point of $\mathcal{R}_{\text{col}}$ and interior point of $\mathcal{R}_{\text{coop}}$;

(ii) (R_1, R_2) is an intersecting point of the boundaries of $\mathcal{R}_{\text{col}}$ and $\mathcal{R}_{\text{coop}}$.

In the former case, Lemma 14 which characterizes the set of corner points of $\mathcal{R}_{\text{col}}$, implies there exists a pair $(d_1, d_2) \in \widetilde{\mathcal{D}}$ such that $(R_1, R_2) = (R_1(d_1, d_2), R_2(d_1, d_2))$. Also $(R_1, R_2) \in \mathcal{R}_{\text{coop}}$ implies

$$(d_1 + d_2)(T - (d_1 + d_2)) \log_2 q = R_1 + R_2 \leq k(T - k) \log_2 q.$$

Note that the function $f(x) \triangleq x(T - x)$ is an increasing function of x for $x \in (0, T/2)$. Therefore, $d_1 + d_2 \leq k = \min\{m_1 + m_2, n\}$, and hence $(d_1, d_2) \in \mathcal{D}^*$, which implies that $(R_1, R_2) \in \mathcal{R}^*$.

In the latter case, it follows from Lemma 15 that (R_1, R_2) should be either a corner point of $\mathcal{R}_{\text{col}}$ for which the above argument holds, or of the form $(R_1, R_2) = (R_1(d_1, d_2), R_2(d_1, d_2))$ with $d_1 d_2 = 0$. Again $(R_1, R_2) \in \mathcal{R}_{\text{coop}}$, which implies that $d_1 + d_2 \leq k = \min\{m_1, m_2, n\}$, and $(R_1, R_2) \in \mathcal{R}^*$. This completes the proof. ■

***Corollary** 1:* The number of corner points of the rate region $\mathcal{R}^*$ excluding the point $(0, 0)$ is equal to

$$\min\left[m_1, (n - m_2)^+\right] + \min\left[m_2, (n - m_1)^+\right] + 2 - 1_{[n \geq m_1 + m_2]}.$$

Proof: By Lemma 14 the set of corner points of region $\mathcal{R}_{\text{col}}$ correspond to the pairs (d_1, d_2) which belong to the set $\{(0, m_2)...(m_1, m_2)...(m_1, 0)\}$. In this case the number of corner points excluding $(R_1, R_2) = (0, 0)$ is $m_1 + m_2 + 1$.

However the final rate region is the intersection of $\mathcal{R}_{\text{col}}$ and $\mathcal{R}_{\text{coop}}$, where the later one includes all the rate pairs with sum smaller than $k(T - k) \log_2 q$, $k = \min[m_1 + m_2, n]$, see Proposition 1.

Lemma 15 explains how these two regions intersect with each other. In this case, the corner points correspond to the pairs (d_1, d_2) which belong to the set $\{(0, m_2), \ldots, (\alpha, m_2), (m_1, \beta), \ldots, (m_1, 0)\}$ where $\alpha = \min[m_1, (n - m_2)^+]$ and $\beta = \min[m_2, (n - m_1)^+]$. So the number of corner points excluding $(0, 0)$ is

$$\alpha + \beta + 2 - 1_{n \geq m_1 + m_2},$$

where $1_{n \geq m_1 + m_2}$ takes into account the case where two points (α, m_2) and (m_1, β) overlap with each other. ■

VI. Conclusions

In this paper, we used a random matrix channel to model the problem of multicasting over a packet network that employs randomized network coding. We calculated the capacity of this channel for the case where the finite field of operation $\mathbb{F}_q$ is large, but showed through simulation results fast convergence for small values of q. We prove that use of subspace coding, proposed for algebraic coding in [6], [7], is optimal for this channel. Moreover, we showed that the capacity achieving distribution for very

small packet lengths uses subspaces of all dimensions, while as the packet length increases, the number of required dimensions in the optimal distribution decreases. In particular, the choice of the subspace dimension used in the seminal work of Koetter and Kschischang [6] is indeed optimal for large enough packet size. We extended our work to the case of multiple access with two sources, where we used a coloring argument to derive an outer bound for the capacity that we believe is interesting in itself. We showed that in all the cases we examined, the throughput benefits subspace coding offers as compared to the use of coding vectors go to zero as the alphabet size q increases, and thus use of coding vectors is (asymptotically) optimal.

ACKNOWLEDGEMENTS

The work of S. Mohajer and C. Fragouli was supported in part by the ERC Starting Investigator grant # 240317. The work of M. Jafari Siavoshani and C. Fragouli was supported in part by the Swiss National Science Foundation through the grant # PP002-110483. We would like to thank the anonymous reviewers for detailed comments that greatly enhanced the paper. In particular, one of the reviewers suggested an alternate proof for Theorem 2, which we have included in the paper in §IV-B. Our original proof of the result is used in the proof of Theorem 3 which gives a non-asymptotic characterization.

APPENDIX A
PROOFS

Proof of Theorem 1: To prove the theorem, we start with $I(X;Y)$ for the channel Ch_m, stated in (16), where the channel transition probability is given in (15). We will show that for each input distribution $P_X(x)$ there exists an input distribution $P_{\Pi_X}(\pi_x)$ for the channel Ch_s such that $I(X;Y) = I(\Pi_Y;\Pi_X)$ and vice versa.

