Capacity Analysis of Linear Operator Channels Over Finite Fields

Shenghao Yang, *Member, IEEE*, Siu-Wai Ho, *Member, IEEE*, Jin Meng, and En-Hui Yang, *Fellow, IEEE*

***Abstract*—Motivated by communication through a network employing linear network coding, capacities of linear operator channels (LOCs) with arbitrarily distributed transfer matrices over finite fields are studied. Both the Shannon capacity C and the subspace coding capacity C_{SS} are analyzed. By establishing and comparing lower bounds on C and upper bounds on C_{SS}, various necessary conditions and sufficient conditions such that $C = C_{\mathrm{SS}}$ are obtained. A new class of LOCs such that $C = C_{\mathrm{SS}}$ is identified, which includes LOCs with uniform-given-rank transfer matrices as special cases. It is also demonstrated that C_{SS} is strictly less than C for a broad class of LOCs. In general, an optimal subspace coding scheme is difficult to find because it requires to solve the maximization of a nonconcave function. However, for an LOC with a unique subspace degradation, C_{SS} can be obtained by solving a convex optimization problem over rank distribution. Classes of LOCs with a unique subspace degradation are characterized. Since LOCs with uniform-given-rank transfer matrices have unique subspace degradations, some existing results on LOCs with uniform-given-rank transfer matrices are explained from a more general way.**

***Index Terms*—Linear operator channel, network coding, subspace coding.**

I. INTRODUCTION

FIX a finite field $\mathbb{F}$ with q elements. A linear operator channel (LOC), also called a multiplicative matrix channel, with input random variable $X \in \mathbb{F}^{T \times M}$ and output random variable $Y \in \mathbb{F}^{T \times N}$ is given by

$$Y = XH, \tag{1}$$

where $H \in \mathbb{F}^{M \times N}$ is called a *transfer matrix*. We assume that X and H are independent, and the transfer matrices in different channel uses are independent and follow the same distribution. For both the transmitter and receiver, the distribution of H is given a priori, but the instances of H are unknown.

A LOC is used to model communication through a network employing linear network coding [1], [2]. Consider a network coding scenario where the source node encodes its message into batches (also called generations, classes or chunks), each of which contains M packets of T symbols [3], [4]. Intermediate network nodes generate new packets by taking linear combinations of the packages among the same batch. There may be packet loss and network topological dynamics during the transmission. The finally received N packets of a batch are all linear combinations of the original packets of the batch. Such a network transmission can be modeled by a LOC.

Coding problems for LOCs have been studied for various scenarios. If T is much larger than M, parts of X can be used to transmit an identity matrix so that the receiver can recover the instances of H. Such a scheme, called *channel training*, has been widely used for random linear network coding [5] and is asymptotically optimal when T goes to infinity. The maximum achievable rate of channel training (by multiple uses of the channel) can be achieved using random linear codes [6], and a channel training scheme with low encoding/decoding complexity has been proposed [7], [8] by generalizing fountain codes. However, if T is not much larger than M, the overhead used to explicitly recover the instances of H is dominating, and hence different coding schemes must be studied.

We call the vector space spanned by the column vectors of a matrix $\mathbf{X}$ the column space of the matrix, denoted by $\langle \mathbf{X} \rangle$. For a LOC, with probability one $\langle Y \rangle$ is a subspace of $\langle X \rangle$. Koetter and Kschischang [9] defined a channel with subspaces as input and output to capture this property, and discussed subspace codes for one use of this subspace channel. They defined the minimum distance of a subspace code in terms of a subspace distance between codewords, and used the minimum distance to characterize the error (or erasure) correction capability of the subspace code. Thereafter, subspace coding has generated a lot of research interests (see [10]–[12]) and the study of subspace coding has also been extended from one use to multiple uses of the channel [13], [14].

In this paper, we are interested in the achievable rates of coding schemes when the error probability goes to zero asymptotically. Most existing works on subspace coding try to design large codebooks with large minimum distances.

Manuscript received December 31, 2012; revised June 17, 2013; accepted February 14, 2014. Date of publication May 29, 2014, date of current version July 10, 2014. S. Yang was supported in part by the National Basic Research Program of China under Grant 2011CBA00300 and Grant 2011CBA00301, and in part by the National Natural Science Foundation of China under Grant 61033001 and Grant 61361136003. S.-W. Ho was supported by the Australian Research Council through an Australian Post-Doctoral Fellowship, Discovery Project, under Grant DP1094571. J. Meng and E.-H. Yang were supported in part by the Natural Sciences and Engineering Research Council of Canada under Grant RGPIN203035-11 and in part by the Canada Research Chairs Program. This paper was presented at the 2010 IEEE Information Theory Workshop and the 2010 IEEE International Symposium on Information Theory.

S. Yang is with the Institute for Theoretical Computer Science, Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing 100084, China (e-mail: shyang@tsinghua.edu.cn).

S.-W. Ho is with the Institute for Telecommunications Research, University of South Australia, Adelaide, SA 5000, Australia (e-mail: siuwai.ho@unisa.edu.au).

J. Meng and E.-H. Yang are with the Department of Electrical and Computer Engineering, Waterloo University, Waterloo, ON N2L 3G1, Canada (e-mail: j4meng@uwaterloo.ca; ehyang@uwaterloo.ca).

Communicated by J. Belfiore, Associate Editor for Coding Theory.

Digital Object Identifier 10.1109/TIT.2014.2326976

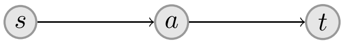

Fig. 1. In this network, s is the source node, t is the destination node, and a is the intermediate node that does not demand the file.

However, subspace codes designed under the minimum distance criteria may not have a good performance for multiple uses of a LOC [6].

Towards better understanding of the coding problems and identifying new directions to study coding for LOCs, an information theoretic study of LOCs becomes necessary. Existing works have studied several classes of distributions of H. When $M = N$, Silva *et al.* [15] studied the case that H is uniformly chosen from all full rank $M \times M$ matrices. Siavoshani *et al.* [16] studied the case that H contains uniformly i.i.d. components. Nóbrega *et al.* [17], [18] studied LOCs with *uniform-given-rank* transfer matrices, which include the transfer matrices studied in [15] and [16] as special cases. For all the above special distributions of H, it is shown that $I(X;Y) = I(\langle X\rangle;\langle Y\rangle)$ for any input X, which in turn implies that using subspaces for encoding and decoding indeed achieves the Shannon capacity of these special LOCs; in addition, the Shannon capacity of these LOCs can be found by maximizing over input rank distribution.

However, many typical scenarios in linear network coding cannot be covered by those special cases studied in the existing literature. Even though the transfer matrix is full rank with high probability for random linear network coding when both the field size and the maximum flow from the source node to the destination node are sufficiently large [5], such a transfer matrix may not have the uniform distribution studied in [15]. The transfer matrix studied in [16] can be formed by using random linear network coding in the intermediate node in Fig. 1, where node a caches M packets transmitted by node s before encoding, and transmits N independent random linear combinations of these M packets. But if we take the packet loss during the transmission on both links into consideration, the transfer matrix will not have independent components since a packet loss will force a row/column to be zero. Moreover, encoding after collecting M packets introduces delay, so it is more practical to apply random linear network coding in a causal way: the intermediate node keeps transmitting the linear combinations of the packets it has received [3], [19], which results in a transfer matrix of the form (take $M = 4$ as an example)

$$\begin{bmatrix} h_{1,1} & h_{1,2} & h_{1,3} & h_{1,4} & h_{1,5} & \cdots \\ 0 & h_{2,2} & h_{2,3} & h_{2,4} & h_{2,5} & \cdots \\ 0 & 0 & 0 & h_{3,4} & h_{3,5} & \cdots \\ 0 & 0 & 0 & 0 & h_{4,5} & \cdots \end{bmatrix},$$

where i) all nonzero rows are above any rows of all zeros, ii) the leading coefficient (the first nonzero component from the left) of a nonzero row is not to the left of the leading coefficient of the row above it, iii) all nonzero components are i.i.d. over a finite field. But such a transfer matrix is even not uniform-given-rank. Furthermore, subspace coding is not capacity achieving in general. For example, when H is an $M \times M$ identity matrix and $T = 1$, the Shannon capacity is $M \log q$ bits per use and the subspace coding capacity is 1 bit.

In this paper, we are motivated to study LOCs with arbitrarily distributed transfer matrices. We analyze both the Shannon capacity and subspace coding capacity of LOCs, and we try to answer the following questions: How to achieve or approach the Shannon capacity of a LOC? What is the performance of subspace coding and when is subsapce coding optimal? How to design subspace coding for general LOCs? Our results are for general values of T, M, N and q.

We first discuss some symmetry properties of LOCs, which lead to the discovery that there exists a *uniform-given-row-space* input distribution achieving the Shannon capacity C of a LOC (Theorem 1). We then derive an upper bound and a lower bound on the Shannon capacity C, where the lower bound is tight for *row-space-symmetric* LOCs (Theorem 2) and is in general at least as good as the lower bound derived using uniform-given-rank transfer matrices in [17] and [18].

We then turn our attention to the subspace coding capacity C_{SS} of a LOC. Note that a LOC has matrices as input and output, while subspace coding uses subspaces for encoding and decoding. A general way to study subspace coding for a LOC is to look at a subspace degradation of the LOC, which is induced by a transition probability from subspaces to matrices. The subspace degradations induced by a LOC are not unique in general, and finding an optimal subspace degradation involves maximizing a non-concave function, which is in general difficult to solve. We study subspace coding with uniform-given-row-space input distributions to obtain a lower bound on the subspace coding capacity (Theorem 3), where the lower bound is further shown to be tight for LOCs with a unique subspace degradation. Optimal uniform-given-row-space input distributions for subspace coding are characterized (Lemma 8 and Theorem 4), and the maximum achievable rate of constant-rank uniform-given-row-space input distribution is given explicitly. For a LOC with a unique subspace degradation, the subspace coding capacity can be obtained by solving a convex optimization over the input rank distribution (Theorem 5), which generalizes the similar result obtained for LOCs with uniform-given-rank transfer matrices in [17] and [18]. For row-space symmetric LOCs, an upper bound on C_{SS} is also obtained (Lemma 12).

To compare C_{SS} with C, we characterize, for both LOCs with a unique subspace degradation and row-space-symmetric LOCs, necessary conditions and sufficient conditions for $C_{\mathrm{SS}} = C$ (Theorem 6 and 7). Subspace coding is not Shannon capacity achieving for both classes of LOCs if certain Markov conditions are not satisfied. On the other hand, subspace coding is capacity achieving for *degraded* LOCs, which has $I(X;Y) = I(\langle X\rangle;\langle Y\rangle)$ for all input distributions. A degraded LOC has a unique subspace degradation and is also row-space symmetric (Theorem 8). The LOCs studied in [15]–[18] are all degraded. We further characterize a new class of degraded LOCs, called *rank-symmetric* LOCs, and show that a LOC with a uniform-given-rank transfer matrix is always rank symmetric, but not vice versa when $T < M$ (Theorem 9).

The relationship among the classes of LOCs characterized in this paper is demonstrated in Fig. 2. Note that when $T \geq M$, a row-space-symmetric LOC always has a unique subspace

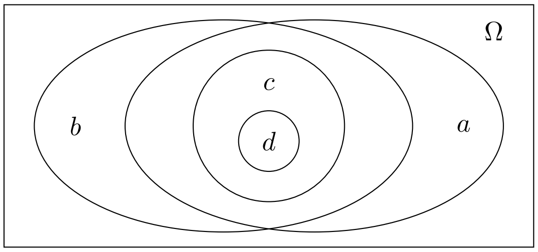

Fig. 2. The Venn diagram about LOCs. Ω is the set of all LOCs. In additional to all LOCs, we study four subsets of LOCs: a is the set of row-space-symmetric LOCs; b is the set of LOCs with a unique subspace degradation; c is the set of degraded LOCs; and d is the set of rank-symmetric LOCs. Note that $a \subset b$ when $T \geq M$, and d includes the LOCs studied in [15]–[18].

degradation, but when $T < M$, a row-space-symmetric LOC may not have a unique subspace degradation.

The rest of this paper is organized as follows. After introducing some notations and mathematical results in Section II, we discuss symmetry properties of LOCs and bounds on C in Section III. Subspace coding for LOCs is studied in Section IV. The comparison between C and C_{SS} is made in Section IV. Finally, conclusion remarks are drawn in Section VI.

II. PRELIMINARIES

Readers can skip this section and come back later when these definitions/results are referred to.

A. Counting in Projective Space

Let $\mathbb{F}$ be the finite field with q elements. Regard the vectors in $\mathbb{F}^t$ as column vectors. For a matrix $\mathbf{X}$, let $\mathrm{rk}(\mathbf{X})$ be the rank of $\mathbf{X}$, let $\mathbf{X}^\top$ be the transpose of $\mathbf{X}$, and let $\langle \mathbf{X} \rangle$ be the subspace spanned by the columns of $\mathbf{X}$. We call $\langle \mathbf{X} \rangle$ and $\langle \mathbf{X}^\top \rangle$ the column space and the row space of $\mathbf{X}$, respectively.

For a matrix $\mathbf{B}$ and a set of matrices $\mathcal{A}$, we define

$$\mathbf{B} + \mathcal{A} \triangleq \{\mathbf{B} + \mathbf{D} : \mathbf{D} \in \mathcal{A}\},$$

and

$$\mathbf{B}\mathcal{A} \triangleq \{\mathbf{B}\mathbf{D} : \mathbf{D} \in \mathcal{A}\}. \tag{2}$$

The multiplication $\mathcal{A}\mathbf{B}$ can be similarly defined.

The *projective space* $\mathrm{Pj}(\mathbb{F}^t)$ is the collection of all subspaces of $\mathbb{F}^t$. If V is a subspace of U, we write $V \leq U$. Define

$$\mathrm{Pj}(m, \mathbb{F}^t) \triangleq \{V : V \leq \mathbb{F}^t, \dim(V) \leq m\}.$$

This paper involves some counting results in projective spaces, some of which have been discussed in previous works (see [9], [12], and [20]–[23] and the reference therein). A self-contained discussion can be found in [24].

Let $\mathrm{Fr}(\mathbb{F}^{m\times r})$ be the set of full rank matrices in $\mathbb{F}^{m\times r}$. Define

$$\chi_r^m \triangleq \begin{cases} (q^m - 1)(q^m - q)\cdots(q^m - q^{r-1}) & 0 < r \leq m \\ 1 & r = 0 \end{cases} \tag{3}$$

For $r \leq m$, it is well-known that $|\mathrm{Fr}(\mathbb{F}^{m\times r})| = \chi_r^m$. Define

$$\zeta_r^m \triangleq \chi_r^m q^{-mr}. \tag{4}$$

Since the number of $m \times r$ matrices is q^{mr}, ζ_r^m is equal to the probability that a randomly chosen $m \times r$ matrix is full rank.

The *Grassmannian* $\mathrm{Gr}(r, \mathbb{F}^t)$ is the set of all r-dimensional subspaces of $\mathbb{F}^t$. Thus $\mathrm{Pj}(m, \mathbb{F}^t) = \bigcup_{r\leq m} \mathrm{Gr}(r, \mathbb{F}^t)$. The *Gaussian binomial* [20]

$$\begin{bmatrix} m \\ r \end{bmatrix} \triangleq \frac{\chi_r^m}{\chi_r^r}$$

is the number of r-dimensional subspaces of $\mathbb{F}^m$, i.e., $|\mathrm{Gr}(r, \mathbb{F}^m)| = \begin{bmatrix} m \\ r \end{bmatrix}$. Let

$$\chi_r^{m,n} \triangleq \frac{\chi_r^m \chi_r^n}{\chi_r^r},$$

which is the number of $m \times n$ matrices with rank r [21]. So we have

$$\sum_r \chi_r^{m,n} = q^{mn}. \tag{5}$$

The following counting result is a special case of [12, Lemma 2].

Lemma 1: Let V be an s-dimensional subspace of $\mathbb{F}^t$. For any integer r with $s \leq r \leq t$,

$$|\{U \in \mathrm{Gr}(r, \mathbb{F}^t) : V \leq U\}| = \begin{bmatrix} t-s \\ r-s \end{bmatrix} = \begin{bmatrix} t \\ r \end{bmatrix} \frac{\chi_s^r}{\chi_s^t}.$$

B. Probability Distribution Over Matrices and Subspaces

For a discrete random variable X, we use p_X to denote its probability mass function (PMF). For two random variables X and Y defined on discrete alphabets $\mathcal{X}$ and $\mathcal{Y}$, respectively, we write a transition probability (matrix) from $\mathcal{X}$ to $\mathcal{Y}$ as $P_{Y|X}(\mathbf{Y}|\mathbf{X})$, $\mathbf{X} \in \mathcal{X}$ and $\mathbf{Y} \in \mathcal{Y}$. We say a transition matrix is *deterministic* if all its entries are either zero or one. When it is clear from the context, we may omit the subscript of p_X and $P_{Y|X}$ to simplify the notations. Let $\mathcal{H}(X)$ be the entropy[1] of X and $I(X;Y)$ be the mutual information between X and Y. We take logarithms to the base 2.

For the sake of reference and comparison, we define three classes of conditionally uniform distributions that will be used in the paper.

Definition 1 (Uniform-Given-Row-Space Distribution (α-Type Distribution)): A PMF p over $\mathbb{F}^{m\times n}$ is *uniform-given-row-space* if $p(\mathbf{X}) = p(\mathbf{X}')$ whenever $\langle \mathbf{X}^\top \rangle = \langle \mathbf{X}'^\top \rangle$. In other words, a random matrix $X \in \mathbb{F}^{m\times n}$ is uniform given row space if

$$p_X(\mathbf{X}) = \frac{p_{\langle X^\top \rangle}(\langle \mathbf{X}^\top \rangle)}{\chi^m_{\mathrm{rk}(\mathbf{X})}}.$$

Definition 2 (Uniform-Given-Rank Distribution): A PMF p over $\mathbb{F}^{m\times n}$ is *uniform-given-rank* if $p(\mathbf{X}) = p(\mathbf{X}')$ whenever $\mathrm{rk}(\mathbf{X}) = \mathrm{rk}(\mathbf{X}')$. In other words, a random matrix $X \in \mathbb{F}^{m\times n}$ is uniform-given-rank if

$$p_X(\mathbf{X}) = \frac{p_{\mathrm{rk}(X)}(\mathrm{rk}(\mathbf{X}))}{\chi^{m,n}_{\mathrm{rk}(\mathbf{X})}}.$$

---

[1] The calligraphic $\mathcal{H}$ is used to denote entropy to make a distinction to the notion of the transfer matrix H.

$$\begin{tikzcd}
 & \langle X^\top\rangle \arrow[dr] & & & \langle Y^\top\rangle \arrow[dr] & \\
\mathrm{rk}(X) \arrow[ur] \arrow[dr] & & X \arrow[r] & Y \arrow[ur] \arrow[dr] & & \mathrm{rk}(Y) \\
 & \langle X\rangle \arrow[ur] & & & \langle Y\rangle \arrow[ur] &
\end{tikzcd}$$

Fig. 3. Random variables and Markov chains related to LOC(H, T). All the random variables in a directed path form a Markov chain. For example, $\mathrm{rk}(X) \to \langle X\rangle \to X \to Y \to \langle Y\rangle \to \mathrm{rk}(Y)$ forms a Markov chain.

Definition 3 (Uniform-Given-Dimension Distribution): A PMF p over $\mathrm{Pj}(\mathbb{F}^T)$ is *uniform-given-dimension* if $p(V) = p(V')$ whenever $\dim(V) = \dim(V')$.

A uniform-given-rank distribution is also uniform-given-row-space. If X has a uniform-given-row-space distribution, $\langle X\rangle$ has a uniform-given-dimension distribution. Further define two classes of transition matrices with certain symmetry properties as follows.

Definition 4 (Row-Space-Symmetric Transition Matrix): A transition matrix $P(\cdot|\cdot) : \mathbb{F}^{t\times m} \to \mathbb{F}^{t\times n}$ is said to be *row-space-symmetric* if

$$P(\mathbf{Y}|\mathbf{X}) = P(\mathbf{Y}'|\mathbf{X}')$$

whenever $\langle \mathbf{Y}\rangle \le \langle \mathbf{X}\rangle$, $\langle \mathbf{Y}'\rangle \le \langle \mathbf{X}'\rangle$, $\langle \mathbf{X}^\top\rangle = \langle \mathbf{X}'^\top\rangle$ and $\langle \mathbf{Y}^\top\rangle = \langle \mathbf{Y}'^\top\rangle$. In other words, the transition probability $P(\mathbf{Y}|\mathbf{X})$, $\langle \mathbf{Y}\rangle \le \langle \mathbf{X}\rangle$, is determined by the row spaces of the input and output matrices.

Definition 5 (Rank-Symmetric Transition Matrix): A transition matrix $P(\cdot|\cdot) : \mathbb{F}^{t\times m} \to \mathbb{F}^{t\times n}$ is said to be *rank-symmetric* if

$$P(\mathbf{Y}|\mathbf{X}) = P(\mathbf{Y}'|\mathbf{X}')$$

whenever $\langle \mathbf{Y}\rangle \le \langle \mathbf{X}\rangle$, $\langle \mathbf{Y}'\rangle \le \langle \mathbf{X}'\rangle$, $\mathrm{rk}(\mathbf{X}) = \mathrm{rk}(\mathbf{X}')$ and $\mathrm{rk}(\mathbf{Y}) = \mathrm{rk}(\mathbf{Y}')$. In other words, the transition probability $P(\mathbf{Y}|\mathbf{X})$, $\langle \mathbf{Y}\rangle \le \langle \mathbf{X}\rangle$, is determined by the ranks of the input and output matrices.

III. CAPACITY OF LINEAR OPERATOR CHANNELS

A LOC defined in (1), denoted by LOC(H, T), is a *discrete memoryless channel* (DMC). The dimensions of the transfer matrices discussed in this paper are $M \times N$ unless otherwise specified. Under the assumption that H and X are independent, the transition probability $P_{Y|X}(\mathbf{Y}|\mathbf{X})$ is given by

$$P_{Y|X}(\mathbf{Y}|\mathbf{X}) = \Pr\{\mathbf{X}H = \mathbf{Y}\}.$$

The *(Shannon) capacity* of LOC(H, T) is

$$C = C(H, T) = \max_{p_X} I(X; Y).$$

The input X, the output Y, their row/column spaces and their ranks form Markov chains shown in Fig. 3.

In this section, we first introduce the essential technique of this paper—some symmetry properties of LOCs. We then investigate the input distributions that achieve the Shannon capacity, and give upper and lower bounds on the Shannon capacity.

A. Symmetry Properties

The following lemma demonstrates an intrinsic symmetry property of LOCs. A matrix is said to have full column (row) rank if its rank is equal to its number of columns (rows).

Lemma 2: For LOC(H, T), if $\mathbf{X} = \mathbf{BD}$ and $\mathbf{Y} = \mathbf{BE}$ where $\mathbf{B}$ has full column rank, then

$$P_{Y|X}(\mathbf{Y}|\mathbf{X}) = \Pr\{\mathbf{X}H = \mathbf{Y}\} = \Pr\{\mathbf{D}H = \mathbf{E}\}.$$

Proof: The lemma follows from $P_{Y|X}(\mathbf{Y}|\mathbf{X}) = \Pr\{\mathbf{BD}H = \mathbf{BE}\} = \Pr\{\mathbf{D}H = \mathbf{E}\}$, where the last equality follows because $\mathbf{B}$ has full column rank. ■

Recall that a DMC is defined to be *symmetric* [25] if the set of outputs can be partitioned into subsets in such a way that for each subset the matrix of transition probabilities (using inputs as rows and outputs of the subset as columns) has the property that each row is a permutation of each other row and each column (if more than one) is a permutation of each other column. The transition matrix of a LOC satisfies properties similar to these of a symmetric channel, but in general, a LOC is not a symmetric channel.

Lemma 3: The transition matrix of LOC(H, T) satisfies the following properties:

1) *For $\mathbf{X}_1, \mathbf{X}_2 \in \mathbb{F}^{T\times M}$ with $\langle \mathbf{X}_1^\top\rangle = \langle \mathbf{X}_2^\top\rangle$ and $V \le \mathbb{F}^N$, the vector $(P_{Y|X}(\mathbf{Y}|\mathbf{X}_1) : \mathbf{Y} \in \mathbb{F}^{T\times N}, \langle \mathbf{Y}^\top\rangle = V)$ is a permutation of the vector $(P_{Y|X}(\mathbf{Y}|\mathbf{X}_2) : \mathbf{Y} \in \mathbb{F}^{T\times N}, \langle \mathbf{Y}^\top\rangle = V)$;*
2) *For $\mathbf{Y}_1, \mathbf{Y}_2 \in \mathbb{F}^{T\times N}$ with $\langle \mathbf{Y}_1^\top\rangle = \langle \mathbf{Y}_2^\top\rangle$ and $U \le \mathbb{F}^M$, the vector $(P_{Y|X}(\mathbf{Y}_1|\mathbf{X}) : \mathbf{X} \in \mathbb{F}^{T\times M}, \langle \mathbf{X}^\top\rangle = U)$ is a permutation of the vector $(P_{Y|X}(\mathbf{Y}_2|\mathbf{X}) : \mathbf{X} \in \mathbb{F}^{T\times M}, \langle \mathbf{X}^\top\rangle = U)$.*

Proof: Let $\phi(V) = \{\mathbf{Y} \in \mathbb{F}^{T\times N}, \langle \mathbf{Y}^\top\rangle = V\}$. To prove 1), we show that there exists a bijection $f : \phi(V) \to \phi(V)$ such that $\Pr\{\mathbf{X}_1 H = \mathbf{Y}\} = \Pr\{\mathbf{X}_2 H = f(\mathbf{Y})\}$. Since $\langle \mathbf{X}_1^\top\rangle = \langle \mathbf{X}_2^\top\rangle$, there exists a full rank matrix $\mathbf{T}$ such that $\mathbf{X}_2 = \mathbf{T}\mathbf{X}_1$. Define $f : \phi(V) \to \phi(V)$ as $f(\mathbf{Y}) = \mathbf{TY}$. Since $\mathbf{T}$ is a full rank square matrix, f is a bijection. The claim in 1) is verified by $\Pr\{\mathbf{X}_2 H = f(\mathbf{Y})\} = \Pr\{\mathbf{X}_2 H = \mathbf{TY}\} = \Pr\{\mathbf{T}^{-1}\mathbf{X}_2 H = \mathbf{Y}\} = \Pr\{\mathbf{X}_1 H = \mathbf{Y}\}$, where the second equality follows from Lemma 2.

The proof of 2) is similar and hence omitted. ■

For the input matrices with different row spaces, the rows of the transition matrix are usually not a permutation of each other. The following result implied by Lemma 3 will be used in this paper.

Lemma 4: For a LOC, if $\langle \mathbf{X}^\top\rangle = \langle \mathbf{X}'^\top\rangle$, then

$$P_{\langle Y^\top\rangle|X}(V|\mathbf{X}) = P_{\langle Y^\top\rangle|X}(V|\mathbf{X}')$$

and

$$P_{\mathrm{rk}(Y)|X}(s|\mathbf{X}) = P_{\mathrm{rk}(Y)|X}(s|\mathbf{X}').$$

Proof: Since

$$P_{\langle Y^\top\rangle|X}(V|\mathbf{X}) = \sum_{\mathbf{Y}:\langle \mathbf{Y}^\top\rangle = V} P_{Y|X}(\mathbf{Y}|\mathbf{X}),$$

the first equality follows from 1) in Lemma 3. The second equality follows from the first one. ■

B. Uniform-Given-Row-Space Input Distributions

The intrinsic symmetry property of LOCs implies that the capacity-achieving input distributions should have certain symmetry property, which is characterized in the following theorem. Recall the definition of uniform-given-row-space distribution in Definition 1.

Theorem 1: There exists a uniform-given-row-space input distribution that maximizes $I(X; Y)$ for any LOC.

Proof: Let p be an optimal input distribution for LOC(H, T). For $\Phi \in \mathrm{Fr}(\mathbb{F}^{T\times T})$, define p^Φ as $p^\Phi(\mathbf{X}) = p(\Phi\mathbf{X})$. First p^Φ is a PMF because $0 \le p^\Phi(\mathbf{X}) = p(\Phi\mathbf{X}) \le 1$ and $\sum_{\mathbf{X}\in\mathbb{F}^{T\times M}} p^\Phi(\mathbf{X}) = 1$.

We show that p^Φ also achieves the capacity of the LOC. For the simplicity of the notations, we write $p' = p^\Phi$. Let p_Y and p'_Y be the PMF of Y when the input distributions are p and p', respectively. We have

$$\begin{aligned} p'_Y(\mathbf{Y}) &= \sum_{\mathbf{X}\in\mathbb{F}^{T\times M}} p'(\mathbf{X}) P_{Y|X}(\mathbf{Y}|\mathbf{X}) \\ &= \sum_{\mathbf{X}\in\mathbb{F}^{T\times M}} p(\Phi\mathbf{X}) P_{Y|X}(\Phi\mathbf{Y}|\Phi\mathbf{X}) \qquad (6) \\ &= \sum_{\mathbf{X}'\in\mathbb{F}^{T\times M}} p(\mathbf{X}') P_{Y|X}(\Phi\mathbf{Y}|\mathbf{X}') \qquad (7) \\ &= p_Y(\Phi\mathbf{Y}), \end{aligned}$$

where (6) follows from Lemma 2 and $p'(\mathbf{X}) = p(\Phi\mathbf{X})$, and (7) follows by letting $\mathbf{X}' = \Phi\mathbf{X}$ and noting $\Phi\mathbb{F}^{T\times M} = \mathbb{F}^{T\times M}$. Therefore,

$$\begin{aligned} &I(X; Y)|_{p'} \\ &= \sum_{\mathbf{X}\in\mathbb{F}^{T\times M}} p'(\mathbf{X}) \sum_{\mathbf{Y}\in\mathbb{F}^{T\times N}} P(\mathbf{Y}|\mathbf{X}) \log \frac{P(\mathbf{Y}|\mathbf{X})}{p'_Y(\mathbf{Y})} \\ &= \sum_{\mathbf{X}\in\mathbb{F}^{T\times M}} p(\Phi\mathbf{X}) \sum_{\mathbf{Y}\in\mathbb{F}^{T\times N}} P(\Phi\mathbf{Y}|\Phi\mathbf{X}) \log \frac{P(\Phi\mathbf{Y}|\Phi\mathbf{X})}{p(\Phi\mathbf{Y})} \\ &= \sum_{\mathbf{X}'\in\mathbb{F}^{T\times M}} p(\mathbf{X}') \sum_{\mathbf{Y}'\in\mathbb{F}^{T\times N}} P(\mathbf{Y}'|\mathbf{X}') \log \frac{P(\mathbf{Y}'|\mathbf{X}')}{p(\mathbf{Y}')} \\ &= I(X; Y)|_p, \end{aligned}$$

where the second equality follows from $p'_Y(\mathbf{Y}) = p_Y(\Phi\mathbf{Y})$ and $P(\mathbf{Y}|\mathbf{X}) = P(\Phi\mathbf{Y}|\Phi\mathbf{X})$ (see Lemma 2).

Define p^* as

$$p^*(\mathbf{X}) = \frac{1}{|\mathrm{Fr}(\mathbb{F}^{T\times T})|} \sum_{\Phi\in\mathrm{Fr}(\mathbb{F}^{T\times T})} p^\Phi(\mathbf{X}).$$

Since mutual information is a concave function of the input distribution [25],

$$\begin{aligned} I(X; Y)|_{p^*} &\ge \frac{1}{|\mathrm{Fr}(\mathbb{F}^{T\times T})|} \sum_{\Phi\in\mathrm{Fr}(\mathbb{F}^{T\times T})} I(X; Y)|_{p^\Phi} \\ &= C(H, T). \end{aligned}$$

Thus, p^* is also an optimal input distribution for the channel. The proof is completed by noting that p^* is uniform-given-row-space. ■

Theorem 1 reveals that uniform-given-row-space input distributions can match the intrinsic symmetry of LOCs. We will show more applications of uniform-given-row-space input distributions in this paper.

In the remaining part of this subsection, we discuss how the calculation of the transition matrix and the channel capacity can be simplified by the symmetry properties and uniform-given-row-space input distributions. To compute the channel capacity of LOC(H, T), the first step is to compute the matrix of transition probabilities using the distribution of H. A straightforward computation of the transition matrix from the distribution of H requires the calculation of $q^{T(M+N)}$ components of the transition matrix. But using the symmetry properties, this number can be reduced to

$$\sum_{k=0}^{\min\{T,M\}} \begin{bmatrix} M \\ k \end{bmatrix} q^{kN} < \begin{cases} cq^{MN} & M \le \min\{T, N\} \\ c'q^{L(M+N-L)} & \text{otherwise}, \end{cases}$$

where $L = \min\{T, (M + N)/2\}$, c and c' are constants (ref. Appendix A).

The input distribution of a LOC has q^{TM} probability masses. To find an optimal input distribution, a straightforward approach needs to determine $q^{TM} - 1$ out of them. Theorem 1 enable us to focus on uniform-given-row-space input distributions, which is determined by a PMF over $\mathrm{Pj}(\min\{M, T\}, \mathbb{F}^M)$. Thus the number of probability masses to determine can be reduced to

$$\sum_{k=0}^{\min\{M,T\}} \begin{bmatrix} M \\ k \end{bmatrix} < \begin{cases} \Theta_1 q^{M^2/4} & \text{for } T \ge M/2 \\ \Theta_2 q^{T(M-T)} & \text{otherwise}, \end{cases}$$

where Θ_1 and Θ_2 are constants (ref. Appendix A). Note that those computations are still complicated for relatively large M and T.

C. Upper and Lower Bounds on C

We derive bounds on $I(X; Y)$ with the addition of two terms: one corresponds to the intrinsic symmetry and another one is the achievable rate of the channel given by $P_{\langle Y^\top\rangle|\langle X^\top\rangle}$. Note that in Lemma 3, the symmetry property only holds for input (output) matrices sharing the same row space. Roughly, the transition matrix $P_{\langle Y^\top\rangle|\langle X^\top\rangle}$ captures some property of a LOC that is opposite to the intrinsic symmetry.

The transition matrix $P_{\langle Y^\top\rangle|\langle X^\top\rangle}$ is solely determined by p_H. By Lemma 4, $P_{\langle Y^\top\rangle|\langle X^\top\rangle}(V|U) = P_{\langle Y^\top\rangle|X}(V|\mathbf{X})$ for any $\mathbf{X}$ with $\langle \mathbf{X}^\top\rangle = U$. Let

$$J(\mathrm{rk}(X); \mathrm{rk}(Y)) \triangleq \sum_{s,r} p_{\mathrm{rk}(X)\,\mathrm{rk}(Y)}(r, s) \log \frac{\chi_s^T}{\chi_s^r}. \qquad (8)$$

Note that $J(\mathrm{rk}(X); \mathrm{rk}(Y))$ is always nonnegative (see the definition of χ_s^r in (3)), and $p_{\mathrm{rk}(X)\,\mathrm{rk}(Y)}(r, s)$ can be solely derived from $p_{\langle X^\top\rangle}$ and $P_{\langle Y^\top\rangle|\langle X^\top\rangle}$ as

$$p_{\mathrm{rk}(X)\,\mathrm{rk}(Y)}(r, s) = \sum_{U\in\mathrm{Gr}(r,\mathbb{F}^M)} P_{\mathrm{rk}(Y)|\langle X^\top\rangle}(s|U) p_{\langle X^\top\rangle}(U),$$

where $P_{\mathrm{rk}(Y)|\langle X^\top\rangle}(s|U) = \sum_{V:\dim(V)=s} P_{\langle Y^\top\rangle|\langle X^\top\rangle}(V|U)$.

Theorem 2: Consider LOC(H, T) with input X and output Y. For a uniform-given-row-space input distribution,

$$I(X; Y) \ge J(\mathrm{rk}(X); \mathrm{rk}(Y)) + I(\langle X^\top\rangle; \langle Y^\top\rangle) \qquad (9)$$

and

$$I(X;Y) \le J(\mathrm{rk}(X);\mathrm{rk}(Y)) + I(\langle X^\top\rangle;\langle Y^\top\rangle) + \sum_{s,r} p_{\mathrm{rk}(X)\,\mathrm{rk}(Y)}(r,s)\log \chi_s^r, \tag{10}$$

where the equality in (9) *holds when the LOC has a row-space-symmetric transition matrix.*

Proof: Fix a uniform-given-row-space input distribution p_X. Let Y^* be a random matrix over $\mathbb{F}^{T\times N}$ with transition probability

$$P_{Y^*|X}(\mathbf{Y}|\mathbf{X}) = \begin{cases} \frac{P_{\langle Y^\top\rangle|\langle X^\top\rangle}(\langle \mathbf{Y}^\top\rangle|\langle \mathbf{X}^\top\rangle)}{\chi_{\mathrm{rk}(\mathbf{Y})}^{\mathrm{rk}(\mathbf{X})}} & \langle \mathbf{Y}\rangle \le \langle \mathbf{X}\rangle, \\ 0 & \text{otherwise.} \end{cases}$$

Note that $P_{Y^*|X}$ is row-space symmetric.

The proofs of the following claims are given in Appendix B.

Claim 1: For a uniform-given-row-space input distribution p_X, $I(X;Y) \ge I(X;Y^*)$, with equality when $P_{Y^*|X} = P_{Y|X}$.

We can show Claim 1 using the property that for fixed p_X, mutual information $I(X;Y)$ is a convex function of the transition probabilities. We can further show the following claim by directly applying the definition of $P_{Y^*|X}$.

Claim 2: For a uniform-given-row-space input distribution p_X,

$$\mathcal{H}(Y^*|X) = \sum_{s\le r} p_{\mathrm{rk}(X)\,\mathrm{rk}(Y)}(r,s)\log \chi_s^r + \mathcal{H}(\langle Y^\top\rangle|\langle X^\top\rangle) \tag{11}$$

and

$$\mathcal{H}(Y^*) = \sum_{s} p_{\mathrm{rk}(Y)}(s)\log \chi_s^T + \mathcal{H}(\langle Y^\top\rangle). \tag{12}$$

By Claim 2,

$$\begin{aligned} I(X;Y^*) &= \mathcal{H}(Y^*) - \mathcal{H}(Y^*|X) \\ &= \sum_{s\le r} p_{\mathrm{rk}(X)\,\mathrm{rk}(Y)}(r,s)\log \frac{\chi_s^T}{\chi_s^r} + I(\langle X^\top\rangle;\langle Y^\top\rangle), \end{aligned}$$

which, together with Claim 1, proves (9).