We know that $P_{Y|X}(y|x) = P_{Y|X}(y|x')$ if $\langle x\rangle = \langle x'\rangle$. So we can write

$$I(X;Y) = \sum_{\pi_x\in\widetilde{\mathcal{X}},\ y\in\mathcal{Y}} P_{\Pi_X}(\pi_x)P_{Y|\Pi_X}(y|\pi_x)\log_2\left(\frac{P_{Y|\Pi_X}(y|\pi_x)}{P_Y(y)}\right),$$

where we choose $P_{\Pi_X}(\pi_x) = \sum_{x\in\mathcal{X}:\langle x\rangle=\pi_x} P_X(x)$ and define

$$P_{Y|\Pi_X}(y|\pi_x) \triangleq \begin{cases} q^{-n\dim(\pi_x)} & \langle y\rangle \sqsubseteq \pi_x, \\ 0 & \text{otherwise.}\end{cases}$$

Then expanding $I(X;Y)$ we have

$$I(X;Y) = \sum_{\pi_x\in\widetilde{\mathcal{X}}} P_{\Pi_X}(\pi_x)\sum_{\pi_y\in\widetilde{\mathcal{Y}}}\sum_{\substack{y\in\mathcal{Y},\\ \langle y\rangle=\pi_y}} P_{Y|\Pi_X}(y|\pi_x)\log_2\left(\frac{P_{Y|\Pi_X}(y|\pi_x)}{P_Y(y)}\right).$$

Now using the symmetry properties of $P_{Y|\Pi_X}(y|\pi_x)$ we can simplify $I(X;Y)$. In fact $P_{Y|\Pi_X}(y_1|\pi_x) = P_{Y|\Pi_X}(y_2|\pi_x)$ and $P_Y(y_1) = P_Y(y_2)$ if $\langle y_1\rangle = \langle y_2\rangle$. So we can remove the summation over y and write

$$I(X;Y) = \sum_{\pi_x \in \widetilde{\mathcal{X}}} P_{\Pi_X}(\pi_x) \sum_{\pi_y \in \widetilde{\mathcal{Y}}} \psi(T,n,\pi_y) P_{Y|\Pi_X}(y|\pi_x) \log_2 \left(\frac{P_{Y|\Pi_X}(y|\pi_x)}{P_Y(y)} \right),$$

for some matrix y such that $\langle y\rangle = \pi_y$. Remember that $\psi(T,n,\pi_y)$ is defined in Definition 3, §II. Defining $P_{\Pi_Y|\Pi_X}(\pi_y|\pi_x) \triangleq \psi(T,n,\pi_y)\ P_{Y|\Pi_X}(y|\pi_x)\big|_{\text{for some } y:\langle y\rangle=\pi_y}$, we can write

$$I(X;Y) = \sum_{\pi_x \in \widetilde{\mathcal{X}}, \pi_y \in \widetilde{\mathcal{Y}}} P_{\Pi_X}(\pi_x) P_{\Pi_Y|\Pi_X}(\pi_y|\pi_x) \log_2 \frac{P_{\Pi_Y|\Pi_X}(\pi_y|\pi_x)}{P_{\Pi_Y}(\pi_y)} = I(\Pi_X;\Pi_Y).$$

Based on the above discussion going back from the channel Ch_s to Ch_m is very easy. It is sufficient to choose

$$P_X(x) = \frac{P_{\Pi_X}(\pi_x)}{\psi(T,m,\pi_x)}, \quad \forall x:\ \langle x\rangle = \pi_x,$$

for all $\pi_x \in \widetilde{\mathcal{X}}$. This completes the proof. ■

Proof of Lemma 2: We want to count the number of different matrices $\mathbf{X} \in \mathbb{F}_q^{n\times T}$ such that $\langle \mathbf{X}\rangle = \pi_d$ where π_d is an specific d dimensional subspace of $\mathbb{F}_q^T$.

We know that we can decompose $\mathbf{X}$ as

$$\mathbf{X} = \mathbf{A}\mathbf{B}, \quad \mathbf{A} \in \mathbb{F}_q^{n\times d}, \mathbf{B} \in \mathbb{F}_q^{d\times T},$$

where $\mathbf{A}$ and $\mathbf{B}$ are full rank matrices. Let us fix $\mathbf{B}$ such that $\langle \mathbf{B}\rangle = \pi_d$. Now for every two different full rank matrices $\mathbf{A}$ and $\mathbf{A}'$ we would obtain different matrices $\mathbf{X} = \mathbf{A}\mathbf{B}$ and $\mathbf{X}' = \mathbf{A}'\mathbf{B}$ such that $\mathbf{X} \neq \mathbf{X}'$ and $\langle \mathbf{X}\rangle = \langle \mathbf{X}'\rangle = \pi_d$. So the number of different $\mathbf{X}$ where $\langle \mathbf{X}\rangle = \pi_d$ is equal to the number of full rank $n \times d$ matrices over $\mathbb{F}$ which is equal to $\prod_{i=0}^{d-1}(q^n - q^i)$, and we are done. ■

Proof of Lemma 8: Let $P_{\Pi_X}(\pi_x)$ be the optimal input distribution of the channel Ch_s with transition probabilities given in (6). For a fixed dimension $0 \leq d \leq \min[m,T]$, and an arbitrary permutation

$$\sigma : \left\{1, 2, \ldots, \begin{bmatrix} T \\ d \end{bmatrix}\right\} \to \left\{1, 2, \ldots, \begin{bmatrix} T \\ d \end{bmatrix}\right\}$$

which acts on subspaces of dimension d, define $P_\sigma(\pi_x)$ as

$$P_\sigma(\pi_x) = \begin{cases} P_{\Pi_X}(\sigma(\pi_x)) & \text{if } \dim(\pi_x) = d, \\ P_{\Pi_X}(\pi_x) & \text{if } \dim(\pi_x) \neq d. \end{cases}$$