To prove (10), we have

$$\begin{aligned} \mathcal{H}(Y) &= \sum_s \sum_{V\in \mathrm{Gr}(s,\mathbb{F}^N)} \sum_{\mathbf{Y}:\langle \mathbf{Y}^\top\rangle = V} p_Y(\mathbf{Y})\log\frac{1}{p_Y(\mathbf{Y})} \\ &\le \sum_s \sum_{V\in \mathrm{Gr}(s,\mathbb{F}^N)} p_{\langle Y^\top\rangle}(V)\log\frac{\chi_s^T}{p_{\langle Y^\top\rangle}(V)} \qquad (13) \\ &= \mathcal{H}(Y^*), \qquad (14) \end{aligned}$$

where (13) is derived by the log-sum inequality (see [26]) and (14) is obtained by (12). Then,

$$\begin{aligned} &I(X;Y) - I(X;Y^*) \\ &\quad = \mathcal{H}(Y) - \mathcal{H}(Y|X) - \mathcal{H}(Y^*) + \mathcal{H}(Y^*|X) \\ &\quad \le \mathcal{H}(Y^*|X) - \mathcal{H}(Y|X) \\ &\quad \le \sum_{s\le r} p_{\mathrm{rk}(X)\,\mathrm{rk}(Y)}(r,s)\log \chi_s^r, \end{aligned}$$

where the last inequality follows from (11) and $\mathcal{H}(Y|X) \ge \mathcal{H}(\langle Y^\top\rangle|X) = \mathcal{H}(\langle Y^\top\rangle|\langle X^\top\rangle)$ since $p_{\langle Y^\top\rangle|X}(V|\mathbf{X})$ depends on $\mathbf{X}$ only through $\langle \mathbf{X}^\top\rangle$ (see Lemma 4). ■

The lower bound in the above theorem can be achieved by a coding scheme employing a superposition structure, which includes a cloud code and a set of satellite codes, each of which corresponds to a cloud center. The rate $I(\langle X^\top\rangle;\langle Y^\top\rangle)$ can be achieved by the cloud code, while the rate $J(\mathrm{rk}(X);\mathrm{rk}(Y))$ can be achieved by the satellite codes. Readers are referred to [27] for detailed discussion of this coding scheme.

Our lower bound is at least as good as the lower bound obtained in [17] and [18], where a transfer matrix is converted to a uniform-given-rank transfer matrix. We will compare these two bounds at the end of Section V-C.

In the following sections, we will see that the quantity $J(\mathrm{rk}(X);\mathrm{rk}(Y))$ is also related to the coding rate of subspace coding. In the definition of $J(\mathrm{rk}(X);\mathrm{rk}(Y))$, the inverse of the term $\frac{\chi_s^T}{\chi_s^r}$ has the following meaning. Let V be an s-dimensional subspaces of $\mathbb{F}^T$, which can be regarded as the column space of the output matrix. The number of r-dimensional subspaces of $\mathbb{F}^T$ is $\begin{bmatrix} T \\ r \end{bmatrix}$; and by Lemma 1, the number of r-dimensional subspaces of $\mathbb{F}^T$ that include V, which are the possible column spaces of the input matrix, is $\begin{bmatrix} T \\ r \end{bmatrix}\frac{\chi_s^r}{\chi_s^T}$. Thus, the fraction of the number of r-dimensional subspaces of $\mathbb{F}^T$ that include V as a subspace is exactly $\frac{\chi_s^r}{\chi_s^T}$.

Let us look at another property of the quantity $J(\mathrm{rk}(X);\mathrm{rk}(Y))$. Let

$$\epsilon(T,q) \triangleq \sum_s p_{\mathrm{rk}(H)}(s)\log\frac{\zeta_s^T}{\zeta_s^M},$$

where ζ_r^m is defined in (4).

Lemma 5: If $T \ge M$ and $p_{\mathrm{rk}(X)}(M) = 1$,

$$J(\mathrm{rk}(X);\mathrm{rk}(Y)) = (T-M)\,\mathrm{E}[\mathrm{rk}(H)]\log q + \epsilon(T,q),$$

where $0 \le \epsilon(T,q) < 1.8$ for all T and q.

Proof: When $T \ge M$ and $p_{\mathrm{rk}(X)}(M) = 1$,

$$\begin{aligned} J(\mathrm{rk}(X);\mathrm{rk}(Y)) &= \sum_s p_{\mathrm{rk}(H)}(s)\log q^{(T-M)s}\frac{\zeta_s^T}{\zeta_s^M} \qquad (15) \\ &= (T-M)\,\mathrm{E}[\mathrm{rk}(H)]\log q + \epsilon(T,q), \end{aligned}$$

where (15) follows that $\mathrm{rk}(H) = \mathrm{rk}(Y)$ since X has full column rank.

The lower bound on $\epsilon(T,q)$ holds due to $T \ge M$, and the upper bound on $\epsilon(T,q)$ is obtained by bounding ζ_r^m using a constant given in [22]. ■

The above lemma tells us that when $T > M$, $J(\mathrm{rk}(X);\mathrm{rk}(Y))$ is larger than $(T-M)\,\mathrm{E}[\mathrm{rk}(H)]\log q$, which is the maximum achievable rate of channel training [6]. (Recall that in channel training, M rows of X are used to recover the transfer matrix in the receiver.) We know that subspace coding can in general do better than channel training [9]. The lower bound in Theorem 2 implies that the rate gain is at least

$$\max_{p_X:\text{uniform-given-row-space},\, p_{\mathrm{rk}(X)}(M)=1} \epsilon(T,q).$$

(Note that $I(\langle X^\top\rangle;\langle Y^\top\rangle) = 0$ when $p_{\mathrm{rk}(X)}(M) = 1$ since $\langle X^\top\rangle = \mathbb{F}^M$ is deterministic when $\mathrm{rk}(X) = M$.)

Example 1: Consider LOC$(H, 1)$, for which channel training is not useful. We have

$$\begin{aligned}
&J(\mathrm{rk}(X); \mathrm{rk}(Y)) \\
&= \sum_{r\in\{0,1\}} \sum_{s\in\{0,1\}:s\leq r} p_{\mathrm{rk}(X)\,\mathrm{rk}(Y)}(r, s) \log \frac{\chi_s^1}{\chi_s^r} \\
&= p_{\mathrm{rk}(X)\,\mathrm{rk}(Y)}(0, 0) \log \frac{\chi_0^1}{\chi_0^0} + p_{\mathrm{rk}(X)\,\mathrm{rk}(Y)}(1, 0) \log \frac{\chi_0^1}{\chi_0^1} \\
&\quad + p_{\mathrm{rk}(X)\,\mathrm{rk}(Y)}(1, 1) \log \frac{\chi_1^1}{\chi_1^1} = 0.
\end{aligned}$$

The lower bound in Theorem 2 for LOC$(H, 1)$ becomes $I(\langle X^\top\rangle; \langle Y^\top\rangle)$, and the gap between the upper bound and the lower bound is $p_{\mathrm{rk}(X)\,\mathrm{rk}(Y)}(1, 1) \log(q-1)$. Note that $I(\langle X^\top\rangle; \langle Y^\top\rangle)$ can be large. For example, when H is the $M \times M$ identity matrix, $I(\langle X^\top\rangle; \langle Y^\top\rangle) = \log|\mathrm{Gr}(1, \mathbb{F}^M)| = \log \frac{q^M-1}{q-1} \geq (M-1)\log q$. ♢

D. Properties of Row-Space-Symmetric LOCs

We call a LOC *row-space-symmetric* if its transition matrix is row-space-symmetric. The lower bound in Theorem 2 is tight for row-space-symmetric LOCs. We introduce some properties of such LOCs to be used in other sections.

By definition, a LOC is row-space symmetric if and only if for any $\mathbf{X}$ and $\mathbf{Y}$ with $\langle \mathbf{Y}\rangle \leq \langle \mathbf{X}\rangle$,

$$P_{Y|X}(\mathbf{Y}|\mathbf{X}) = \frac{1}{\chi_{\mathrm{rk}(\mathbf{Y})}^{\mathrm{rk}(\mathbf{X})}} P_{\langle Y^\top\rangle|\langle X^\top\rangle}(\langle \mathbf{Y}^\top\rangle|\langle \mathbf{X}^\top\rangle),$$

where $\chi_{\mathrm{rk}(\mathbf{Y})}^{\mathrm{rk}(\mathbf{X})}$ is the number of $\mathbf{Y}_1$ such that $\langle \mathbf{Y}_1\rangle \leq \langle \mathbf{X}\rangle$ and $\langle \mathbf{Y}_1^\top\rangle = \langle \mathbf{Y}^\top\rangle$, and $P_{\langle Y^\top\rangle|\langle X^\top\rangle}(\langle \mathbf{Y}^\top\rangle|\langle \mathbf{X}^\top\rangle)$ is only determined by p_H.

Lemma 6: When $T \geq M$, LOC(H, T) being row-space-symmetric implies that H is uniform-given-row-space.

Proof: Let $\mathbf{X}$ and $\mathbf{X}'$ be two full-rank input matrices with $\langle \mathbf{X}^\top\rangle = \langle \mathbf{X}'^\top\rangle$. Let $\mathbf{H}$ and $\mathbf{H}'$ be two transfer matrices with $\langle \mathbf{H}^\top\rangle = \langle \mathbf{H}'^\top\rangle$. Since $T \geq M$, we have $\langle (\mathbf{XH})^\top\rangle = \langle \mathbf{H}^\top\rangle$ and $\langle (\mathbf{XH}')^\top\rangle = \langle \mathbf{H}'^\top\rangle$. Hence $\langle (\mathbf{XH})^\top\rangle = \langle (\mathbf{XH}')^\top\rangle$. By the definition of row-space-symmetric LOCs, we have $p_{Y|X}(\mathbf{XH}|\mathbf{X}) = p_{Y|X}(\mathbf{X}'\mathbf{H}'|\mathbf{X}')$, which implies $p_H(\mathbf{H}) = p_H(\mathbf{H}')$. ■

When $T < M$, it is not necessary that the transfer matrix of a row-space-symmetric LOC satisfies the above constraint.

Example 2: We denote a LOC with $T = 1$ over the binary field $\mathbb{F}_2$ as LOC$_2(H, 1)$, where $H \in \mathbb{F}_2^{M\times N}$ is the transfer matrix. The input and the output of LOC$_2(H, 1)$ are in the same set $\mathbb{F}_2^{1\times M}$. Since the mapping from X to $\langle X^T\rangle$ in this special case is a bijection, we have $P_{Y|X}(\mathbf{Y}|\mathbf{X}) = P_{\langle Y^\top\rangle|\langle X^\top\rangle}(\langle \mathbf{Y}^\top\rangle|\langle \mathbf{X}^\top\rangle)$ for any $\mathbf{X}, \mathbf{Y} \in \mathbb{F}_2^{1\times M}$ with $\langle \mathbf{Y}\rangle \leq \langle \mathbf{X}\rangle$. Hence, LOC$_2(H, 1)$ is row-space-symmetric for any distribution of H.

IV. SUBSPACE CODING CAPACITY OF LOCS

One of the intrinsic properties of LOCs is that $\langle Y\rangle \leq \langle X\rangle$. If we restrict to the column spaces of the input and output of a LOC, it is possible that simpler encoding/decoding schemes can be developed. This approach, called subspace coding, was first adopted by Koetter and Kschischang [9] for random linear network coding. In this section, we characterize the asymptotic performance of subspace coding with multiple uses of the channel when the error probability goes to zero.

We discuss how to characterize the maximum achievable rate of subspace coding (also known as the *subspace coding capacity*) and provide lower bounds on the subspace coding capacity. We also introduce an important class of LOCs, for which the optimal subspace coding scheme is relatively easier to find.

A. Optimal Subspace Degradations

A LOC is a matrix channel, so it must be converted to a subspace channel to use subspace coding. An n-block subspace code is a subset of $(\mathrm{Pj}(\min\{T, M\}, \mathbb{F}^T))^n$. To apply a subspace code to a LOC, the subspaces in a codeword need to be converted to matrices. For $U \in \mathrm{Pj}(\min\{T, M\}, \mathbb{F}^T)$, this conversion can be done by a transition probability $P_{X|\langle X\rangle}(\cdot|U)$. The decoding of a subspace code also uses only the column spaces spanned by the received matrices. Given a transition matrix $P_{X|\langle X\rangle}$, we have a new channel with input $\langle X\rangle$ and output $\langle Y\rangle$.

Definition 6: For LOC(H, T) with a given a transition probability $P_{X|\langle X\rangle}$, we have a new channel law given by

$$P_{\langle Y\rangle|\langle X\rangle}(V|U) = \sum_{\mathbf{X}} P_{\langle Y\rangle|X}(V|\mathbf{X}) P_{X|\langle X\rangle}(\mathbf{X}|U). \tag{16}$$

This channel takes subspaces as input and output and is called a *subspace degradation* of LOC(H, T) with respect to $P_{X|\langle X\rangle}$.

The capacity of the subspace degradation of LOC(H, t) w.r.t. $P_{X|\langle X\rangle}$ is $\max_{p_{\langle X\rangle}} I(\langle Y\rangle; \langle X\rangle)$. Therefore, the subspace coding capacity of LOC(H, T) is

$$\begin{aligned}
C_{\mathrm{SS}} = C_{\mathrm{SS}}(H, T) &\triangleq \max_{P_{X|\langle X\rangle}} \max_{p_{\langle X\rangle}} I(\langle X\rangle; \langle Y\rangle) \\
&= \max_{p_X} I(\langle X\rangle; \langle Y\rangle).
\end{aligned} \tag{17}$$

To verify (17), we see that for given $P_{X|\langle X\rangle}$ and $p_{\langle X\rangle}$, the PMF of X is given by $p_X(\mathbf{X}) = p_{\langle X\rangle}(\langle \mathbf{X}\rangle) P_{X|\langle X\rangle}(\mathbf{X}|\langle \mathbf{X}\rangle)$. On the other hand, fix a distribution p_X. The distribution $p_{\langle X\rangle}$ can be derived, and the distribution $P_{X|\langle X\rangle}(\cdot|U)$ can be derived for any U with $p_{\langle X\rangle}(U) \neq 0$. If $p_{\langle X\rangle}(U) = 0$, the distribution $P_{X|\langle X\rangle}(\cdot|U)$ does not appear in the maximization of $I(\langle X\rangle; \langle Y\rangle)$.

When $P_{X|\langle X\rangle}$ is fixed, $P_{\langle Y\rangle|\langle X\rangle}(V|U)$ is also fixed (see (16)), and hence $I(\langle X\rangle; \langle Y\rangle)$ is a concave function of $p_{\langle X\rangle}$. On the other hand, when $p_{\langle X\rangle}$ is fixed, $P_{\langle Y\rangle|\langle X\rangle}(V|U)$ is a linear function of $P_{X|\langle X\rangle}$ (see (16)) and $I(\langle X\rangle; \langle Y\rangle)$ is a convex function of $P_{\langle Y\rangle|\langle X\rangle}(V|U)$, and hence $I(\langle X\rangle; \langle Y\rangle)$ is a convex function of $P_{X|\langle X\rangle}$. (We can similarly argue that $I(\langle X\rangle; \langle Y\rangle)$ is not concave in p_X in general.) Hence, finding an optimal subspace coding scheme involves maximizing a non-concave function, which is in general a difficult problem due to computational complexity.

Recall that a transition matrix is *deterministic* if all its entries are either zero or one. We can simplify the problem of finding an optimal subspace degradation by considering only deterministic transition matrices.

Lemma 7: There exists an optimal subspace degradation w.r.t. a deterministic transition matrix $P_{X|\langle X\rangle}$.

Proof: Consider a procedure as follows. Fix $p_{\langle X\rangle}$ and $P^0_{X|\langle X\rangle}$ that achieve $C_{\mathrm{SS}}(H,T)$. If $P^0_{X|\langle X\rangle}$ is deterministic, the procedure stops. Otherwise, there must exist $U\in\mathrm{Gr}(\min\{T,M\},\mathbb{F}^T)$ such that $P^0_{X|\langle X\rangle}(\mathbf{X}|U)<1$ for all input $\mathbf{X}$ with $\langle\mathbf{X}\rangle=U$.

For each $\mathbf{X}$ with $\langle\mathbf{X}\rangle=U$, define $P^{\mathbf{X}}_{X|\langle X\rangle}$ as $P^{\mathbf{X}}_{X|\langle X\rangle}(\cdot|U')=P^0_{X|\langle X\rangle}(\cdot|U')$ for $U'\neq U$ and $P^{\mathbf{X}}_{X|\langle X\rangle}(\mathbf{X}|U)=1$. We can write

$$P^0_{X|\langle X\rangle}(\cdot|\cdot)=\frac{1}{\chi^M_{\dim(U)}}\sum_{\mathbf{X}:\langle\mathbf{X}\rangle=U}P^0_{X|\langle X\rangle}(\mathbf{X}|U)P^{\mathbf{X}}_{X|\langle X\rangle}(\cdot|\cdot).$$

Since $I(\langle X\rangle;\langle Y\rangle)$ is a convex function of $P_{X|\langle X\rangle}$, there exists $\mathbf{X}_0$ with $\langle\mathbf{X}_0\rangle=U$ such that

$$I(\langle X\rangle;\langle Y\rangle)\big|_{P^{\mathbf{X}_0}_{X|\langle X\rangle}}\geq I(\langle X\rangle;\langle Y\rangle)\big|_{P^0_{X|\langle X\rangle}}.$$

Hence the subspace degradation associated with $P^{\mathbf{X}_0}_{X|\langle X\rangle}$ is also optimal. We then repeat the above procedure with $P^{\mathbf{X}_0}_{X|\langle X\rangle}$ in place of $P^0_{X|\langle X\rangle}$.

The above procedure must stop in finite steps since $\mathrm{Gr}(\min\{T,M\},\mathbb{F}^T)$ has finite elements. $P^0_{X|\langle X\rangle}$ in the final step is deterministic. ■

Lemma 7 enables us to focus on a finite set of deterministic transition matrices $P_{X|\langle X\rangle}$ to find the optimal subspace degradation. For small T, it is possible to numerically evaluate all the deterministic transition matrices $P_{X|\langle X\rangle}$.

Example 3: We use $\mathrm{LOC}(H,1)$ as an example to show how to evaluate the subspace coding capacity. The input and output of a subspace degradation can be two subspaces $\langle 0\rangle\triangleq\{0\}$ and $\langle 1\rangle\triangleq\{0,1\}$. By Lemma 7, we only need to consider subspace degradations with $P_{X|\langle X\rangle}(\mathbf{X}|\langle 1\rangle)=1$ for certain $\mathbf{X}\in\mathbb{F}^{1\times M}\setminus\{\mathbf{0}\}$, where

$$P_{\langle Y\rangle|\langle X\rangle}(\langle 0\rangle|\langle 1\rangle)=P_{Y|X}(\mathbf{0}|\mathbf{X}).$$

Since $P_{\langle Y\rangle|\langle X\rangle}(\langle 1\rangle|\langle 0\rangle)=0$, the subspace degradations of $\mathrm{LOC}(H,1)$ are Z-channels with the crossover probability given by $P_{\langle Y\rangle|\langle X\rangle}(\langle 0\rangle|\langle 1\rangle)$. We know that the capacity of Z-channel is a decreasing function of the crossover probability. So the best subspace degradation is the one with the smallest $P_{\langle Y\rangle|\langle X\rangle}(\langle 0\rangle|\langle 1\rangle)$. Therefore, the best subspace degradation can be found by evaluating $P_{Y|X}(\mathbf{0}|\mathbf{X})$ for $\mathbf{X}\in\mathbb{F}^{1\times M}\setminus\{\mathbf{0}\}$. Since different $\mathbf{X}$ spanning the same row space only need to be calculated once, we need to consider $|\mathrm{Gr}(1,\mathbb{F}^M)|=\frac{q^M-1}{q-1}$ inputs.

Since the input/output of a subspace degradation is binary, the maximum achievable rate of subspace coding for $\mathrm{LOC}(H,1)$ is at most 1 bit per use, which is much smaller than the lower bound characterized in Example 1.

B. Lower Bound on Subspace Coding Capacity

Since it is difficult to find an optimal subspace degradation in general, we consider in this section the achievable rate of subspace coding for uniform-given-row-space input distributions to get a lower bound on the subspace coding capacity. We will show (in the next subsection) that the lower bound to be obtained is exactly the subspace coding capacity for certain important special cases.

Theorem 3: For a LOC with uniform-given-row-space input distributions,

$$I(\langle X\rangle;\langle Y\rangle)=J(\mathrm{rk}(X);\mathrm{rk}(Y))+I(\mathrm{rk}(X);\mathrm{rk}(Y)),\tag{18}$$

where $J(\mathrm{rk}(X);\mathrm{rk}(Y))$ *is defined in* (8)*; and hence*

$$C_{\mathrm{SS}}\geq\max_{p_{\langle X^\top\rangle}}[J(\mathrm{rk}(X);\mathrm{rk}(Y))+I(\mathrm{rk}(X);\mathrm{rk}(Y))].$$

Proof: Fix a uniform-given-row-space input p_X. For $U\in\mathrm{Gr}(m,\mathbb{F}^t)$, let

$$A(m,U)=\{\mathbf{X}\in\mathbb{F}^{t\times m}:\langle\mathbf{X}\rangle=U\}.$$

The set $A(m,U)$ has several properties that will be used in the proof. For a full-column-rank matrix $\mathbf{B}$ with $\langle\mathbf{B}\rangle=U$, we have

$$A(m,U)=\{\mathbf{BD}:\mathbf{D}\in\mathrm{Fr}(\mathbb{F}^{\dim(U)\times m})\}=\mathbf{B}\,\mathrm{Fr}(\mathbb{F}^{\dim(U)\times m}).$$

Thus, $|A(m,U)|=|\mathrm{Fr}(\mathbb{F}^{\dim(U)\times m})|=\chi^m_{\dim(U)}$. For $\Phi\in\mathrm{Fr}(\mathbb{F}^{t\times t})$, $\langle\Phi\mathbf{B}\rangle=\Phi U$. So $A(m,\Phi U)=\Phi\mathbf{B}\,\mathrm{Fr}(\mathbb{F}^{r\times M})=\Phi A(m,U)$.

Fix any $V,V',U,U'\in\mathrm{Pj}(\mathbb{F}^T)$ satisfying $V\leq U$, $V'\leq U'$, $\dim(U)=\dim(U')=r$ and $\dim(V)=\dim(V')=s$. We show that there exists a full rank $T\times T$ matrix such that $\Phi V=V'$ and $\Phi U=U'$. Find a basis $\{\mathbf{b}_i:i=1,\ldots,s\}$ of V, extend the basis of V to a basis $\{\mathbf{b}_i:i=1,\ldots,r\}$ of U, and further extend the basis of U to a basis $\{\mathbf{b}_i:i=1,\ldots,T\}$ of $\mathbb{F}^T$. Similarly, find a basis $\{\mathbf{b}'_i:i=1,\ldots,T\}$ of $\mathbb{F}^T$ such that $\{\mathbf{b}'_i:i=1,\ldots,r\}$ is a basis of U and $\{\mathbf{b}'_i:i=1,\ldots,s\}$ is a basis of V. Consider the linear equation system

$$\Phi\mathbf{b}_i=\mathbf{b}'_i,\quad i=1,\ldots,T.$$

The unique solution of the above system satisfies $\Phi V=V'$ and $\Phi U=U'$. Using the above notations, we have

$$\begin{aligned}
&p_{\langle X\rangle\langle Y\rangle}(U,V)\\
&\quad=\sum_{\mathbf{X}\in A(M,U)}p_X(\mathbf{X})\sum_{\mathbf{Y}\in A(N,V)}P_{Y|X}(\mathbf{Y}|\mathbf{X})\\
&\quad=\sum_{\mathbf{X}\in A(M,U)}p_X(\Phi\mathbf{X})\sum_{\mathbf{Y}\in A(N,V)}P_{Y|X}(\Phi\mathbf{Y}|\Phi\mathbf{X})\qquad(19)\\
&\quad=\sum_{\mathbf{X}\in A(M,\Phi U)}p_X(\mathbf{X})\sum_{\mathbf{Y}\in A(N,\Phi V)}P_{Y|X}(\mathbf{Y}|\mathbf{X})\\
&\quad=p_{\langle X\rangle\langle Y\rangle}(\Phi U,\Phi V)\\
&\quad=p_{\langle X\rangle\langle Y\rangle}(U',V'),
\end{aligned}$$

where in (19) $p_X(\mathbf{X})=p_X(\Phi\mathbf{X})$ follows that p_X is uniform-given-row-space, and $P_{Y|X}(\Phi\mathbf{Y}|\Phi\mathbf{X})=P_{Y|X}(\mathbf{Y}|\mathbf{X})$ follows from Lemma 2. Then it can be verified that for $V,U\leq\mathbb{F}^T$ with $V\leq U$, $\dim(U)=r$ and $\dim(V)=s$,

$$p_{\langle X\rangle\langle Y\rangle}(U,V)=\frac{p_{\mathrm{rk}(X)\,\mathrm{rk}(Y)}(r,s)}{\begin{bmatrix}T\\r\end{bmatrix}\begin{bmatrix}r\\s\end{bmatrix}}.\tag{20}$$

Similarly, we can show that $p_{\langle X\rangle}(U)=p_{\langle X\rangle}(U')$ for U, $U'\leq\mathbb{F}^T$ with $\dim(U)=\dim(U')=r$, which implies

$$p_{\langle X\rangle}(U)=\frac{p_{\mathrm{rk}(X)}(r)}{\begin{bmatrix}T\\r\end{bmatrix}}.\tag{21}$$

Moreover, for $V \leq \mathbb{F}^T$ with $\dim(V) = s$,

$$\begin{aligned} p_{\langle Y\rangle}(V) &= \sum_{r\geq s} \sum_{U:V\leq U,\dim(U)=r} p_{\langle X\rangle\langle Y\rangle}(U,V) \\ &= \sum_{r\geq s} \frac{p_{\mathrm{rk}(X)\,\mathrm{rk}(Y)}(r,s)}{\begin{bmatrix}T\\r\end{bmatrix}\begin{bmatrix}r\\s\end{bmatrix}} \sum_{U:V\leq U,\dim(U)=r} 1 \\ &= \sum_{r\geq s} \frac{p_{\mathrm{rk}(X)\,\mathrm{rk}(Y)}(r,s)}{\begin{bmatrix}T\\r\end{bmatrix}\begin{bmatrix}r\\s\end{bmatrix}} \begin{bmatrix}T\\r\end{bmatrix} \frac{\chi_s^r}{\chi_s^T} \qquad (22) \\ &= \frac{p_{\mathrm{rk}(Y)}(s)}{\begin{bmatrix}T\\s\end{bmatrix}}, \qquad (23) \end{aligned}$$

where (22) is obtained by Lemma 1.

Substituting (20), (21) and (23) into $I(\langle X\rangle; \langle Y\rangle)$ completes the proof. ■

1) Optimal Uniform-Given-Row-Space Input Distribution for Subspace Coding: Define

$$C_{\mathrm{USS}} \triangleq \max_{p_{\langle X^\top\rangle}} [J(\mathrm{rk}(X); \mathrm{rk}(Y)) + I(\mathrm{rk}(X); \mathrm{rk}(Y))],$$

which is the maximum achievable rate of subspace coding using uniform-given-row-space input distribution. Rewrite $p_{\langle X^\top\rangle}(U) = p_{\mathrm{rk}(X)}(\dim(U)) P_{\langle X^\top\rangle|\,\mathrm{rk}(X)}(U|\dim(U))$. By treating $p_{\mathrm{rk}(X)}$ and $P_{\langle X^\top\rangle|\,\mathrm{rk}(X)}$ as variables, we can rewrite the above maximization problem as

$$\max_{p_{\mathrm{rk}(X)}} \max_{P_{\langle X^\top\rangle|\,\mathrm{rk}(X)}} [I(\mathrm{rk}(X); \mathrm{rk}(Y)) + J(\mathrm{rk}(X); \mathrm{rk}(Y))]. \qquad (24)$$

Both $I(\mathrm{rk}(X); \mathrm{rk}(Y))$ and $J(\mathrm{rk}(X); \mathrm{rk}(Y))$ depend on $P_{\mathrm{rk}(Y)|\,\mathrm{rk}(X)}$. We have

$$\begin{aligned} &P_{\mathrm{rk}(Y)|\,\mathrm{rk}(X)}(s|r) \\ &\quad= \sum_{U\in\mathrm{Gr}(r,\mathbb{F}^M)} P_{\mathrm{rk}(Y)|\langle X^\top\rangle}(s|U) P_{\langle X^\top\rangle|\,\mathrm{rk}(X)}(U|r), \end{aligned}$$

in which $P_{\mathrm{rk}(Y)|\langle X^\top\rangle}(s|U)$ is a function of p_H and is not related to $p_{\mathrm{rk}(X)}$ and $P_{\langle X^\top\rangle|\,\mathrm{rk}(X)}$ (see Lemma 4). The formulation of $J(\mathrm{rk}(X); \mathrm{rk}(Y))$ can be rewritten as

$$\begin{aligned} &J(\mathrm{rk}(X); \mathrm{rk}(Y)) \\ &= \sum_r p_{\mathrm{rk}(X)}(r) \sum_{U\in\mathrm{Gr}(r,\mathbb{F}^M)} P_{\langle X^\top\rangle|\,\mathrm{rk}(X)}(U|r) R_{H,T}(U), \qquad (25) \end{aligned}$$

where

$$R(U) = R_{H,T}(U) \triangleq \sum_s P_{\mathrm{rk}(Y)|\langle X^\top\rangle}(s|U) \log \frac{\chi_s^T}{\chi_s^{\dim(U)}} \qquad (26)$$

is only related to the distribution of H. Note that $R(U)$ is the achievable rate of subspace coding for the uniform-given-row-space input with $p_{\langle X^\top\rangle}(U) = 1$.

For fixed $p_{\mathrm{rk}(X)}$, $I(\mathrm{rk}(X); \mathrm{rk}(Y))$ is a convex function of $P_{\langle X^\top\rangle|\,\mathrm{rk}(X)}$, and $J(\mathrm{rk}(X); \mathrm{rk}(Y))$ is a linear function of $P_{\langle X^\top\rangle|\,\mathrm{rk}(X)}$. For fixed $P_{\langle X^\top\rangle|\,\mathrm{rk}(X)}$, $I(\mathrm{rk}(X); \mathrm{rk}(Y))$ is a concave function of $p_{\mathrm{rk}(X)}$, and $J(\mathrm{rk}(X); \mathrm{rk}(Y))$ is a linear function of $p_{\mathrm{rk}(X)}$. Therefore, $[I(\mathrm{rk}(X); \mathrm{rk}(Y)) + J(\mathrm{rk}(X); \mathrm{rk}(Y))]$ is not concave for $p_{\mathrm{rk}(X)}$ and $P_{\langle X^\top\rangle|\,\mathrm{rk}(X)}$. The following lemma characterizes a special optimizer of (24).

Lemma 8: There exists a deterministic transition matrix $P_{\langle X^\top\rangle|\,\mathrm{rk}(X)}$ achieving C_{USS}.

Proof: The proof is similar to the one of Lemma 7, and hence omitted. ■

2) Optimal Uniform-Given-Row-Space Input Distribution for Large T and q: We can further narrow down the range to search an optimal uniform-given-row-space input distribution when both T and q are large. For a random matrix H, define

$$\mathrm{rk}^*(H) \triangleq \max\{r : \Pr\{\mathrm{rk}(H) = r\} > 0\}.$$

Theorem 4: There exists T_0 and R_0 as functions of M and the rank distribution of H, such that when $T \geq T_0$ and $(T - M)\log q \geq R_0$, C_{USS} is achieved by the uniform-given-row-space input distribution with $\Pr\{\mathrm{rk}(X) \geq \mathrm{rk}^*(H)\} = 1$.

Proof: By Lemma 8, there exists a uniform-given-row-space input achieving C_{USS} such that $p_{\langle X^\top\rangle}(U(r)) = p_{\mathrm{rk}(X)}(r)$ for all $r \leq \min\{M, T\}$, where $\dim(U(r)) = r$. In other words, for r such that $p_{\mathrm{rk}(X)}(r) > 0$, $P_{\langle X^\top\rangle|\,\mathrm{rk}(X)}(U(r)|r) = 1$. We show by contradiction that $\Pr\{\mathrm{rk}(X) \geq \mathrm{rk}^*(H)\} = 1$ for sufficiently large T.

Consider an input distribution with $\Pr\{\mathrm{rk}(X) < \mathrm{rk}^*(H)\} > 0$. By Theorem 3 and (25),

$$C_{\mathrm{USS}} = I(\mathrm{rk}(X); \mathrm{rk}(Y)) + \sum_r p_{\mathrm{rk}(X)}(r) R(U(r)).$$

Define a uniform-given-row-space input distribution p'_X with $p'_{\langle X^\top\rangle}(U(r)) = p'_{\mathrm{rk}(X)}(r) = p_{\mathrm{rk}(X)}(r)$ for $\mathrm{rk}^*(H) \leq r < M$ and $p'_{\langle X^\top\rangle}(U(M)) = p'_{\mathrm{rk}(X)}(M) = p_{\mathrm{rk}(X)}(M) + \sum_{k<\mathrm{rk}^*(H)} p_{\mathrm{rk}(X)}(k)$. We have that

$$\begin{aligned} &I(\langle X\rangle; \langle Y\rangle)|_{p'_X} - C_{\mathrm{USS}} \\ &\geq \sum_{r=0}^{\mathrm{rk}^*(H)-1} p_{\mathrm{rk}(X)}(r)[R(\mathbb{F}^M) - R(U(r))] \\ &\quad - I(\mathrm{rk}(X); \mathrm{rk}(Y))|_{p_X} \\ &> \sum_{r=0}^{\mathrm{rk}^*(H)-1} p_{\mathrm{rk}(X)}(r)\Theta(T, r, H)\log q \\ &\quad - I(\mathrm{rk}(X); \mathrm{rk}(Y))|_{p_X}, \qquad (27) \end{aligned}$$

where the last inequality follows from Lemma 14 in Appendix C with

$$\begin{aligned} \Theta(T, r, H) = (T - M) &\sum_{k:k>r} \Pr\{\mathrm{rk}(H) \geq k\} \\ &- r(M - r) + \log_q \zeta_r^r. \end{aligned}$$

The quantity $\Theta(T, r, H)$ is a lower bound on $(R(\mathbb{F}^M) - R(U))/\log q$ with $\dim(U) = r$ and it is positive when T is sufficiently large.

Fix a sufficiently large T such that $\Theta(T, r, H) > 0$ for $r < \mathrm{rk}^*(H)$. Since $\Pr\{\mathrm{rk}(X) < \mathrm{rk}^*(H)\} > 0$ by assumption, we see that when $(T - M)\log q$ is sufficiently large, the RHS of (27) becomes positive, a contradiction to $C_{\mathrm{SS}}(H, T) \geq I(\langle X\rangle; \langle Y\rangle)$ for any input distribution. ■

3) Constant-Rank Uniform-Given-Row-Space Input Distributions: An input distribution with $p_{\mathrm{rk}(X)}(r) = 1$ is called a *constant-rank or rank-r input distribution*. Note that for a subspace degradation, using rank-r input is corresponding to using r-dimensional subspace coding.

For a constant-rank uniform-given-row-space input distribution, we always have $I(\mathrm{rk}(X); \mathrm{rk}(Y)) = 0$. So, together with (18), an optimal contant-rank uniform-given-row-space input distribution for subspace coding can be found by maximizing $J(\mathrm{rk}(X);\mathrm{rk}(Y))$. Define

$$\begin{aligned} C_{\mathrm{CUSS}} &\triangleq \max_{p_X:\text{constant-rank uniform-given-row-space}} J(\mathrm{rk}(X);\mathrm{rk}(Y)) \\ &= \max_{U\in \mathrm{Pj}(\min\{M,T\},\mathbb{F}^M)} R(U), \end{aligned} \tag{28}$$

where (28) follows from (25).

Since $I(\mathrm{rk}(X);\mathrm{rk}(Y)) \leq \log(\min\{T,M,N\}+1)$, the loss of rate by using constant-rank uniform-given-row-space input distribution is small when

$$C_{\mathrm{CUSS}} \gg \log(\min\{T,M,N\}+1). \tag{29}$$

By Lemma 5, we know that when $T > M$,

$$C_{\mathrm{CUSS}} > (T-M)\,\mathrm{E}[\mathrm{rk}(H)]\log q.$$

So when $(T-M)\,\mathrm{E}[\mathrm{rk}(H)]\log q \gg \log(\min\{T,M,N\}+1)$, it is reasonable to use constant-dimensional subspace coding.

Example 4: Consider $T-1=M=N=64$, $\mathrm{E}[\mathrm{rk}(H)]=32$, and $q=256$. We can calculate that $J(\mathrm{rk}(X);\mathrm{rk}(Y)) > 256$, while $\log(\min\{T,M,N\}+1) \approx 5$. So the loss of rate by using constant-rank uniform-given-row-space input distribution is small.

The following corollary is a direct result of Theorem 4 with the condition that $\mathrm{rk}^*(H)=M$.

Corollary 1: For a transfer matrix H with $\mathrm{rk}^*(H)=M$, when both T and $(T-M)\log q$ are sufficiently large, the optimal value of (24) is achieved by the uniform-given-row-space input with $p_{\mathrm{rk}(X)}(M)=1$, and the optimal value is $R(\mathbb{F}^M)=\sum_s p_{\mathrm{rk}(H)}(s)\log\frac{\chi_s^T}{\chi_s^M}$.

C. LOCs With a Unique Subspace Degradation

Now let us turn to LOCs with a unique subspace degradation, i.e., $P_{\langle Y\rangle|\langle X\rangle}$ is invariant with respect to $P_{X|\langle X\rangle}$. For such LOCs, we do not have the issue of finding an optimal subspace degradation—a subspace U can be converted to any matrix $\mathbf{X}$ with $\langle \mathbf{X}\rangle = U$. This property makes it easier to apply subspace coding on LOCs with a unique subspace degradation. As we will further show in this paper, all LOCs studied in existing literature have a unique subspace degradation, and some results previous obtained for special cases are actually shared by all LOCs with a unique subspace degradation.