Also define $P^*(\pi_x) = \frac{1}{\begin{bmatrix} T \\ d \end{bmatrix}!} \sum_\sigma P_\sigma(\pi_x)$ where the summation is over all possible permutations. Rewriting the mutual information in (17) as a function of the input distribution and the transition probabilities,

$I(P_{\Pi_X}(\pi_x), P_{\Pi_Y|\Pi_X}(\pi_y|\pi_x))$, we have

$$\begin{aligned} I(P^*(\pi_x),&P_{\Pi_Y|\Pi_X}(\pi_y|\pi_x)) \\ &= I\left(\frac{1}{\begin{bmatrix}T\\d\end{bmatrix}!}\sum_{\sigma} P_\sigma(\pi_x), P_{\Pi_Y|\Pi_X}(\pi_y|\pi_x)\right) \\ &\overset{(a)}{\geq} \frac{1}{\begin{bmatrix}T\\d\end{bmatrix}!}\sum_{\sigma} I(P_\sigma(\pi_x), P_{\Pi_Y|\Pi_X}(\pi_y|\pi_x)) \\ &\overset{(b)}{=} I(P_{\Pi_X}(\pi_x), P_{\Pi_Y|\Pi_X}(\pi_y|\pi_x)) \end{aligned}$$

where (a) is due to concavity of the mutual information with respect to the input distribution, and (b) holds because $I(P_\sigma(\pi_x), P_{\Pi_Y|\Pi_X}(\pi_y|\pi_x)) = I(P_{\Pi_X}(\pi_x), P_{\Pi_Y|\Pi_X}(\pi_y|\pi_x))$ for all σ, since the permutation only permutes the terms in a summation in (17).

Note that $P^*(\pi_x)$ assigns equal probabilities to all subspaces with dimension d, and the above-mentioned inequality shows that it is as good as the optimal input distribution. A similar argument holds for all $0 \leq d \leq \min[m, T]$. Therefore, a dimensional-uniform distribution achieves the capacity of the channel. ■

Proof of Lemma 9: Assuming an optimal input probability distribution of the form (18), the probability of receiving a specific subspace $\Pi_Y = \pi_y$ at the receiver can be written as

$$\begin{aligned} P_{\Pi_Y}(\pi_y) &= \sum_{\pi_x \in \widetilde{\mathcal{X}}} P_{\Pi_Y|\Pi_X}(\pi_y|\pi_x) P_{\Pi_X}(\pi_x) \\ &= \sum_{\substack{\pi_x \in \widetilde{\mathcal{X}}, \\ \pi_y \sqsubseteq \pi_x}} \psi(T, n, \pi_y) q^{-nd_x} \frac{\alpha_{d_x}}{\begin{bmatrix}T\\d_x\end{bmatrix}}. \end{aligned}$$

Splitting the summation into two, we can write

$$P_{\Pi_Y}(\pi_y) = \psi(T, n, \pi_y) \sum_{d_x = d_y}^{\min[m,T]} \sum_{\substack{\pi_x \in \widetilde{\mathcal{X}}, \\ \dim(\pi_x) = d_x, \\ \pi_y \sqsubseteq \pi_x}} \frac{q^{-nd_x}\alpha_{d_x}}{\begin{bmatrix}T\\d_x\end{bmatrix}}, \tag{49}$$

where $d_y = \dim(\pi_y)$. Using the following result, Lemma 16, we can replace the second summation in (49).

***Lemma** 16:* Let π_y be a fixed subspace of $\mathbb{F}_q^T$ with dimension d_y. Then the number of different subspaces $\pi_x \in \mathbb{F}_q^T$ with dimension d_x, $d_y \leq d_x \leq T$, that contain π_y is equal to $\begin{bmatrix}T-d_y\\d_x-d_y\end{bmatrix}$.

Proof: This lemma can be proved by applying [24, Lemma 2] with proper choice of the parameters.

■

Using Lemma 16 we can rewrite (49) as

$$P_{\Pi_Y}(\pi_y) = \psi(T,n,\pi_y) \sum_{d_x=d_y}^{\min[m,T]} \begin{bmatrix} T-d_y \\ d_x-d_y \end{bmatrix} \frac{q^{-nd_x}\alpha_{d_x}}{\begin{bmatrix} T \\ d_x \end{bmatrix}}$$

$$\stackrel{(a)}{=} \frac{\psi(T,n,\pi_y)}{\begin{bmatrix} T \\ d_y \end{bmatrix}} \sum_{d_x=d_y}^{\min[m,T]} \begin{bmatrix} d_x \\ d_y \end{bmatrix} q^{-nd_x}\alpha_{d_x}$$

$$= \frac{\psi(n,d_y)}{\begin{bmatrix} T \\ d_y \end{bmatrix}} \sum_{d_x=d_y}^{\min[m,T]} \begin{bmatrix} d_x \\ d_y \end{bmatrix} q^{-nd_x}\alpha_{d_x}, \tag{50}$$

where (a) follows from the following result, Lemma 17.