By definition, a LOC has a unique subspace degradation if and only if for any V,

$$P_{\langle Y\rangle|X}(V|\mathbf{X}) = P_{\langle Y\rangle|X}(V|\mathbf{X}') \text{ whenever } \langle \mathbf{X}'\rangle = \langle \mathbf{X}\rangle. \tag{30}$$

If H is uniform-given-row-space, then the transition matrix of $\mathrm{LOC}(H,T)$ satisfies (30), and hence $\mathrm{LOC}(H,T)$ has a unique subspace degradation. Therefore, the LOCs studied in [17] and [18] with uniform-given-rank transfer matrices have a unique subspace degradation. Since a row-space-symmetric LOC has a uniform-given-row-space transfer matrix when $T \geq M$ (see Lemma 6), we have the following lemma.

Lemma 9: When $T \geq M$, a row-space-symmetric LOC has a unique subspace degradation.

When $T < M$, a row-space-symmetric LOC may not have a unique subspace degradation.

Example 5: Consider $\mathrm{LOC}(H,1)$. By (30), $\mathrm{LOC}(H,1)$ has a unique subspace degradation if and only if for any nonzero $x_1, x_2 \in \mathbb{F}^{1\times M}$,

$$P_{\langle Y\rangle|X}(\langle 0\rangle|x_1) = P_{\langle Y\rangle|X}(\langle 0\rangle|x_2) \tag{31}$$

$$P_{\langle Y\rangle|X}(\langle 1\rangle|x_1) = P_{\langle Y\rangle|X}(\langle 1\rangle|x_2). \tag{32}$$

However, (31) implies (32) since

$$P_{\langle Y\rangle|X}(\langle 1\rangle|x) = 1 - P_{\langle Y\rangle|X}(\langle 0\rangle|x).$$

The equalities in (31) give linear constraints on the distribution of H, from which we can find the set of H such that $\mathrm{LOC}(H,1)$ has a unique subspace degradation.

More examples of LOCs with a unique subspace degradation will be provided in Section V-C.

Lemma 10: A LOC has a unique subspace degradation if and only if

$$P_{\langle Y\rangle|X}(V|\mathbf{X}) = P_{\langle Y\rangle|X}(V'|\mathbf{X}') \tag{33}$$

whenever $\dim(V)=\dim(V')$, $\mathrm{rk}(\mathbf{X})=\mathrm{rk}(\mathbf{X}')$, $V \leq \langle \mathbf{X}\rangle$ *and* $V' \leq \langle \mathbf{X}'\rangle$.

Proof: The sufficient condition holds since (33) implies (30). We prove the necessary condition as follows. Fix a full column-rank matrix $\mathbf{B}_0$ such that $\langle \mathbf{B}_0\rangle = V$. Since $V \leq \langle \mathbf{X}\rangle$, we can find full rank matrix $\mathbf{B}_1$ and $\mathbf{D}$ such that $[\mathbf{B}_0\mathbf{B}_1]\mathbf{D} = \mathbf{X}$. Therefore,

$$\begin{aligned} &P_{\langle Y\rangle|X}(V|\mathbf{X}) \\ &\quad= P_{\langle Y\rangle|X}\left(V|[\mathbf{B}_0\mathbf{B}_1]\mathbf{D}\right) \\ &\quad= \sum_{\mathbf{Y}:\langle \mathbf{Y}\rangle = V} P_{Y|\langle X\rangle}\left(\mathbf{Y}|[\mathbf{B}_0\mathbf{B}_1]\mathbf{D}\right) \\ &\quad= \sum_{\mathbf{E}\in \mathrm{Fr}(\mathbb{F}^{\dim(V)\times N})} P_{Y|X}\left([\mathbf{B}_0\mathbf{B}_1]\begin{bmatrix}\mathbf{E}\\ \mathbf{0}\end{bmatrix}\middle|[\mathbf{B}_0\mathbf{B}_1]\mathbf{D}\right) \\ &\quad= \sum_{\mathbf{E}\in \mathrm{Fr}(\mathbb{F}^{\dim(V)\times N})} \Pr\left\{\mathbf{D}H = \begin{bmatrix}\mathbf{E}\\ \mathbf{0}\end{bmatrix}\right\}, \end{aligned} \tag{34}$$

where (34) follows from Lemma 2. If (30) holds, then (34) holds for any full row-rank $\mathrm{rk}(\mathbf{X}) \times M$ matrix $\mathbf{D}$, and hence (33) holds. ■

Recall the definition of uniform-given-dimension distributions over $\mathrm{Pj}(\mathbb{F}^T)$ in Definition 3.

Theorem 5: For a LOC with a unique subspace degradation, the capacity of the subspace degradation can be achieved by a uniform-given-dimension distribution, and

$$C_{\mathrm{SS}} = \max_{p_{\mathrm{rk}(X)}}[J(\mathrm{rk}(X);\mathrm{rk}(Y)) + I(\mathrm{rk}(X);\mathrm{rk}(Y))]. \tag{35}$$

Proof: For a LOC with a unique subspace degradation, $P_{\langle Y\rangle|\langle X\rangle}$ is well defined without specifying $p_{X|\langle X\rangle}$. So considering $p_{\langle X\rangle}$ is sufficient for $I(\langle X\rangle;\langle Y\rangle)$. We first show that there exists a uniform-given-dimension input distribution that maximizes $I(\langle X\rangle;\langle Y\rangle)$.

Fix a LOC with a unique subspace degradation. Let p be a distribution over $\mathrm{Pj}(\mathbb{F}^T)$ achieving the capacity of the subspace degradation, i.e., p achieves C_{SS}. For $\Phi \in \mathrm{Fr}(\mathbb{F}^{T\times T})$, define p^Φ as $p^\Phi(U) = p(\Phi U)$, where ΦU is defined in (2). First p^Φ is a PMF because $0 \leq p^\Phi(U) = p(\Phi U) \leq 1$ and $\sum_{U\in \mathrm{Pj}(\mathbb{F}^T)} p^\Phi(U) = 1$.

We show that p^{Φ} also achieves the capacity. For the simplicity of the notations, we write $p' = p^{\Phi}$. Let $p_{\langle Y\rangle}$ and $p'_{\langle Y\rangle}$ be the PMF of $\langle Y\rangle$ when the input distributions are p and p', respectively. We have

$$\begin{aligned} p'_{\langle Y\rangle}(V) &= \sum_{U\in \mathrm{Pj}(\mathbb{F}^T):V\le U} p'(U)P_{\langle Y\rangle|\langle X\rangle}(V|U) \\ &= \sum_{U\in \mathrm{Pj}(\mathbb{F}^T):V\le U} p(\Phi U)P_{\langle Y\rangle|\langle X\rangle}(\Phi V|\Phi U) \quad (36)\\ &= \sum_{U'\in \mathrm{Pj}(\mathbb{F}^T):\Phi V\le U'} p(U')P_{\langle Y\rangle|\langle X\rangle}(\Phi V|U') \\ &= p_{\langle Y\rangle}(\Phi V), \end{aligned}$$

where (36) follows from $p'(U) = p(\Phi U)$ and Lemma 10. Therefore,

$$\begin{aligned} &I(\langle X\rangle;\langle Y\rangle)|_{p'} \\ &= \sum_{U\in \mathrm{Pj}(\mathbb{F}^T)} p'(U) \sum_{V\in \mathrm{Pj}(\mathbb{F}^T):V\le U} P(V|U)\log\frac{P(V|U)}{p'_{\langle Y\rangle}(V)} \\ &= \sum_{U\in \mathrm{Pj}(\mathbb{F}^T)} p(\Phi U) \sum_{V\in \mathrm{Pj}(\mathbb{F}^T):V\le U} P(\Phi V|\Phi U) \\ &\quad\times \log\frac{P(\Phi V|\Phi U)}{p(\Phi V)} \\ &= \sum_{U'\in \mathrm{Pj}(\mathbb{F}^T)} p(U') \sum_{V'\in \mathrm{Pj}(\mathbb{F}^T):V'\le U'} P(V'|U')\log\frac{P(V'|U')}{p(V')} \\ &= I(\langle X\rangle;\langle Y\rangle)|_{p}, \end{aligned}$$

which implies that p' also achieves the subspace coding capacity.

Define p^* on $\mathrm{Pj}(\mathbb{F}^T)$ as

$$p^*(U) = \frac{1}{|\mathrm{Fr}(\mathbb{F}^{T\times T})|}\sum_{\Phi\in \mathrm{Fr}(\mathbb{F}^{T\times T})} p^{\Phi}(U).$$

Note that p^* is uniform-given-dimension. Since mutual information is a concave function of the input distribution [25],

$$I(\langle X\rangle;\langle Y\rangle)|_{p^*} \ge \frac{1}{|\mathrm{Fr}(\mathbb{F}^{T\times T})|}\sum_{\Phi\in \mathrm{Fr}(\mathbb{F}^{T\times T})} I(\langle X\rangle;\langle Y\rangle)|_{p^{\Phi}}.$$

Thus, p^* is also an optimal input distribution for the subspace channel.

Note that for a uniform-given-dimension LOC,

$$p_{\langle X\rangle}(V) = \frac{p_{\mathrm{rk}(X)}(\dim(V))}{\begin{bmatrix} T \\ \dim(V)\end{bmatrix}}.$$

So C_{SS} can be found by only optimizing over the input rank distribution $p_{\mathrm{rk}(X)}$.

If X is a uniform-given-row-space distribution, then $\langle X\rangle$ is uniform-given-dimension. For any uniform-given-dimension distribution p on $\mathrm{Pj}(\mathbb{F}^T)$ we can find a uniform-given-row-space distribution p' on $\mathbb{F}^{T\times M}$ such that $p(U) = \sum_{\mathbf{X}:\langle\mathbf{X}\rangle=U} p'(\mathbf{X})$. Hence, by Theorem 3 and the fact that a uniform-given-row-space input distribution p_X can be determined by $p_{\langle X^{\top}\rangle}$, we get

$$C_{\mathrm{SS}} = \max_{p_{\langle X^{\top}\rangle}}[J(\mathrm{rk}(X);\mathrm{rk}(Y)) + I(\mathrm{rk}(X);\mathrm{rk}(Y))].$$

Fix $U, U' \le \mathbb{F}^M$ with $\dim(U) = \dim(U')$. Find $\mathbf{X}_U$ and $\mathbf{X}_{U'}$ with $\langle\mathbf{X}_U\rangle = \langle\mathbf{X}_{U'}\rangle$, $\langle\mathbf{X}_U^{\top}\rangle = U$ and $\langle\mathbf{X}_{U'}^{\top}\rangle = U'$. By Lemma 4, $P_{\mathrm{rk}(Y)|\langle X^{\top}\rangle}(s|U) = P_{\mathrm{rk}(Y)|X}(s|\mathbf{X}_U)$ and $P_{\mathrm{rk}(Y)|\langle X^{\top}\rangle}(s|U') = P_{\mathrm{rk}(Y)|X}(s|\mathbf{X}_{U'})$. Further by Lemma 10,

$$\begin{aligned} P_{\mathrm{rk}(Y)|X}(s|\mathbf{X}_U) &= \sum_{V\in \mathrm{Gr}(s,\langle\mathbf{X}_U\rangle)} P_{\langle V\rangle|X}(V|\mathbf{X}_U) \\ &= \sum_{V\in \mathrm{Gr}(s,\langle\mathbf{X}_{U'}\rangle)} P_{\langle V\rangle|X}(V|\mathbf{X}_{U'}) \\ &= P_{\mathrm{rk}(Y)|X}(s|\mathbf{X}_{U'}). \end{aligned}$$

Therefore, $P_{\mathrm{rk}(Y)|\langle X^{\top}\rangle}(s|U) = P_{\mathrm{rk}(Y)|\langle X^{\top}\rangle}(s|U')$. Thus, $P_{\mathrm{rk}(Y)|\,\mathrm{rk}(X)}$ only depends on the distribution of H, and hence $p_{\mathrm{rk}(X)\,\mathrm{rk}(Y)}$ depends on $\langle X^{\top}\rangle$ only through $\mathrm{rk}(X)$. The proof is completed. ■

The above theorem implies that input distributions p_X with $p_{\langle X\rangle}$ being uniform-given-dimension achieve the subspace coding capacity for LOCs with a unique subspace degradation. Since only the input rank affects the subspace coding capacity, it has no penalty if we only consider uniform-given-rank input distributions for subspace coding.

Now, consider the computation of C_{SS} for LOCs with a unique subspace degradation, i.e., solving the maximization in (35). The problem is simpler than the one of computing C_{USS} (see (24)) since we do not need to optimize $P_{\langle X^{\top}\rangle|\,\mathrm{rk}(x)}$. The proof of Theorem 5 implies

$$P_{\mathrm{rk}(Y)|\,\mathrm{rk}(x)}(s|r) = P_{\mathrm{rk}(Y)|\langle X^{\top}\rangle}(s|U) = P_{\mathrm{rk}(Y)|X}(s|\mathbf{X}) \quad (37)$$

for any $U \in \mathrm{Gr}(r,\mathbb{F}^M)$ and any $\mathbf{X}$ with $\mathrm{rk}(\mathbf{X}) = r$. The optimization in (35) is convex and has $\min\{M,T\}$ variables.

Similar to $R_{H,T}(U)$ (defined in (26)), by abuse of notations, we define for LOCs with a unique subspace degradation

$$R(r) = R_{H,T}(r) = \sum_{s} P_{\mathrm{rk}(Y)|\,\mathrm{rk}(X)}(s|r)\log\frac{\chi_s^T}{\chi_s^r}.$$

Actually, $R_{H,T}(\dim(U)) = R_{H,T}(U)$ and hence we can rewrite

$$J(\mathrm{rk}(X);\mathrm{rk}(Y)) = \sum_{r} p_{\mathrm{rk}(X)}(r)R(r).$$

When applying on LOC with a unique subspace degradations, the same result of Theorem 4 still holds (with C_{SS} in place of C_{USS}) and the proof can be simplified by using $R(r)$ instead of $R(U)$. Similar to the discussion around (28), the maximum achievable rate of constant-rank input distributions is

$$\max_{p_{\mathrm{rk}(X)}} J(\mathrm{rk}(X);\mathrm{rk}(Y)) = \max_{r} R(r).$$

Example 6: Let's apply the above general discussion on LOCs with uniform-given-rank transfer matrices. Assume that $p_{\mathrm{rk}(H)}$ is known. To compute $P_{\mathrm{rk}(Y)|\,\mathrm{rk}(X)}(s|r)$, we choose the input matrix

$$\mathbf{X}^{(r)} = \begin{bmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix}.$$

For transfer matrix $H = \begin{bmatrix} H_1 \\ H_2 \end{bmatrix}$ where H_1 has r rows and H_2 has $M-r$ rows, the output matrix is $\begin{bmatrix} H_1 \\ \mathbf{0} \end{bmatrix}$. So

$$\begin{aligned} &P_{\mathrm{rk}(Y)|\,\mathrm{rk}(X)}(s|r) \\ &\quad = P_{\mathrm{rk}(Y)|X}(s|\mathbf{X}^{(r)}) = \Pr\{\mathrm{rk}(H_1) = s\} \\ &\quad = \sum_{k=s}^{\min\{M,N\}} \Pr\{\mathrm{rk}(H_1) = s|\,\mathrm{rk}(H) = k\} p_{\mathrm{rk}(H)}(k). \end{aligned}$$

Since the transfer matrix is uniform-given-rank, we have

$$\begin{aligned} &\Pr\{\mathrm{rk}(H_1) = s|\,\mathrm{rk}(H) = k\} \\ &\quad = \frac{|\{\mathbf{H} \in \mathbb{F}^{M\times N} : \mathrm{rk}(\mathbf{H}_1) = s, \mathrm{rk}(\mathbf{H}) = k\}|}{|\{\mathbf{H} \in \mathbb{F}^{M\times N} : \mathrm{rk}(\mathbf{H}) = k\}|}, \end{aligned}$$

where the RHS can be counted using the techniques introduced in Preliminaries. After computing $P_{\mathrm{rk}(Y)|\,\mathrm{rk}(X)}(s|r)$, $R(r)$ can be computed accordingly. Then the subspace coding capacity, as well as an optimal input rank distribution, can be obtained by solving

$$\begin{aligned} \max_{p(r)} \quad & \sum_r p(r)R(r) + \sum_r p(r) \sum_s P_{\mathrm{rk}(Y)|\,\mathrm{rk}(X)}(s|r) \\ & \times \log \frac{P_{\mathrm{rk}(Y)|\,\mathrm{rk}(X)}(s|r)}{\sum_{r'} p(r')P_{\mathrm{rk}(Y)|\,\mathrm{rk}(X)}(s|r')} \\ \text{s.t.} \quad & p(r) \geq 0, \ \sum_r p(r) = 1. \end{aligned}$$

We would not go into the details about solving the above optimization problem. Readers are referred to [17] and [18] to find more results about LOCs with uniform-given-rank transfer matrices.

V. SHANNON CAPACITY VS SUBSPACE CODING CAPACITY

In this section, we discuss some necessary conditions and sufficient conditions for a LOC such that $C = C_{\mathrm{SS}}$ as applications of the results obtained in the previous sections. A new class of LOCs such that $C = C_{\mathrm{SS}}$ is explicitly characterized.

A. Unique Subspace Degradation

Theorem 2 says

$$C \geq C_L \triangleq \max_{p_{\langle X^\top\rangle}} \left[J(\mathrm{rk}(X); \mathrm{rk}(Y)) + I(\langle X^\top\rangle; \langle Y^\top\rangle)\right],$$

and for a LOC with a unique subspace degradation Theorem 5 shows

$$C_{\mathrm{SS}} = \max_{p_{\langle X^\top\rangle}} [J(\mathrm{rk}(X); \mathrm{rk}(Y)) + I(\mathrm{rk}(X); \mathrm{rk}(Y))].$$

The above bounds imply a necessary condition such that $C = C_{\mathrm{SS}}$.

Theorem 6: Consider a LOC with a unique subspace degradation. If $C = C_{SS}$, then for certain $p_{\langle X^\top\rangle}$ that achieves C_L, $\langle X^\top\rangle \to \mathrm{rk}(X) \to \mathrm{rk}(Y) \to \langle Y^\top\rangle$ is a Markov chain. In other words, subspace coding is not capacity achieving if the LOC does not satisfy the Markov condition $\langle X^\top\rangle \to \mathrm{rk}(X) \to \mathrm{rk}(Y) \to \langle Y^\top\rangle$ for any $p_{\langle X^\top\rangle}$ achieving C_L.

Proof: Fix a LOC with a unique subspace degradation and $C = C_{\mathrm{SS}}$. If there is no $p_{\langle X^\top\rangle}$ achieving C_L and C_{SS} simultaneously, $C > C_{\mathrm{SS}}$. Consider a distribution $p^*_{\langle X^\top\rangle}$ of $\langle X^\top\rangle$ that achieves C_L and C_{SS} simultaneously, for which we have $I(\langle X^\top\rangle; \langle Y^\top\rangle) = I(\mathrm{rk}(X); \mathrm{rk}(Y))$, which implies $I(\langle X^\top\rangle; \langle Y^\top\rangle|\,\mathrm{rk}(Y)) = 0$ and $I(\langle X^\top\rangle; \mathrm{rk}(Y)|\,\mathrm{rk}(X)) = 0$. So both $\langle X^\top\rangle \to \mathrm{rk}(X) \to \mathrm{rk}(Y)$ and $(\langle X^\top\rangle, \mathrm{rk}(X)) \to \mathrm{rk}(Y) \to \langle Y^\top\rangle$ form Markov chains. Hence $\langle X^\top\rangle \to \mathrm{rk}(X) \to \mathrm{rk}(Y) \to \langle Y^\top\rangle$ is a Markov chain. ■

We know that for a distribution $p_{\langle X^\top\rangle}$, $\langle X^\top\rangle \to \mathrm{rk}(X) \to \mathrm{rk}(Y) \to \langle Y^\top\rangle$ is a Markov chain if and only if

$$\begin{aligned} &p_{\mathrm{rk}(X)}(r)p_{\mathrm{rk}(Y)}(s)p_{\langle X^\top\rangle\,\mathrm{rk}(X)\langle Y^\top\rangle\,\mathrm{rk}(Y)}(U, r, V, s) \\ &\quad = p_{\langle X^\top\rangle\,\mathrm{rk}(X)}(U, r)p_{\mathrm{rk}(X)\,\mathrm{rk}(Y)}(r, s)p_{\langle Y^\top\rangle\,\mathrm{rk}(Y)}(V, s), \\ &\qquad \forall r, s, U, V, \end{aligned}$$

which is equivalent to

$$\begin{aligned} &p_{\mathrm{rk}(X)}(r)p_{\mathrm{rk}(Y)}(s)p_{\langle X^\top\rangle\langle Y^\top\rangle}(U, V) \\ &\quad = p_{\langle X^\top\rangle}(U)p_{\mathrm{rk}(X)\,\mathrm{rk}(Y)}(r, s)p_{\langle Y^\top\rangle}(V), \\ &\qquad \forall r \geq s, \dim(U) = r, \dim(V) = s. \end{aligned} \tag{38}$$

For U such that $p_{\langle X^\top\rangle}(U) > 0$, the equality in (38) becomes

$$p_{\mathrm{rk}(Y)}(s)P_{\langle Y^\top\rangle|\langle X^\top\rangle}(V|U) = p_{\mathrm{rk}(Y)|\,\mathrm{rk}(X)}(s|r)p_{\langle Y^\top\rangle}(V),$$

where $r = \dim(U) \geq \dim(V) = s$. Thus, for each V, among all $U \in \mathrm{Gr}(r, \mathbb{F}^M)$ with $p_{\langle X^\top\rangle}(U) > 0$, $p_{\langle Y^\top\rangle|\langle X^\top\rangle}(V|U)$ are the same. Therefore, we can have the following lemma.

Lemma 11: If $p^*_{\langle X^\top\rangle}$ achieves C_L and satisfies the Markov chain $\langle X^\top\rangle \to \mathrm{rk}(X) \to \mathrm{rk}(Y) \to \langle Y^\top\rangle$, then there exists $p'_{\langle X^\top\rangle}$ achieving C_L such that

1) *for each r there exists at most one $U_r \in \mathrm{Gr}(r, \mathbb{F}^M)$ such that $p'_{\langle X^\top\rangle}(U_r) > 0$; and*
2) *the Markov chain $\langle X^\top\rangle \to \mathrm{rk}(Y) \to \langle Y^\top\rangle$ holds.*

Proof: Let

$$f(p_{\langle X^\top\rangle}) = J(\mathrm{rk}(X); \mathrm{rk}(Y)) + I(\langle X^\top\rangle; \langle Y^\top\rangle).$$

Then $C_L = \max_{p_{\langle X^\top\rangle}} f(p_{\langle X^\top\rangle})$. We have

$$\begin{aligned} \frac{\partial f(p_{\langle X^\top\rangle})}{\partial p_{\langle X^\top\rangle}(U)} = & \sum_{s=0}^{\dim(U)} \sum_{V\in\mathrm{Gr}(s,\mathbb{F}^N)} P(V|U) \log \frac{P(V|U)}{p_{\langle Y^\top\rangle}(V)} \\ & + R(U) - \log e. \end{aligned} \tag{39}$$

Let $p'_{\langle X^\top\rangle}$ be a distribution on $\mathrm{Pj}(\mathbb{F}^M)$ such that for each r with $p^*_{\mathrm{rk}(X)}(r) > 0$, there exists $U_r \in \mathrm{Gr}(r, \mathbb{F}^M)$ such that $p'_{\langle X^\top\rangle}(U_r) = \sum_{U\in\mathrm{Gr}(r,\mathbb{F}^M)} p^*_{\langle X^\top\rangle}(U)$ and $p^*_{\langle X^\top\rangle}(U_r) > 0$.

Let $p^*_{\langle Y^\top\rangle}$ and $p'_{\langle Y^\top\rangle}$ be the distribution of $\langle Y^\top\rangle$ with respect to $p^*_{\langle X^\top\rangle}$ and $p'_{\langle X^\top\rangle}$, respectively. We have

$$\begin{aligned} &p^*_{\langle Y^\top\rangle}(V) \\ &\quad = \sum_r \sum_{U\in\mathrm{Gr}(r,\mathbb{F}^M):p_{\langle X^\top\rangle}(U)>0} P_{\langle Y^\top\rangle|\langle X^\top\rangle}(V|U)p_{\langle X^\top\rangle}(U) \\ &\quad = \sum_r \sum_{U\in\mathrm{Gr}(r,\mathbb{F}^M):p_{\langle X^\top\rangle}(U)>0} P_{\langle Y^\top\rangle|\langle X^\top\rangle}(V|U_r)p_{\langle X^\top\rangle}(U) \end{aligned}$$

$$\begin{aligned} &= \sum_r P_{\langle Y^\top\rangle|\langle X^\top\rangle}(V|U_r)\, p'_{\langle X^\top\rangle}(U_r) \\ &= p'_{\langle Y^\top\rangle}(V), \end{aligned}$$

where the second equality follows from the discussion after (38). By checking the KKT condition [28], $p'_{\langle X^\top\rangle}$ achieves C_L since $\frac{\partial f(p_{\langle X^\top\rangle})}{\partial p_{\langle X^\top\rangle}(U)}\Big|_{p'_{\langle X^\top\rangle}} = \frac{\partial f(p_{\langle X^\top\rangle})}{\partial p_{\langle X^\top\rangle}(U)}\Big|_{p^*_{\langle X^\top\rangle}}$.

For r with $p_{\langle X^\top\rangle}(r) > 0$, we further have

$$\begin{aligned} &p_{\mathrm{rk}(Y)|\,\mathrm{rk}(X)}(s|r) \\ &\quad= \sum_{U \in \mathrm{Gr}(r,\mathbb{F}^M): p_{\langle X^\top\rangle}(U)>0} P_{\mathrm{rk}(Y)|\langle X^\top\rangle}(s|U)\, p_{\langle X^\top\rangle|\,\mathrm{rk}(X)}(U|r) \\ &\quad= P_{\mathrm{rk}(Y)|\langle X^\top\rangle}(s|U_r) \\ &\quad= p'_{\mathrm{rk}(Y)|\,\mathrm{rk}(X)}(s|r). \end{aligned}$$

Therefore, $p'_{\langle X^\top\rangle}$ also satisfies the Markov condition in (38). Note that since in this case, the distributions of $\mathrm{rk}(X)$ and $\langle X^\top\rangle$ are the same, we equivalently have the Markov condition $\langle X^\top\rangle \to \mathrm{rk}(Y) \to \langle Y^\top\rangle$. ■

Using Lemma 11, the sufficient condition in Theorem 6 can be refined, and more explicit necessary conditions can be obtained for special cases.

Example 7: Suppose that $C = C_{\mathrm{SS}}$ for certain $\mathrm{LOC}(H, 1)$ with a unique subspace degradation. Fix $p'_{\langle X^\top\rangle}$ such that 1) $p'_{\langle X^\top\rangle}$ achieves C_L and 2) $p'_{\langle X^\top\rangle}(U_1) = 1 - p_{\mathrm{rk}(X)}(0)$ for $U_1 \in \mathrm{Gr}(1, \mathbb{F}^M)$. The existence of such $p_{\langle X^\top\rangle}$ is guaranteed by Theorem 6 and Lemma 11. Using (39), we have for $U \in \mathrm{Gr}(1, \mathbb{F}^M)$

$$\begin{aligned} &\frac{\partial f(p_{\langle X^\top\rangle})}{\partial p_{\langle X^\top\rangle}(U)}\Big|_{p=p'} \\ &= \sum_{s=0}^{1} \sum_{V \in \mathrm{Gr}(s,\mathbb{F}^N)} P(V|U) \log \frac{P(V|U)}{p_{\langle Y^\top\rangle}(V)} - \log e \\ &= P(\langle \mathbf{0}\rangle|U) \log \frac{P(\langle \mathbf{0}\rangle|U)}{p_{\langle Y^\top\rangle}(\langle \mathbf{0}\rangle)} \\ &\quad + \sum_{V \in \mathrm{Gr}(1,\mathbb{F}^N)} P(V|U) \log \frac{P(V|U)}{P(V|U_1)\, p_{\langle X^\top\rangle}(U_1)} - \log e \\ &= -\log p'_{\langle X^\top\rangle}(U_1) - \log e + D_{\mathrm{KL}}(P(\cdot|U)||P(\cdot|U_1)) \\ &\quad + P(\langle \mathbf{0}\rangle|U) \log \frac{P(\langle \mathbf{0}\rangle|U_1)\, p'_{\langle X^\top\rangle}(U_1)}{p_{\langle Y^\top\rangle}(\langle \mathbf{0}\rangle)}, \end{aligned}$$

where D_{KL} is the Kullback-Leibler divergence (see [26]). By (31), $P_{\langle Y^\top\rangle|\langle X^\top\rangle}(\langle \mathbf{0}\rangle|U) = P_{\langle Y^\top\rangle|\langle X^\top\rangle}(\langle \mathbf{0}\rangle|U_1)$ for all $U \in \mathrm{Gr}(1, \mathbb{F}^M)$. Since $p_{\langle X^\top\rangle}$ achieves C_L, by the KKT condition, we have for all $U \neq U_1$

$$\frac{\partial f(p_{\langle X^\top\rangle})}{\partial p_{\langle X^\top\rangle}(U)}\Big|_{p=p'} \leq \frac{\partial f(p_{\langle X^\top\rangle})}{\partial p_{\langle X^\top\rangle}(U_1)}\Big|_{p=p'}$$

which implies $D_{\mathrm{KL}}(P(\cdot|U)||P(\cdot|U_1)) = 0$. Therefore, for $\mathrm{LOC}(H, 1)$ with a unique subspace degradation, a necessary condition such that subspace coding is capacity achieving is that for each $V \in \mathrm{Gr}(1, \mathbb{F}^N)$, $P_{\langle Y^\top\rangle|\langle X^\top\rangle}(V|U) = P_{\langle Y^\top\rangle|\langle X^\top\rangle}(V|U')$ for all $U, U' \in \mathrm{Gr}(1, \mathbb{F}^M)$.

We can get a stronger result if the LOC with a unique subspace degradation is also row-space symmetric. Note that when $T \geq M$, a row-space-symmetric LOC has a unique subspace degradation (see Lemma 9).

Corollary 2: For a row-space-symmetric LOC which has a unique subspace degradation, $C = C_{\mathrm{SS}}$ if and only if for certain $p_{\langle X^\top\rangle}$ that achieves C, $\langle X^\top\rangle \to \mathrm{rk}(X) \to \mathrm{rk}(Y) \to \langle Y^\top\rangle$ is a Markov chain.

Proof: For a row-space-symmetric LOC, $C = C_L$. So the necessary condition follows from Theorem 6. On the other hand, assume $\langle X^\top\rangle \to \mathrm{rk}(X) \to \mathrm{rk}(Y) \to \langle Y^\top\rangle$ is a Markov chain for certain $p_{\langle X^\top\rangle}$ that achieves C. So $I(\langle X^\top\rangle; \langle Y^\top\rangle) = I(\mathrm{rk}(X); \mathrm{rk}(Y))$ for this $p_{\langle X^\top\rangle}$. Therefore $C = J(\mathrm{rk}(X); \mathrm{rk}(Y)) + I(\langle X^\top\rangle; \langle Y^\top\rangle) = J(\mathrm{rk}(X); \mathrm{rk}(Y)) + I(\mathrm{rk}(X); \mathrm{rk}(Y)) \leq C_{\mathrm{SS}}$, which implies $C = C_{\mathrm{SS}}$. ■

Example 8: Following Example 2, we discuss $\mathrm{LOC}_2(H, 1)$ with a unique subspace degradation. We know H satisfies (31). Since $\mathrm{LOC}_2(H, 1)$ is row-space-symmetric, we can apply the necessary and sufficient for $C = C_{\mathrm{SS}}$ given in Corollary 2. Similar to the discussion in Example 7, we have that for $\mathrm{LOC}_2(H, 1)$ with a unique subspace degradation, $C = C_{\mathrm{SS}}$ if and only if for each $y \in \mathbb{F}^{1\times N}$

$$P_{Y|X}(y|x_1) = P_{Y|X}(y|x_2), \ \forall x_1, x_2 \in \mathbb{F}^{1\times M}. \tag{40}$$

We will connect the above condition to another class of LOCs to be discussed.

B. Row-Space-Symmetric LOCs ($T < M$)

When $T \geq M$, a row-space-symmetric LOC has a unique subspace degradation (see Lemma 9). Hence, we can apply the results in the last subsection. But when $T < M$, a row-space-symmetric LOC may not have a unique subspace degradation. The following lemma gives an upper bound on the subspace coding capacity of row-space-symmetric LOCs.

Lemma 12: For a row-space-symmetric LOC,

$$C_{\mathrm{SS}} \leq \max_{p_{\langle X^\top\rangle}} \left[J(\mathrm{rk}(X); \mathrm{rk}(Y)) + I(\langle X^\top\rangle; \mathrm{rk}(Y)) \right].$$

Proof: See Appendix D. ■

Theorem 7: Consider a row-space-symmetric LOC.

1) *(Necessary condition) If $C = C_{SS}$, then for certain $p_{\langle X^\top\rangle}$ that achieves C, $\langle X^\top\rangle \to \mathrm{rk}(Y) \to \langle Y^\top\rangle$ is a Markov chain. In other words, subspace coding is not capacity achieving if the LOC does not satisfy the Markov condition $\langle X^\top\rangle \to \mathrm{rk}(Y) \to \langle Y^\top\rangle$ for all $p_{\langle X^\top\rangle}$ achieving C.*
2) *(Sufficient condition) If for certain $p_{\langle X^\top\rangle}$ that achieves C, $\langle X^\top\rangle \to \mathrm{rk}(X) \to \mathrm{rk}(Y) \to \langle Y^\top\rangle$ is a Markov chain, then $C = C_{SS}$.*

Proof: We first prove the necessary condition. Fix a row-space-symmetric LOC such that $C = C_{\mathrm{SS}}$. By Lemma 12,

$$C_{\mathrm{SS}} \leq R^U \triangleq \max_{p_{\langle X^\top\rangle}} \left[J(\mathrm{rk}(X); \mathrm{rk}(Y)) + I(\langle X^\top\rangle; \mathrm{rk}(Y)) \right].$$

On the other hand, by Theorem 2,

$$C = \max_{p_{\langle X^\top\rangle}} \left[J(\mathrm{rk}(X); \mathrm{rk}(Y)) + I(\langle X^\top\rangle; \langle Y^\top\rangle) \right].$$

Since $I(\langle X^\top\rangle;\langle Y^\top\rangle) \geq I(\langle X^\top\rangle;\mathrm{rk}(Y))$ for any $p_{\langle X^\top\rangle}$, if there exists no $p_{\langle X^\top\rangle}$ achieving C and R^U simultaneously, $C > R^U \geq C_{\mathrm{SS}}$, a contradiction to $C = C_{\mathrm{SS}}$. Fix $p_{\langle X^\top\rangle}$ that achieves C and R^U simultaneously. We have $I(\langle X^\top\rangle;\langle Y^\top\rangle) = I(\langle X^\top\rangle;\mathrm{rk}(Y))$, which implies $I(\langle X^\top\rangle;\langle Y^\top\rangle|\,\mathrm{rk}(Y)) = 0$, i.e., $\langle X^\top\rangle \to \mathrm{rk}(Y) \to \langle Y^\top\rangle$ is a Markov chain.

Now we show the sufficient condition. Fix a $p_{\langle X^\top\rangle}$ that achieves C and for which $\langle X^\top\rangle \to \mathrm{rk}(X) \to \mathrm{rk}(Y) \to \langle Y^\top\rangle$ is a Markov chain. So $I(\langle X^\top\rangle;\langle Y^\top\rangle) = I(\mathrm{rk}(X);\mathrm{rk}(Y))$ for this $p_{\langle X^\top\rangle}$. Thus

$$\begin{aligned} C &= J(\mathrm{rk}(X);\mathrm{rk}(Y)) + I(\langle X^\top\rangle;\langle Y^\top\rangle) \\ &= J(\mathrm{rk}(X);\mathrm{rk}(Y)) + I(\mathrm{rk}(X);\mathrm{rk}(Y)) \leq C_{\mathrm{SS}}, \end{aligned}$$

where the last inequality follows from Theorem 3. Therefore $C = C_{\mathrm{SS}}$. ■

To verify the sufficient condition given in Theorem 7, we do not need to check all input distributions that achieve C. If $p_{\langle X^\top\rangle}$ satisfies the sufficient condition in Theorem 7, we can apply Lemma 11 on $p_{\langle X^\top\rangle}$ since $C = C_L$ for row-space-symmetric LOCs, and obtain that $p'_{\langle X^\top\rangle}$ satisfies the sufficient condition and has the structure defined in Lemma 11-1). Therefore, we only need to check the sufficient condition for input distributions with the structure that for each r, $p_{\mathrm{rk}(X)}(r) = p_{\langle X^\top\rangle}(U_r)$ for certain $U_r \in \mathrm{Gr}(r,\mathbb{F}^M)$.

Example 9: Since $\mathrm{LOC}_2(H,1)$ is row-space-symmetric for any H (see Example 2), we can use Theorem 7 to characterize a sufficient condition such that $C = C_{\mathrm{SS}}$.