***Lemma** 17:* The following relation for the Gaussian number holds [26], [25]

$$\begin{bmatrix} T-d_y \\ d_x-d_y \end{bmatrix} \begin{bmatrix} T \\ d_y \end{bmatrix} = \begin{bmatrix} T \\ d_x \end{bmatrix} \begin{bmatrix} d_x \\ d_y \end{bmatrix}.$$

Now we can simplify the mutual information $I(\Pi_X;\Pi_Y)$ in (17) as follows. Using (6), (18), and (50) for $I(\Pi_X;\Pi_Y)$ we can write

$$I(\Pi_X;\Pi_Y) = \sum_{\pi_x\in\widetilde{\mathcal{X}},\pi_y\in\widetilde{\mathcal{Y}}} P_{\Pi_X}(\pi_x)P_{\Pi_Y|\Pi_X}(\pi_y|\pi_x)\log_2\left(\frac{P_{\Pi_Y|\Pi_X}(\pi_y|\pi_x)}{P_{\Pi_Y}(\pi_y)}\right)$$

$$= \sum_{d_x=0}^{\min[m,T]} \sum_{d_y=0}^{\min[n,d_x]} \sum_{\substack{\pi_x\in\widetilde{\mathcal{X}},\\ \dim(\pi_x)=d_x}} \sum_{\substack{\pi_y\in\widetilde{\mathcal{Y}},\\ \dim(\pi_y)=d_y,\\ \pi_y\sqsubseteq\pi_x}} \frac{\alpha_{d_x}\psi(n,d_y)q^{-nd_x}}{\begin{bmatrix} T \\ d_x \end{bmatrix}} \log_2\left(\frac{q^{-nd_x}}{f(d_y)}\right),$$

where

$$f(d_y) \triangleq \frac{P_{\Pi_Y}(\pi_y)}{\psi(n,d_y)} = \frac{1}{\begin{bmatrix} T \\ d_y \end{bmatrix}} \sum_{d_x=d_y}^{\min[m,T]} \begin{bmatrix} d_x \\ d_y \end{bmatrix} q^{-nd_x}\alpha_{d_x}, \tag{51}$$

because $P_{\Pi_Y}(\pi_y)$ only depends on d_y. Now observe that the two inner most summations depend on π_x and π_y only through their dimensions. So we can write

$$I(\Pi_X;\Pi_Y) = \sum_{d_x=0}^{\min[m,T]} \alpha_{d_x}q^{-nd_x} \sum_{d_y=0}^{\min[n,d_x]} \psi(n,d_y)\begin{bmatrix} d_x \\ d_y \end{bmatrix} \log_2\left(\frac{q^{-nd_x}}{f(d_y)}\right).$$

Then using Lemma 4 in §II-B we can further simplify the mutual information and write

$$I(\Pi_X;\Pi_Y) = -\sum_{d_x=0}^{\min[m,T]} \alpha_{d_x} nd_x \log_2 q$$

$$- \sum_{d_x=0}^{\min[m,T]} \alpha_{d_x}q^{-nd_x} \sum_{d_y=0}^{\min[n,d_x]} \psi(n,d_y)\begin{bmatrix} d_x \\ d_y \end{bmatrix} \log_2(f(d_y)), \tag{52}$$

that is the assertion of Lemma 9. ■

Proof of Lemma 10: By taking the partial derivative of the mutual information with respect to α_k, we have that

$$\begin{aligned} I'_k \triangleq & \frac{\partial I(\Pi_X;\Pi_Y)}{\partial \alpha_k} \\ = & -nk\log_2 q - \sum_{d_y=0}^{\min[n,k]} \psi(n,d_y) \begin{bmatrix} k \\ d_y \end{bmatrix} q^{-nk} \log_2\left(f(d_y)\right) \\ & - \sum_{d_x=0}^{\min[m,T]} \alpha_{d_x} \sum_{d_y=0}^{\min[n,d_x,k]} \psi(n,d_y) \begin{bmatrix} d_x \\ d_y \end{bmatrix} q^{-nd_x} \frac{\begin{bmatrix} k \\ d_y \end{bmatrix} q^{-nk} \log_2 e}{\begin{bmatrix} T \\ d_y \end{bmatrix} f(d_y)}. \end{aligned}$$

$$\begin{aligned} I'_k = & -nk\log_2 q - \sum_{d_y=0}^{\min[n,k]} \psi(n,d_y) \begin{bmatrix} k \\ d_y \end{bmatrix} q^{-nk} \log_2\left(f(d_y)\right) \\ & - \sum_{d_y=0}^{\min[n,k]} \frac{\begin{bmatrix} k \\ d_y \end{bmatrix} \psi(n,d_y) q^{-nk}}{f(d_y)} \underbrace{\sum_{d_x=d_y}^{\min[m,T]} \alpha_{d_x} \frac{\begin{bmatrix} d_x \\ d_y \end{bmatrix}}{\begin{bmatrix} T \\ d_y \end{bmatrix}} q^{-nd_x}}_{f(d_y)} \log_2 e \\ \stackrel{(a)}{=} & -nk\log_2 q - \sum_{d_y=0}^{\min[n,k]} \psi(n,d_y) \begin{bmatrix} k \\ d_y \end{bmatrix} q^{-nk} \log_2\left(f(d_y)\right) - \log_2 e, \end{aligned}$$

where to derive (a) we use Lemma 4 in §II-B. ■

Proof of Lemma 11: For convenience we rewrite (24) again

$$\log_2\left(f(d_y)\right) = -d_y T \log_2 q + O(q^{-1}) + \log_2\left(\sum_{d_x=d_y}^{\min[m,T]} q^{-(n-d_y)d_x} \alpha_{d_x}\right). \tag{53}$$

We prove the assertion in two steps for every k. First, let us assume that the α_i's are such that we have $\log_2\left(f(\min[n,k])\right) = o(q)$. Then using (53) one can conclude that

$$\sum_{d_x=\min[n,k]}^{\min[m,T]} q^{-(n-d_y)d_x} \alpha_{d_x} = 2^{-o(q)},$$

so we should have $\alpha_i = 2^{-o(q)}$ for $\min[n,k] \leq i \leq \min[m,T]$. We know that $0 \leq \alpha_i \leq 1$, and $\sum_{i=0}^{\min[m,T]} \alpha_i = 1$, so $\exists j:\ \alpha_j = \Omega(1)$. So we can deduce that

$$\log_2(f(d_y)) = \begin{cases} o(q) & j < d_y \leq \min[n,k], \\ \Theta(\log q) & 0 \leq d_y \leq j, \end{cases}$$

where j, $0 \leq j \leq \min[n,k]$, is the largest index such that $\alpha_j = \Omega(1)$. So in this case the dominating term in the summation of (23) is the one obtained for $d_y = \min[n,k]$ because the order difference between each term inside the summation of (23) is at least of order $\Theta(q)$.

Now, for the second case, let us assume that the α_i's are such that we have $\log_2\left(f(\min[n,k])\right) = \Omega(q)$. We will show that this assumption leads to a contradiction. Using (53) we can write

$$\sum_{d_x=\min[n,k]}^{\min[m,T]} q^{-(n-d_y)d_x}\alpha_{d_x} = 2^{-\Omega(q)},$$

so we should have $\alpha_i = 2^{-\Omega(q)}$ for $\min[n,k] \leq i \leq \min[m,T]$. As before, we find the asymptotic behavior of $\log_2(f(d_y))$ for different values of d_y but in this case we should make finer regimes for $\log_2(f(d_y))$. The asymptotic behavior of α_i, $0 \leq i \leq \min[n,k]$, is either $2^{-\Omega(q)}$ or $2^{-o(q)}$. So we can write

$$\log_2(f(d_y)) = \begin{cases} \Omega(q) & l < d_y \leq \min[n,k], \\ o(q) & j < d_y \leq l, \\ \Theta(\log q) & 0 \leq d_y \leq j, \end{cases}$$

where l, $0 \leq l \leq \min[n,k]$, is the largest index such that $\alpha_i = 2^{-o(q)}$ which means that $\alpha_i = 2^{-\Omega(q)}$ for $l < i \leq \min[m,T]$. As before j, $0 \leq j \leq \min[n,k]$, is the largest index such that $\alpha_j = \Omega(1)$. Now we check the Kuhn-Tucker conditions, (21), for I'_k and I'_j. From the above argument we have that $I'_k = \Omega(q)$ and $I'_j = \Theta(\log q)$. We know that $\alpha_j = \Omega(1) > 0$, so we have $I'_j = \Theta(\log q) = \lambda$. On the other hand, we have $I'_k = \Omega(q) \leq \lambda$, which is a contradiction implying the second case cannot occur. This completes the proof. ■

Proof of Lemma 12: There are $\begin{bmatrix} d_1 \\ d_{12} \end{bmatrix} \doteq q^{d_{12}(d_1-d_{12})}$ different choices for the intersection of π_1 and π_2. We have to choose $d_2 - d_{12}$ basis vectors for the rest of the subspace. This can be done in

$$\frac{\left(q^T - q^{d_1}\right)\left(q^T - q^{d_1+1}\right)\ldots\left(q^T - q^{d_1+d_2-d_{12}-1}\right)}{\left(q^{d_2} - q^{d_{12}}\right)\left(q^{d_2} - q^{d_{12}+1}\right)\ldots\left(q^{d_2} - q^{d_2-1}\right)} \doteq q^{(d_2-d_{12})(T-d_2)}$$

ways. So we have $n(d_1,d_2,d_{12}) \doteq q^{d_{12}(d_1-d_{12})+(d_2-d_{12})(T-d_2)}$. The proof follows from the results in [24, Lemma 2], by proper choice of parameters. Independently, an alternate proof of this lemma appeared in our paper [17]. ■

Proof of Lemma 13: Define $\pi = \pi_1 + \pi_2$, where $\dim(\pi) = \dim(\pi_1) + \dim(\pi_2) - \dim(\pi_1 \cap \pi_2) = d_1 + d_2 - d_{12} \triangleq d$. The proof of this lemma is similar to that of Lemma 12, unless we can only choose the last $d_2 - d_{12}$ basis vectors from π instead of $\mathbb{F}_q^T$. Therefore replacing T in Lemma 12 with d, we have $a(\pi_1,\pi_2) \doteq q^{d_{12}(d_1-d_{12})+(d_2-d_{12})(d-d_2)} = q^{d_2(d_1-d_{12})}$. ■

Proof of Lemma 14: Let (R_1,R_2) be a corner point of the region $\mathcal{R}_{\text{col}}$. Since $\mathcal{R}_{\text{col}}$ is the convex hull of a set of primitive regions, there should exist a primitive region $\mathcal{R}(d_1,d_2)$ which contains (R_1,R_2) as a corner point, *i.e.*,

$$\exists (d_1,d_2) \in \mathcal{D}_{\text{col}}, \quad (R_1,R_2) = (R_1(d_1,d_2), R_2(d_1,d_2)).$$