Consider an input distribution p^* with $p^*_{\mathrm{rk}(X)}(0) = p_0$ and $p^*_{\mathrm{rk}(X)}(1) = p^*_{\langle X^\top\rangle}(U_1) = 1 - p_0 = p_1$ for some $U_1 \in \mathrm{Gr}(1,\mathbb{F}^M)$. Hence $p^*_{\langle X^\top\rangle}(U) = 0$ for all $U \neq U_1 \in \mathrm{Gr}(1,\mathbb{F}^M)$. We first check the sufficient condition. By (38), for p^*, the Markov chain in the sufficient condition holds for any choices of p_0, $0 \leq p_0 \leq 1$ and U_1. To satisfy the sufficient condition, we further require that p^* achieves C. Now we assume $0 < p_0 < 1$ since otherwise, p^* would not be capacity achieving unless the channel is trivial. A necessary and sufficient condition such that p^* achieves C is given by the KKT condition:

$$\begin{aligned} C &= \log \frac{1}{p_{\mathrm{rk}(Y)}(0)}, \\ C &= -\log p_1 + P_{\mathrm{rk}(Y)|\langle X^\top\rangle}(0|U_1)\log \frac{P_{\mathrm{rk}(Y)|\langle X^\top\rangle}(0|U_1)p_1}{p_{\mathrm{rk}(Y)}(0)}, \\ C &\geq -\log p_1 + P_{\mathrm{rk}(Y)|\langle X^\top\rangle}(0|U)\log \frac{P_{\mathrm{rk}(Y)|\langle X^\top\rangle}(0|U_1)p_1}{p_{\mathrm{rk}(Y)}(0)} \\ &\quad + D_{\mathrm{KL}}(P_{\langle Y^\top\rangle|\langle X^\top\rangle}(\cdot|U)||P_{\langle Y^\top\rangle|\langle X^\top\rangle}(\cdot|U_1)). \end{aligned}$$

Note that the first two equalities fix p_1 and C as functions of $P_{\mathrm{rk}(Y)|\langle X^\top\rangle}(0|U_1)$. The third inequality gives a constraint for U_1, i.e., for all $U \neq U_1 \in \mathrm{Gr}(1,\mathbb{F}^M)$, we have

$$\begin{aligned} &D_{\mathrm{KL}}(P(\cdot|U)||P(\cdot|U_1)) \\ &\quad \leq (P_{\mathrm{rk}(Y)|\langle X^\top\rangle}(0|U) - P_{\mathrm{rk}(Y)|\langle X^\top\rangle}(0|U_1)) \\ &\quad\quad \times \log \frac{p_{\mathrm{rk}(Y)}(0)}{P_{\mathrm{rk}(Y)|\langle X^\top\rangle}(0|U_1)p_1}. \end{aligned}$$

Substituting the value of p_1, we have,

$$\begin{aligned} &D_{\mathrm{KL}}(P(\cdot|U)||P(\cdot|U_1)) \\ &\quad \leq \frac{P_{\mathrm{rk}(Y)|\langle X^\top\rangle}(0|U) - P_{\mathrm{rk}(Y)|\langle X^\top\rangle}(0|U_1)}{P_{\mathrm{rk}(Y)|\langle X^\top\rangle}(1|U_1)} \\ &\quad\quad \times \log \frac{1}{P_{\mathrm{rk}(Y)|\langle X^\top\rangle}(0|U_1)},\ \forall U \neq U_1 \in \mathrm{Gr}(1,\mathbb{F}^M). \quad (41) \end{aligned}$$

Therefore, if (41) holds, there exists p_1 such that p^* is capacity achieving, and hence $C = C_{\mathrm{SS}}$.

C. Degraded Linear Operator Channels

Definition 7: A LOC is *degraded* if $I(X;Y) = I(\langle X\rangle;\langle Y\rangle)$ for all p_X.

By definition, it is clear that a degraded LOC has $C = C_{\mathrm{SS}}$. Some degraded LOCs have been studied in the literature. When $M = N$, the LOC with H uniformly distributed among all full rank $M \times M$ matrices is degraded [15]. If H contains uniformly i.i.d. components, it was shown that the corresponding LOC is also degraded [16]. LOCs with uniform-given-rank transfer matrices [17], [18] are degraded, and uniform-given-rank transfer matrices include the transfer matrices studied in [15] and [16] as special cases.

In this section, we focus on the general properties of degraded LOCs. Since

$$\begin{aligned} &I(X;Y) \\ &\quad = \sum_{V,U\in\mathrm{Pj}(\mathbb{F}^T)} \sum_{\substack{\mathbf{X},\mathbf{Y}:\\ \langle\mathbf{X}\rangle=U,\langle\mathbf{Y}\rangle=V}} p(\mathbf{X},\mathbf{Y})\log\frac{p(\mathbf{X},\mathbf{Y})}{p_X(\mathbf{X})p_Y(\mathbf{Y})} \\ &\quad \geq \sum_{V,U\in\mathrm{Pj}(\mathbb{F}^T)} p_{\langle X\rangle\langle Y\rangle}(U,V)\log\frac{p_{\langle X\rangle\langle Y\rangle}(U,V)}{p_{\langle X\rangle}(U)p_{\langle Y\rangle}(V)} \\ &\quad = I(\langle X\rangle;\langle Y\rangle), \end{aligned}$$

where the inequality follows from the log-sum inequality (see [26]), a LOC is degraded if and only if

$$\forall \mathbf{Y},\ P_{Y|X}(\mathbf{Y}|\mathbf{X}) = P_{Y|X}(\mathbf{Y}|\mathbf{X}') \text{ if } \langle\mathbf{X}\rangle = \langle\mathbf{X}'\rangle, \quad (42)$$

and for all p_X

$$\forall \mathbf{X},\ \frac{P_{Y|X}(\mathbf{Y}|\mathbf{X})}{p_Y(\mathbf{Y})} = \frac{P_{Y|X}(\mathbf{Y}'|\mathbf{X})}{p_Y(\mathbf{Y}')} \text{ if } \langle\mathbf{Y}\rangle = \langle\mathbf{Y}'\rangle. \quad (43)$$

Example 10: We check when $\mathrm{LOC}_2(H,1)$ is degraded. For this example, (43) holds trivially and (42) is equivalent to

$$P_{Y|X}(y|x_1) = P_{Y|X}(y|x_2),\ \forall y, x_1, x_2 \in \mathbb{F}^{1\times 2}. \quad (44)$$

We have at most six linear constraints on the distribution of H such that $\mathrm{LOC}_2(H,1)$ is degraded.

Note that (44) is equivalent to (40). Hence we can rephrase the conclusion of Example 8 as $\mathrm{LOC}_2(H,1)$ with a unique subspace degradation has $C = C_{\mathrm{SS}}$ if and only if it is degraded. However, a LOC may not be degraded even if $C = C_{\mathrm{SS}}$. As an example, for the distribution of $H_2 \in \mathbb{F}^{2\times 2}$ in Table I, $\mathrm{LOC}_2(H_2,1)$ has multiple subspace degradations, but $C = C_{\mathrm{SS}} = 1$bit. The optimal input distribution has $p_{\mathrm{rk}(X)}(0) = p_{\langle X^\top\rangle}(\langle[1\ 1]^\top\rangle) = 0.5$.

Theorem 8: A degraded LOC has a unique subspace degradation and it is row-space-symmetric.

TABLE I

A DISTRIBUTION OVER $\mathbb{F}_2^{2\times 2}$. EACH NUMBERED CELL IS THE PROBABILITY MASS OF THE MATRIX WHOSE FIRST COLUMN IS THE ROW INDEX OF THE TABLE AND SECOND COLUMN IS THE COLUMN INDEX OF THE TABLE

	$\begin{bmatrix}0\\0\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}1\\1\end{bmatrix}$
$\begin{bmatrix}0\\0\end{bmatrix}$	0	0	$\frac{1}{6}$	0
$\begin{bmatrix}1\\0\end{bmatrix}$	0	0	$\frac{1}{12}$	$\frac{1}{12}$
$\begin{bmatrix}0\\1\end{bmatrix}$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$
$\begin{bmatrix}1\\1\end{bmatrix}$	0	$\frac{1}{12}$	$\frac{1}{12}$	0

TABLE II

A DISTRIBUTION OVER $\mathbb{F}_2^{2\times 2}$. EACH NUMBERED CELL IS THE PROBABILITY MASS OF THE MATRIX WHOSE FIRST COLUMN IS THE ROW INDEX OF THE TABLE AND SECOND COLUMN IS THE COLUMN INDEX OF THE TABLE

	$\begin{bmatrix}0\\0\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}1\\1\end{bmatrix}$
$\begin{bmatrix}0\\0\end{bmatrix}$	0	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$
$\begin{bmatrix}1\\0\end{bmatrix}$	$\frac{1}{12}$	$\frac{1}{6}$	0	0
$\begin{bmatrix}0\\1\end{bmatrix}$	$\frac{1}{12}$	0	$\frac{1}{6}$	0
$\begin{bmatrix}1\\1\end{bmatrix}$	$\frac{1}{12}$	0	0	$\frac{1}{6}$

Proof: Fix a degraded LOC. Since (42) implies the condition given in (30), the subspace degradation is unique for a degraded LOC. Fix full-row-rank $r \times M$ matrices $\mathbf{D}$ and $\mathbf{D}'$, and an $r \times N$ matrix $\mathbf{E}$. By (42), for any full-column-rank matrix $\mathbf{B}$,

$$P_{Y|X}(\mathbf{BE}|\mathbf{BD}) = P_{Y|X}(\mathbf{BE}|\mathbf{BD}').$$

By Lemma 2,

$$\Pr\{\mathbf{D}H = \mathbf{E}\} = \Pr\{\mathbf{D}'H = \mathbf{E}\}. \tag{45}$$

We show that the LOC is row-space symmetric using the above equality.

Fix any input $\mathbf{X}$ and $\mathbf{X}'$ and output $\mathbf{Y}$ and $\mathbf{Y}'$ satisfying $\langle \mathbf{Y} \rangle \leq \langle \mathbf{X} \rangle$, $\langle \mathbf{Y}' \rangle \leq \langle \mathbf{X}' \rangle$, $\langle \mathbf{X}^\top \rangle = \langle \mathbf{X}'^\top \rangle$ and $\langle \mathbf{Y}^\top \rangle = \langle \mathbf{Y}'^\top \rangle$. Then we can write $\mathbf{X} = \mathbf{BD}$, $\mathbf{Y} = \mathbf{BE}$, $\mathbf{X}' = \mathbf{B}'\mathbf{D}$ and $\mathbf{Y}' = \mathbf{B}'\mathbf{E}'$ (see (50) and (51) in Appendix A). Since $\langle \mathbf{E}^\top \rangle = \langle \mathbf{E}'^\top \rangle$, there exists a full-rank square matrix Φ such that $\mathbf{E} = \Phi\mathbf{E}'$. Then we have

$$\begin{aligned}
P(\mathbf{Y}|\mathbf{X}) &= \Pr\{\mathbf{D}H = \mathbf{E}\} && (46)\\
&= \Pr\{\mathbf{D}H = \Phi\mathbf{E}'\}\\
&= \Pr\{\Phi^{-1}\mathbf{D}H = \mathbf{E}'\}\\
&= \Pr\{\mathbf{D}H = \mathbf{E}'\} && (47)\\
&= P(\mathbf{Y}'|\mathbf{X}'), && (48)
\end{aligned}$$

where (46) and (48) follow from Lemma 2, and (47) follows from (45) and $\mathrm{rk}(\Phi^{-1}\mathbf{D}) = \mathrm{rk}(\mathbf{D})$. The proof is completed by noting that (48) is sufficient for a LOC being row-space-symmetric. ■

The above theorem tells us that all the LOCs studied in [15]–[18] have a unique subspace degradation. Now we have a better understanding of why the capacity of these LOCs can be achieved by only optimizing the input rank distribution.

We say a LOC is *rank-symmetric* if its transition matrix is rank-symmetric (see Definition 5). We see that $\mathrm{LOC}(H, T)$ is rank-symmetric if and only if there exists a function $\mu : \mathbb{Z}^+ \times \mathbb{Z}^+ \to [0\ 1]$ such that

$$P_{Y|X}(\mathbf{Y}|\mathbf{X}) = \begin{cases} \mu(\mathrm{rk}(\mathbf{X}), \mathrm{rk}(\mathbf{Y})) & \langle \mathbf{Y} \rangle \leq \langle \mathbf{X} \rangle \\ 0 & \text{otherwise,} \end{cases}$$

where $\mathbb{Z}^+$ is the set of nonnegative integers.

By the definition, we see that a rank-symmetric LOC is also row-space-symmetric (see Definition 4 and Definition 5). The following theorem gives a stronger characterization of rank-symmetric LOCs.

Lemma 13: A rank-symmetric LOC is degraded.

Proof: We can check that (42) and (43) hold for a rank-symmetric LOC. By the definition of rank-symmetric LOCs, we know that $P_{Y|X}(\mathbf{Y}|\mathbf{X})$ only depends on U and V, which verifies (42). By the same property of rank-symmetric LOCs,

$$\begin{aligned}
p_Y(\mathbf{Y}) &= \sum_{\mathbf{X}' : V \leq \langle \mathbf{X}' \rangle} P_{Y|X}(\mathbf{Y}|\mathbf{X}) p_X(\mathbf{X})\\
&= \sum_{U' \in \mathrm{Pj}(\mathbb{F}^T) : V \leq U'} \mu(\dim(U'), \dim(V)) \sum_{\mathbf{X} : \langle \mathbf{X} \rangle = U'} p_X(\mathbf{X})\\
&= \sum_r \mu(r, \dim(V)) \sum_{U' \in \mathrm{Gr}(r, \mathbb{F}^T) : V \leq U'} p_{\langle X \rangle}(U').
\end{aligned}$$

This verifies (43). ■

But a degraded LOC may not be rank-symmetric.

Example 11: Consider $\mathrm{LOC}_2(H_2, 1)$, $H_2 \in \mathbb{F}^{2\times 2}$ as an example. For the distribution of H as given in Table II, we can calculate that

$$\begin{aligned}
P_{Y|X}(z_1|z_i) &= \frac{1}{6}, \quad i = 1, 2, 3\\
P_{Y|X}(z_2|z_i) &= \frac{1}{6}, \quad i = 1, 2, 3\\
P_{Y|X}(z_3|z_i) &= \frac{1}{3}, \quad i = 1, 2, 3\\
P_{Y|X}(z_0|z_i) &= \frac{1}{3}, \quad i = 1, 2, 3,
\end{aligned}$$

where

$$z_0 = [0\ 0], \quad z_1 = [1\ 0], \quad z_2 = [0\ 1], \text{ and } z_3 = [1\ 1]. \tag{49}$$

We can check by (44) that $\mathrm{LOC}_2(H, 1)$ with the distribution of H given in Table II is degraded. But this LOC is not rank symmetric.

The following theorem shows the relation between uniform-given-rank transfer matrices and rank-symmetric LOCs.

Theorem 9: Let H be a random matrix with dimension $M \times N$. i) If $T \geq M$ and $LOC(H, T)$ is rank-symmetric, then H is uniform-given-rank. ii) If H is uniform-given-rank, then $LOC(H, T)$ is rank-symmetric.

Proof: Proof of i). Fix $\mathbf{X} \in \mathbb{F}^{T\times M}$ with $\mathrm{rk}(\mathbf{X}) = M$. The existence of such $\mathbf{X}$ follows from $T \geq M$. For any $\mathbf{Y} \in \mathbb{F}^{T\times N}$, we have a unique $\mathbf{H}$ such that $\mathbf{Y} = \mathbf{X}\mathbf{H}$. Since the LOC is rank-symmetric,

$$\begin{aligned} p_H(\mathbf{H}) &= \Pr\{\mathbf{Y} = \mathbf{X}H\} \\ &= \mu(\mathrm{rk}(\mathbf{X}), \mathrm{rk}(\mathbf{Y})) \\ &= \mu(M, \mathrm{rk}(\mathbf{H})). \end{aligned}$$

Therefore H is uniform-given-rank.

Proof of ii). Fix $\mathbf{X} \in \mathbb{F}^{T\times M}$ and $\mathbf{Y} \in \mathbb{F}^{T\times N}$ with $\mathrm{rk}(\mathbf{X}) = r$, $\mathrm{rk}(\mathbf{Y}) = s$ and $\langle \mathbf{Y}\rangle \leq \langle \mathbf{X}\rangle$. By the similar procedure for obtaining (34), we have

$$P_{Y|X}(\mathbf{Y}|\mathbf{X}) = \Pr\left\{\mathbf{D}H = \begin{bmatrix}\mathbf{E}\\ \mathbf{0}\end{bmatrix}\right\},$$

for certain full row-rank matrices $\mathbf{D}$ and $\mathbf{E}$ are full row-rank matrices satisfying $\langle \mathbf{D}^\top\rangle = \langle \mathbf{X}^\top\rangle$ and $\langle \mathbf{E}^\top\rangle = \langle \mathbf{Y}^\top\rangle$. Fix any full-row-rank matrices $\mathbf{D}' \in \mathbb{F}^{r\times M}$ and $\mathbf{E}' \in \mathbb{F}^{s\times N}$. Find full rank matrices Φ and Ψ such that $\mathbf{D}' = \mathbf{D}\Phi$ and $\mathbf{E}' = \mathbf{E}\Psi$. We have

$$\begin{aligned} \Pr\left\{\mathbf{D}'H = \begin{bmatrix}\mathbf{E}'\\ \mathbf{0}\end{bmatrix}\right\} &= \Pr\left\{\mathbf{D}\Phi H\Psi^{-1} = \begin{bmatrix}\mathbf{E}\\ \mathbf{0}\end{bmatrix}\right\} \\ &= \Pr\left\{\mathbf{D}H = \begin{bmatrix}\mathbf{E}\\ \mathbf{0}\end{bmatrix}\right\}, \end{aligned}$$

where the last equality follows that H is uniform-given-rank. Hence

$$P_{Y|X}(\mathbf{Y}|\mathbf{X}) = \Pr\left\{\mathbf{D}'H = \begin{bmatrix}\mathbf{E}'\\ \mathbf{0}\end{bmatrix}\right\}.$$

So $P_{Y|X}(\mathbf{Y}|\mathbf{X})$ only relates to the ranks of $\mathbf{X}$ and $\mathbf{Y}$, i.e., $\mathrm{LOC}(H, T)$ is rank-symmetric. ■

There exists rank-symmetric LOCs with non-uniform-given-rank transfer matrices.

Example 12: We give an example of a rank-symmetric LOC that has a non-uniform transfer matrix. Consider $\mathrm{LOC}_2(H_2, 1)$, $H_2 \in \mathbb{F}^{2\times 2}$ with

$$p_{H_2}(\mathbf{H}) = \frac{1}{4}, \text{ for } \mathbf{H} = \begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix}, \begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}, \begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix}, \begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}.$$

We can check that H_2 is not uniform-given-rank, but we can verify that $\mathrm{LOC}_2(H, 1)$ is rank-symmetric. ◇

In the last of this section, we verify a claim given in Section III-C. Note that any transfer matrix H can be converted to a uniform-given-rank transfer matrix H^* with the same rank distribution [17], [18], [29] obtained by $H^* = \Phi H\Psi$, where Φ and Ψ are independent uniformly distributed random matrices in $\mathrm{Fr}(\mathbb{F}^{M\times M})$ and $\mathrm{Fr}(\mathbb{F}^{N\times N})$, respectively. Hence $C(H, T) \geq C(H^*, T) = C_{\mathrm{SS}}(H^*, T)$. We show that the lower bound on $C(H, T)$ given in Theorem 2 is at least as good as $C_{\mathrm{SS}}(H^*, T)$.

Let p_X be a uniform-given-rank input distribution that achieves $C_{\mathrm{SS}}(H^*, T)$, the existence of such a distribution is guaranteed by Theorem 5. Let P^* and P be the transition matrices corresponding to H^* and H respectively. For any input matrix $\mathbf{X}'$ with $\mathrm{rk}(\mathbf{X}') = r$, we have

$$\begin{aligned} P^*_{\mathrm{rk}(Y)|\mathrm{rk}(X)}(s|r) &= P^*_{\mathrm{rk}(Y)|X}(s|\mathbf{X}') \\ &= \Pr\{\mathrm{rk}(\mathbf{X}'\Phi H\Psi) = s\} \\ &= \Pr\{\mathrm{rk}(\mathbf{X}'\Phi H) = s\} \\ &= \sum_{\mathbf{X}:\mathrm{rk}(\mathbf{X})=r} \Pr\{\mathrm{rk}(\tilde{X}H) = s, \tilde{X} = \mathbf{X}\} \\ &= \sum_{\mathbf{X}:\mathrm{rk}(\mathbf{X})=r} P_{\mathrm{rk}(Y)|X}(s|\mathbf{X})p_{X|\mathrm{rk}(X)}(\mathbf{X}|r) \\ &= P_{\mathrm{rk}(Y)|\mathrm{rk}(X)}(s|r), \end{aligned}$$

where the first equality follows from (37); and $\tilde{X} = \mathbf{X}'\Phi$ is uniformly distributed among all input matrices with rank r, and has the same distribution of $p_{X|\mathrm{rk}(r)}(\mathbf{X}|r)$. Therefore for p_X,

$$\begin{aligned} I(X;Y)|_{p_H} &\geq J(\mathrm{rk}(X), \mathrm{rk}(Y))|_{P_{\mathrm{rk}(Y)|\mathrm{rk}(X)}} \\ &\quad + I(\langle X^\top\rangle; \langle Y^\top\rangle)|_{P_{\langle Y^\top\rangle|\langle X^\top\rangle}} \\ &\geq J(\mathrm{rk}(X), \mathrm{rk}(Y))|_{P_{\mathrm{rk}(Y)|\mathrm{rk}(X)}} \\ &\quad + I(\mathrm{rk}(X); \mathrm{rk}(Y))|_{P_{\mathrm{rk}(Y)|\mathrm{rk}(X)}} \\ &= J(\mathrm{rk}(X), \mathrm{rk}(Y))|_{P^*_{\mathrm{rk}(Y)|\mathrm{rk}(X)}} \\ &\quad + I(\mathrm{rk}(X); \mathrm{rk}(Y))|_{P^*_{\mathrm{rk}(Y)|\mathrm{rk}(X)}} \\ &= C_{\mathrm{SS}}(H^*, T), \end{aligned}$$

where the first inequality follows from Theorem 2 and the last equality follows from Theorem 5. Thus $C(H, T) \geq C_{\mathrm{SS}}(H^*, T)$.