We will show that any point $(R_1(d_1,d_2), R_2(d_1,d_2))$ is dominated by the segment connecting $(R_1(d_1+1,d_2), R_2(d_1+1,d_2))$ and $(R_1(d_1,d_2+1), R_2(d_1,d_2+1))$. In order to show that, we have to prove that there exists some $\lambda \in [0,1]$, such that

$$\begin{aligned} R_1(d_1,d_2) &< \lambda R_1(d_1+1,d_2) + (1-\lambda) R_1(d_1,d_2+1), \\ R_2(d_1,d_2) &< \lambda R_2(d_1+1,d_2) + (1-\lambda) R_2(d_1,d_2+1). \end{aligned} \tag{54}$$

After a little simplification, (54) can be rewritten as

$$\lambda[T-d_1-d_2-1] < d_1,$$

$$(1-\lambda)[T-d_1-d_2-1] < d_2,$$

$$\text{or} \quad \frac{d_1}{T-1-d_1-d_2} < \lambda < \frac{T-1-d_1-2d_2}{T-1-d_1-d_2}.$$

The last two inequalities can be satisfied for some choice of λ if and only if $d_1+d_2 < (T-1)/2$. Therefore, if we have $d_1 < m_1$, $d_2 < m_2$, and $d_1+d_2 < (T-1)/2$ for some $(d_1,d_2) \in \mathcal{D}_{\text{col}}$, then (d_1+1,d_2) and (d_1,d_2+1) also belong to $\mathcal{D}_{\text{col}}$, and hence, $(R_1(d_1,d_2), R_2(d_1,d_2))$ is an interior point, and cannot be on the boundary of the region. Eliminating such (d_1,d_2) from $\mathcal{D}_{\text{col}}$, we get $\widetilde{\mathcal{D}}$.

It is also easy to show that all of the rate pairs corresponding to $(d_1,d_2) \in \widetilde{\mathcal{D}}$ are on the boundary of $\mathcal{R}_{\text{col}}$. This can be done by comparing the slope of the connecting segment for two consecutive points (according to the order they are appeared in $\widetilde{\mathcal{D}}$). The slopes are

$$\begin{aligned} &\mathcal{S}\{(R_1(t,m_2), R_2(t,m_2)); (R_1(t+1,m_2), R_2(t+1,m_2))\} \\ &\quad = -\frac{m_2}{T-2t-m_2-1} \qquad \text{for } 0 \le t \le m_1 \\ &\mathcal{S}\{(R_1(m_1,t), R_2(m_1,t)); (R_1(m_1,t-1), R_2(m_1,t-1))\} \\ &\quad = -\frac{T-2t-m_1-1}{m_1} \qquad \text{for } 1 \le t \le m_2. \end{aligned}$$

It is easy to check that all the slopes are negative and they are in a decreasing order. Therefore, no point in the set $\widetilde{\mathcal{D}}$ can be an interior point. ■

Proof of Lemma 15: Note that $\mathcal{R}_{\text{col}} \nsubseteq \mathcal{R}_{\text{coop}}$ implies $m_1+m_2 > n$. Since $\mathcal{R}_{\text{col}}$ is a convex region, its boundary intersects with the line $R_1+R_2 = n(T-n)\log_2 q$ in exactly two points (it cannot be only one point, otherwise it would be inside of $\mathcal{R}_{\text{coop}}$). It is easy to verify that the rate points corresponding to $(d_1,d_2) = ((n-m_2)^+, \min[m_2,n])$ and $(d_1,d_2) = (\min[m_1,n], (n-m_1)^+)$ lie on both the boundary of $\mathcal{R}_{\text{col}}$ and the line $R_1+R_2 = n(T-n)\log_2 q$. Therefore this line cannot intersect with the boundary of $\mathcal{R}_{\text{col}}$ in any other point. ■

APPENDIX B
EXTENSION TO PACKET ERASURE NETWORKS

Let us write the capacity for the erasure case as follows

$$\begin{aligned} C_e &= \max_{P_X} I(X;Y,N) \\ &= \max_{P_X} \left[I(X;N) + I(X;Y|N) \right] \\ &\overset{(a)}{=} \max_{P_X} I(X;Y|N) \\ &= \max_{P_X} \mathbb{E}_N \left[I(X;Y) \right], \end{aligned}$$

where (a) follows from the independence of input distribution P_X and the distribution of the number of received packets P_N.

The Upper Bound:

We can write an upper bound for C_e as follows

$$\begin{aligned} C_e &= \max_{P_X} \mathbb{E}_N \left[[I(X;Y)] \right] \\ &\leq \mathbb{E}_N \left[\max_{P_X} I(X;Y) \right] \\ &= \mathbb{E}_N \left[i^*(T-i^*) \log_2 q \right], \end{aligned}$$

where $i^* = \min[m, N, \lfloor T/2 \rfloor]$. From here on let us assume that $m \leq \lfloor T/2 \rfloor$. We thus have that $i^* = N$ and we can write

$$C_e \leq \mathbb{E}_N \left[N(T-N) \log_2 q \right].$$

Let us define $\mu_1 \triangleq \mathbb{E}_N [N]$ and $\mu_2 \triangleq \mathbb{E}_N \left[N^2 \right]$ so we can write

$$C_e \leq (\mu_1 T - \mu_2) \log_2 q.$$

The Lower Bound:

For the lower bound we can write

$$\begin{aligned} C_e &= \max_{P_X} \mathbb{E}_N \left[[I(X;Y)] \right] \\ &\geq \mathbb{E}_N \left[I(X;Y) \right]_{\text{for some } P_X} \\ &= \mathbb{E}_N \left[I(\Pi_X;\Pi_Y) \right]_{\text{for some } P_{\Pi_X}}. \end{aligned}$$

From (19) we know that we can write

$$\begin{aligned} I(\Pi_X;\Pi_Y) = &- \sum_{d_x=0}^{\min[m,T]} \alpha_{d_x} N d_x \log_2 q \\ &- \sum_{d_x=0}^{\min[m,T]} \alpha_{d_x} q^{-Nd_x} \sum_{d_y=0}^{\min[N,d_x]} \psi(N,d_y) \begin{bmatrix} d_x \\ d_y \end{bmatrix} \log_2(f(d_y)), \end{aligned}$$

where

$$f(d_y) \triangleq \frac{1}{\begin{bmatrix} T \\ d_y \end{bmatrix}} \sum_{d_x=d_y}^{\min[m,T]} \begin{bmatrix} d_x \\ d_y \end{bmatrix} q^{-Nd_x} \alpha_{d_x}.$$

Now assume that $m \leq \lfloor T/2 \rfloor$ and choose the input distribution to be $\alpha_k = 1$ for some $0 \leq k \leq m$ and $\alpha_i = 0$ for all $i \neq k$. Then for this input distribution we have

$$\begin{aligned} I(\Pi_X;\Pi_Y) &= -kN \log_2 q - q^{-kN} \sum_{d_y=0}^{\min[N,k]} \psi(N,d_y) \begin{bmatrix} k \\ d_y \end{bmatrix} \log_2(f(d_y)) \\ &= -kN \log_2 q - q^{-kN} \sum_{d_y=0}^{\min[N,k]} \psi(N,d_y) \begin{bmatrix} k \\ d_y \end{bmatrix} \log_2(f(d_y)). \end{aligned}$$

Then assuming q is large we may approximate the above mutual information as follows

$$I(\Pi_X;\Pi_Y) \approx -kN \log_2 q - \sum_{d_y=0}^{\min[N,k]} q^{-(N-d_y)(k-d_y)} \log_2(f(d_y)).$$

The term $(N-d_y)(k-d_y)$ in the summation is maximized for $d_y = \min[N,k]$ and because we had shown before in Lemma 11 that $\log_2(f(d_y)) = \Theta(\log q)$, we can write

$$\begin{aligned} I(\Pi_X;\Pi_Y) &\approx -kN \log_2 q - \log_2(f(\min[N,k])) \\ &\approx -kN \log_2 q - \log_2 \left(q^{\min[N,k](k-T)-Nk} \right) \\ &= \min[N,k](T-k) \log_2 q. \end{aligned}$$

So by choosing $k = m$ we can write the lower bound for C_e as follows

$$\begin{aligned} C_e &\geq \mathbb{E}_N \left[I(\Pi_X;\Pi_Y) \right]_{\text{for some } P_{\Pi_X}} \\ &\approx \mathbb{E}_N \left[N(T-m) \log_2 q \right] \\ &= \mu_1 (T-m) \log_2 q. \end{aligned}$$

REFERENCES

[1] R. Ahlswede, N. Cai, S-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Transactions on Information Theory*, vol. 46, pp. 1204–1216, Jul. 2000.

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

[3] R. Koetter and M. Medard, "An algebraic approach to network coding," *IEEE/ACM Transaction on Networking*, vol. 11, no. 5, pp. 782–795, Oct. 2003.

[4] C. Fragouli and E. Soljanin, "Information flow decomposition for network coding", *IEEE Transactions on Information Theory*, vol. 52, no. 23, pp. 829-848, March 2006.

[5] T. Ho, R. Koetter, M. Medard, M. Effros, J. Shi, and D. Karger, "A random linear network coding approach to multicast," *IEEE Transactions on Information Theory*, vol. 52, pp. 4413–4430, Oct. 2006.

[6] R. Koetter and F. Kschischang, "Coding for errors and erasures in random network coding," *IEEE Transactions on Information Theory*, vol. 54, iss. 8, Aug. 2008.

[7] D. Silva, F. Kschischang and R. Koetter, "A Rank-Metric Approach to Error Control in Random Network Coding," *IEEE Transactions on Information Theory*, vol. 54, no. 9, pp. 3951–3967, Sep. 2008.

[8] M. Jafari Siavoshani, C. Fragouli, and S. Diggavi, "Passive topology discovery for network coded systems," *Information Theory Workshop*, Bergen, Norway, Jul. 2007.

[9] 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.

[10] 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.

[11] K. Price and R. Storn, "Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces," *Journal of Global Optimization*, vol. 11, pp. 341–359, 1997.

[12] P. A. Chou, Y. Wu, and K. Jain, "Practical network coding," *Allerton Conference on Communication, Control, and Computing*, IL, Oct. 2003.

[13] L. Keller, M. Jafari Siavoshani, C. Fragouli, K. Argyraki, and S. Diggavi, "Joint identity-message coding for sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 7, pp. 1083–1093, Sep. 2010. See also *Proc. INFOCOM*, pp. 2177–2185, 2009.

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

[15] T. Cover and J. Thomas, "Elements of Information Theory," Wiley & Sons, New York, Second edition, 2006.

[16] D. Silva, F. R. Kschischang, and R. Koetter, "Communication over finite-field matrix channels," *IEEE Transactions on Information Theory*, vol. 56, iss. 3, pp. 1296–1305, Mar. 2010.

[17] S. Mohajer, M. Jafari Siavoshani, S. N. Diggavi, C. Fragouli, "On the capacity of multisource non-coherent network coding", *Information Theory Workshop*, pp. 130–134, Jun. 2009.

[18] S. Boyd and L. Vandenberghe, "Convex Optimization," *Cambridge University Press*, 2004.

[19] L. Zheng and D. N. C. Tse, "Communication on the Grassmannian manifold: A geometric approach to the non-coherent multiple-antenna channel," *IEEE Transaction on Information Theory*, vol. 48, pp. 359–383, Feb. 2002.

[20] P. Sattari, A. Markopoulou, C. Fragouli, "Multiple source multiple destination topology inference using network coding," *The Workshop on Network Coding, Theory and Applications*, Lausanne, Jun. 2009.

[21] G. Sharma, S. Jaggi, B. K. Dey, "Network Tomography via Network Coding," *Information Theory and Application Workshop*, UCSD, 2007.

[22] J. H. van Lint, R. M. Wilson, "A course in combinatorics," *Cambridge University Press*, Second Edition, 2001.

[23] M. Gadouleau and Z. Yan, "On the decoder error probability of bounded rank-distance decoders for maximum rank distance codes," *IEEE Transactions on Information Theory*, vol. 54, no. 7, pp. 3202–3206, Jul. 2008.

[24] M. Gadouleau and Z. Yan, "Packing and Covering Properties of Subspace Codes for Error Control in Random Linear Network Coding," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2097–2108, May 2010.

[25] G. Andrews, "The theory of partitions," *Encyclopedia of Mathematics and its Applications*, 1976.

[26] E. Gabidulin, "Theory of codes with maximum rank distance," *Problems of Information Transmission*, vol. 21, no. 1, pp. 1–12, Jan. 1985.

[27] Tinyos. http://www.tinyos.net/.

PLACE
PHOTO
HERE

Mahdi Jafari Siavoshani received the Bachelor degree in Communication Systems with a minor in Applied Physics at Sharif University of Technology, Tehran, Iran, in 2005. He was awarded an Excellency scholarship from EPFL, Switzerland, to study a master degree in Communication System finished in 2007. He is currently a PhD student at the same university. His research interests include network coding, coding and information theory, wireless communications, and signal processing.

PLACE
PHOTO
HERE

Soheil Mohajer received the B.S. degree in electrical engineering from the Sharif University of Technology, Tehran, Iran, in 2004, and the M.S. degrees in communication systems from Ecole Polytechnique Fdrale de Lausanne (EPFL), Lausanne, Switzerland, in 2005. He completed his Ph.D. at EPFL in September 2010, and since October 2010 is a post-doctoral researcher at Princeton University. His fields of interests are multiuser information theory, network coding theory, and wireless communication.

PLACE
PHOTO
HERE

Christina Fragouli is a tenure-track Assistant Professor in the School of Computer and Communication Sciences, EPFL, Switzerland. She received the B.S. degree in Electrical Engineering from the National Technical University of Athens, Athens, Greece, in 1996, and the M.Sc. and Ph.D. degrees in electrical engineering from the University of California, Los Angeles, in 1998 and 2000, respectively. She has worked at the Information Sciences Center, AT&T Shannon Labs, Florham Park New Jersey, and the National University of Athens. She also visited Bell Laboratories, Murray Hill, NJ, and DIMACS, Rutgers University. From 2006 to 2007, she was an FNS Assistant Professor in the School of Computer and Communication Sciences, EPFL, Switzerland. She served as an editor for IEEE Communications Letters. She is currently serving as an editor for IEEE Transactions on Information Theory, IEEE Transactions on Communications, Elsevier Computer Communications and IEEE Transactions on Mobile Computing. She was the technical co-chair for the 2009 Network coding symposium in Lausanne and has served on program commmittees of several conferences. She received the Fulbright Fellowship for her graduate studies, the Outstanding Ph.D. Student Award 2000-2001, UCLA, Electrical Engineering Department, the Zonta award 2008 in Switzerland, and the Young Investigator ERC starting grant in 2009. Her research interests are in network information flow theory and algorithms, network coding, and connections between communications and computer science.

PLACE
PHOTO
HERE

Suhas N. Diggavi received the B. Tech. degree in electrical engineering from the Indian Institute of Technology, Delhi, India, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 1998.

After completing his Ph.D., he was a Principal Member Technical Staff in the Information Sciences Center, AT&T Shannon Laboratories, Florham Park, NJ. After that he was on the faculty at the School of Computer and Communication Sciences, EPFL, where he directed the Laboratory for Information and Communication Systems (LICOS). He is currently a Professor, in the Department of Electrical Engineering, at the University of California, Los Angeles. His research interests include wireless communications networks, information theory, network data compression and network algorithms.

He is a recipient of the 2006 IEEE Donald Fink prize paper award, 2005 IEEE Vehicular Technology Conference best paper award and the Okawa foundation research award. He is currently an editor for ACM/IEEE Transactions on Networking and IEEE Transactions on Information Theory. He has 8 issued patents.