VI. CONCLUDING REMARKS

In this paper, we studied upper and lower bounds for both the Shannon capacity and the subspace coding capacity of LOCs. We characterized various classes of LOCs with different properties of these bounds, where row spaces and ranks of input and output matrices play important roles.

Our results provide some guidelines for coding design. Subspace coding is good for LOCs with a unique subspace degradation since otherwise we have difficulty to find an optimal input distribution for subspace coding. For general LOCs, we can use constant-rank uniform-given-row-space input distribution for subspace codes since 1) such an optimal input distribution is relatively easy to compute, and 2) the loss of rate, compared with the subspace coding capacity, can be small for typical parameters. Further, it is not always optimal to uniform input and output of a LOC for applying subspace coding.

We are motivated to consider other coding schemes for LOCs since for many cases either the optimal subspace coding scheme is difficult to find or subspace coding is not capacity achieving. Readers are referred to [27] for a superposition based coding scheme that can achieve rate higher than subspace coding.

APPENDIX A
SYMMETRY PROPERTIES IN CHANNEL CAPACITY OPTIMIZATION

We discuss how the symmetry properties is used to solve the optimization problem for finding the channel capacity of LOCs. This is useful for getting some numerical results. We first introduce some notations that will be used in this section and Appendix B.

Let $\mathbf{B}$ be a $t \times r$ matrix with rank r, i.e., $\mathbf{B}$ is of full column-rank. For a $t \times m$ matrix $\mathbf{A}$ with $\langle \mathbf{A} \rangle \leq \langle \mathbf{B} \rangle$, define $\mathbf{A}/\mathbf{B}$ to be a matrix such that $\mathbf{A} = \mathbf{B}(\mathbf{A}/\mathbf{B})$. The notation "/" is well defined because i) there always exists $\mathbf{C}$ such that $\mathbf{A} = \mathbf{BC}$ since $\langle \mathbf{A} \rangle \leq \langle \mathbf{B} \rangle$ and ii) such $\mathbf{C}$ is unique since $\mathbf{B}$ is full column rank.

Let $\mathbf{X}$ and $\mathbf{Y}$ be the input and output matrices of LOC(H, T), respectively, with $\langle \mathbf{Y} \rangle \leq \langle \mathbf{X} \rangle$. A decomposition of $\mathbf{X}$ and $\mathbf{Y}$ as in Lemma 2 can be found as follows. First, fix a full column rank matrix $\mathbf{B}$ with $\langle \mathbf{X} \rangle = \langle \mathbf{B} \rangle$. Then, $\mathbf{X} = \mathbf{B}(\mathbf{X}/\mathbf{B})$ and $\mathbf{Y} = \mathbf{B}(\mathbf{Y}/\mathbf{B})$. By Lemma 2,

$$P_{Y|X}(\mathbf{Y}|\mathbf{X}) = \Pr\{(\mathbf{X}/\mathbf{B})H = \mathbf{Y}/\mathbf{B}\}.$$

For $U \in \mathrm{Pj}(\mathbb{F}^M)$, let $\mathbf{D}_U$ be a $\dim(U) \times M$ matrix with $\langle \mathbf{D}_U^\top \rangle = U$. For any $\mathbf{X} \in \mathbb{F}^{T \times M}$ and $\mathbf{Y} \in \mathbb{F}^{T \times N}$ with $\langle \mathbf{Y} \rangle \leq \langle \mathbf{X} \rangle$, we can write

$$\mathbf{X} = \mathbf{B}\mathbf{D}_{\langle \mathbf{X}^\top \rangle}, \tag{50}$$

$$\mathbf{Y} = \mathbf{B}\mathbf{E}, \tag{51}$$

where $\mathbf{B}^T = \mathbf{X}^\top / \mathbf{D}^\top_{\langle \mathbf{X}^\top \rangle}$ and $\mathbf{E} = \mathbf{Y}/\mathbf{B}$.

Due to the symmetry properties of the matrix of transition probabilities in Lemma 3, it is not necessary to calculate $P_{Y|X}(\mathbf{Y}|\mathbf{X})$ for all pairs of $\mathbf{X}$ and $\mathbf{Y}$. For each subspace $U \in \mathrm{Pj}(\min\{T, M\}, \mathbb{F}^M)$, we choose one full row rank matrix $\mathbf{D}$ with $\langle \mathbf{D}^\top \rangle = U$ to compute $(\Pr\{\mathbf{D}H = \mathbf{E}\} : \mathbf{E} \in \mathbb{F}^{k \times N})$. Then for any $\mathbf{X}$ and $\mathbf{Y}$ with $\langle \mathbf{Y} \rangle \leq \langle \mathbf{X} \rangle$ and $\langle \mathbf{X}^\top \rangle = U$, we know $P_{Y|X}(\mathbf{Y}|\mathbf{X}) = \Pr\{\mathbf{D}H = \mathbf{Y}/(\mathbf{X}^\top/\mathbf{D}^\top)^\top\}$. The overall complexity of computing the transition matrix is

$$\sum_{k=0}^{\min\{T,M\}} \begin{bmatrix} M \\ k \end{bmatrix} q^{kN} < \begin{cases} c q^{MN} & M \leq \min\{T, N\} \\ c' q^{L(M+N-L)} & \text{otherwise,} \end{cases}$$

where $L = \min\{T, (M+N)/2\}$, c and c' are constants. The inequality for $M \leq \min\{T, N\}$ is obtained as follows.[2] We have

$$\chi_k^N = q^{Nk} \prod_{i=N-k+1}^{N} (1 - q^{-i}) > q^{Nk} \prod_{i=1}^{\infty} (1 - q^{-i}) \geq \kappa q^{Nk}, \tag{52}$$

where $\kappa = \prod_{i=0}^{\infty}(1 - 2^{-i}) \approx 0.28879$ is a constant [22]. Thus

$$\sum_{k=0}^{M} \begin{bmatrix} M \\ k \end{bmatrix} q^{kN} < 1/\kappa \sum_{k=0}^{M} \begin{bmatrix} M \\ k \end{bmatrix} \chi_k^N = 1/\kappa \sum_{k=0} \chi_k^{M,N} = 1/\kappa q^{MN},$$

where the last equality is obtained by (5). When $M > \min\{T, N\}$, we have

$$\begin{aligned} \sum_{k=0}^{\min\{T,M\}} \begin{bmatrix} M \\ k \end{bmatrix} q^{kN} &= \sum_{k=0}^{\min\{T,M\}} \frac{\chi_k^M}{\chi_k^k} q^{kN} \\ &< \sum_{k=0}^{\min\{T,M\}} \frac{1/\kappa q^{Mk}}{q^{k^2}} q^{kN} \\ &= 1/\kappa \sum_{k=0}^{\min\{T,M\}} q^{k(M+N-k)}, \end{aligned} \tag{53}$$

where the inequality is obtained by (52) with M in place of N and $\chi_k^k < q^{k^2}$. Note that the $k(M+N-k)$ in (53) takes its maximum at $k = \min\{\min\{T, M\}, (M+N)/2\} = L$. Then by a technique similar to the one used in [12, Lemma 1], the inequality for $M > \min\{T, N\}$ is obtained, where the constant $c' = 1/\kappa \sum_{i=0}^{\infty} 2^{-i^2} \approx 5.4137$.

After obtaining the transition matrix, we can find an optimal input distribution by solving the maximization problem in Theorem 1, which is equivalent to finding an optimal distribution over $\mathrm{Pj}(\min\{T, M\}, \mathbb{F}^M)$. Since $|\mathrm{Pj}(\min\{M, T\}, \mathbb{F}^M)| = \sum_{k=0}^{\min\{M,T\}} \begin{bmatrix} M \\ k \end{bmatrix}$, we can bound the number of probability masses to determine as

$$\sum_{k=0}^{\min\{M,T\}} \begin{bmatrix} M \\ k \end{bmatrix} < \begin{cases} \Theta_1 q^{M^2/4} & T \geq M/2 \\ \Theta_2 q^{T(M-T)} & \text{otherwise,} \end{cases}$$

where Θ_1 and Θ_2 are constants. The inequality for $T \geq M/2$ is obtained by [12, Lemma 1], while the inequality for $T < M/2$, is obtained by [12, Proposition 1].

APPENDIX B
PROOF OF CLAIMS IN THE PROOF OF THEOREM 2

Proof of Claim 1: Define random variable $Y^{(r)} = Y$ for $r = 0$. For $r = 1, \ldots, \min\{T, M\}$, define random variables $Y^{(r)}$ and $Y^{(r,\Phi)}$ with $\Phi \in \mathrm{Fr}(\mathbb{F}^{r \times r})$ over $\mathbb{F}^{T \times N}$ as follows. For $\langle \mathbf{Y} \rangle \leq \langle \mathbf{X} \rangle$, let

$$P_{Y^{(r,\Phi)}|X}(\mathbf{Y}|\mathbf{X}) = \begin{cases} P_{Y^{(r-1)}|X}(\mathbf{Y}|\mathbf{X}) & \mathrm{rk}(\mathbf{X}) \neq r, \\ \Pr\{\mathbf{D}_{\langle \mathbf{X}^\top \rangle} H = \Phi \mathbf{E}\} & \mathrm{rk}(\mathbf{X}) = r, \end{cases}$$

where $\mathbf{E} = \mathbf{Y}/(\mathbf{X}^\top / \mathbf{D}^\top_{\langle \mathbf{X}^\top \rangle})^\top$ (see (50) and (51)). Random variables $Y^{(r)}$ are over $\mathbb{F}^{T \times N}$ such that for $\langle \mathbf{Y} \rangle \leq \langle \mathbf{X} \rangle$,

$$P_{Y^{(r)}|X}(\mathbf{Y}|\mathbf{X}) = \frac{1}{\chi_r^r} \sum_{\Phi \in \mathrm{Fr}(\mathbb{F}^{r \times r})} P_{Y^{(r,\Phi)}|X}(\mathbf{Y}|\mathbf{X}).$$

Note that when $\mathrm{rk}(\mathbf{X}) > r$,

$$P_{Y^{(r,\Phi)}|X}(\mathbf{Y}|\mathbf{X}) = P_{Y^{(r)}|X}(\mathbf{Y}|\mathbf{X}) = P_{Y|X}(\mathbf{Y}|\mathbf{X}). \tag{54}$$

We will show that for $r = 1, \ldots, \min\{M, T\}$,

$$P_{Y^*|X} = P_{Y^{(\min\{M,T\})}|X}, \tag{55}$$

and

$$I(X; Y^{(r,\Phi)}) = I(X; Y^{(r-1)}). \tag{56}$$

Since for a fixed p_X, mutual information $I(X; Y)$ is a convex function of the transition probabilities, we have

$$I(X; Y^{(r)}) \leq \frac{1}{\chi_r^r} \sum_{\Phi \in \mathrm{Fr}(\mathbb{F}^{r \times r})} I(X; Y^{(r,\Phi)}) = I(X; Y^{(r-1)}).$$

---

[2] This method is suggested by an anonymous reviewer, who is thereby acknowledged.

Then, the lemma is proved by

$$I(X;Y) = I(X;Y^{(0)}) \geq I(X;Y^{(\min\{M,T\})}) = I(X;Y^*).$$

We first prove (55). For $\mathbf{X}$ and $\mathbf{Y}$ with $\mathrm{rk}(\mathbf{X}) = r$, $\mathrm{rk}(\mathbf{Y}) = s$, $\langle \mathbf{Y}^\top \rangle = V$, $\langle \mathbf{X}^\top \rangle = U$ and $\langle \mathbf{Y} \rangle \leq \langle \mathbf{X} \rangle$, by definition,

$$P_{Y^{(\min\{M,T\})}|X}(\mathbf{Y}|\mathbf{X}) = \frac{1}{\chi_r^r} \sum_{\Phi \in \mathrm{Fr}(\mathbb{F}^{r\times r})} \Pr\{\mathbf{D}_U H = \Phi \mathbf{E}\},$$

where $\mathbf{D}_U$ and $\mathbf{E}$ are defined in (50) and (51), respectively. For $\mathbf{E}_0 \in \mathcal{E} \triangleq \{\mathbf{K} \in \mathbb{F}^{r\times N} : \langle \mathbf{K}^\top \rangle = V\}$, let $\mathcal{C}(\mathbf{E}_0) = \{\mathbf{C} \in \mathrm{Fr}(\mathbb{F}^{r\times r}) : \mathbf{C}\mathbf{E} = \mathbf{E}_0\}$. We see that $\{\mathcal{C}(\mathbf{E}_0), \mathbf{E}_0 \in \mathcal{E}\}$ gives a partition of $\mathrm{Fr}(\mathbb{F}^{r\times r})$. Since $\mathcal{C}(\mathbf{E}_0)$ for all $\mathbf{E}_0 \in \mathcal{E}$ have the same cardinality, $|\mathcal{C}(\mathbf{E}_0)| = \frac{|\mathrm{Fr}(\mathbb{F}^{r\times r})|}{|\mathcal{E}|} = \frac{\chi_r^r}{\chi_s^r}$ for all $\mathbf{E}_0 \in \mathcal{E}$. Therefore,

$$\begin{aligned}
&\frac{1}{\chi_r^r} \sum_{\Phi \in \mathrm{Fr}(\mathbb{F}^{r\times r})} \Pr\{\mathbf{D}_U H = \Phi \mathbf{E}\} \\
&= \frac{1}{\chi_r^r} \sum_{\mathbf{E}_0 \in \mathcal{E}} \sum_{\Phi \in \mathcal{C}(\mathbf{E}_0)} \Pr\{\mathbf{D}_U H = \Phi \mathbf{E}\} \\
&= \frac{1}{\chi_r^r} \sum_{\mathbf{E}_0 \in \mathcal{E}} |\mathcal{C}(\mathbf{E}_0)| \Pr\{\mathbf{D}_U H = \mathbf{E}_0\} \\
&= \frac{1}{\chi_s^r} \sum_{\mathbf{E}_0 \in \mathcal{E}} \Pr\{\mathbf{D}_U H = \mathbf{E}_0\} \\
&= \frac{1}{\chi_s^r} \Pr\{\langle (\mathbf{D}_U H)^\top \rangle = V\} \\
&= \frac{1}{\chi_s^r} P_{\langle Y^\top \rangle | \langle X^\top \rangle}(V|U).
\end{aligned}$$

By the definition of Y^*, (55) is proved.

Now we prove (56). First, we have for $i \neq r$,

$$\begin{aligned}
p_{\mathrm{rk}(X),Y^{(r,\Phi)}}(i,\mathbf{Y}) &= \sum_{\mathbf{X}:\mathrm{rk}(\mathbf{X})=i} P_{Y^{(r,\Phi)}|X}(\mathbf{Y}|\mathbf{X}) p_X(\mathbf{X}) \\
&= \sum_{\mathbf{X}:\mathrm{rk}(\mathbf{X})=i} P_{Y^{(r-1)}|X}(\mathbf{Y}|\mathbf{X}) p_X(\mathbf{X}) \\
&= p_{\mathrm{rk}(X),Y^{(r-1)}}(i,\mathbf{Y}), \qquad (57)
\end{aligned}$$

where the second equality is obtained by the definition of $P_{Y^{(r,\Phi)}|X}(\mathbf{Y}|\mathbf{X})$ for $\mathrm{rk}(\mathbf{X}) \neq r$. Specifically, when $r < i$,

$$p_{\mathrm{rk}(X),Y^{(r)}}(i,\mathbf{Y}) = p_{\mathrm{rk}(X),Y^{(r-1)}}(i,\mathbf{Y})$$

by the definition of $P_{Y^{(r)}|X}$ and (57). Recursively applying the above equality, we have that when $r < i$,

$$p_{\mathrm{rk}(X),Y^{(r)}}(i,\mathbf{Y}) = p_{\mathrm{rk}(X),Y}(i,\mathbf{Y}). \qquad (58)$$

We also have

$$\begin{aligned}
&p_{\mathrm{rk}(X),Y^{(r,\Phi)}}(r,\mathbf{Y}) \\
&= \sum_{\mathbf{X}:\mathrm{rk}(\mathbf{X})=r} P_{Y^{(r,\Phi)}|X}(\mathbf{Y}|\mathbf{X}) p_X(\mathbf{X}) \\
&= \sum_{U \in \mathrm{Gr}(r,\mathbb{F}^M)} \sum_{\mathbf{B} \in \mathrm{Fr}(\mathbb{F}^{T\times r})} P_{Y^{(r,\Phi)}|X}(\mathbf{Y}|\mathbf{B}\mathbf{D}_U) p_X(\mathbf{B}\mathbf{D}_U) \\
&= \sum_{U \in \mathrm{Gr}(r,\mathbb{F}^M)} \sum_{\mathbf{B} \in \mathrm{Fr}(\mathbb{F}^{T\times r})} \Pr\{\mathbf{D}_U H = \Phi(\mathbf{Y}/\mathbf{B})\} \frac{p_{\langle X^\top \rangle}(U)}{\chi_r^T} \\
&= \sum_{U \in \mathrm{Gr}(r,\mathbb{F}^M)} \frac{p_{\langle X^\top \rangle}(U)}{\chi_r^T} \sum_{\mathbf{B}' \in \mathrm{Fr}(\mathbb{F}^{T\times r})} \Pr\{\mathbf{D}_U H = \mathbf{Y}/\mathbf{B}'\} \\
&= p_{\mathrm{rk}(X),Y}(r,\mathbf{Y}) = p_{\mathrm{rk}(X),Y^{(r-1)}}(r,\mathbf{Y}), \qquad (59)
\end{aligned}$$

where the third equality follows that p_X is uniform-given-row-space and the definition of $P_{Y^{(r,\Phi)}|X}(\mathbf{Y}|\mathbf{X})$ for $\mathrm{rk}(\mathbf{X}) = r$; the forth equality follows by $\Phi(\mathbf{Y}/\mathbf{B}) = \mathbf{Y}/(\mathbf{B}\Phi^{-1})$ and substituting $\mathbf{B}\Phi^{-1}$ by $\mathbf{B}' \in \mathrm{Fr}(\mathbb{F}^{T\times r})$; and (59) follow from (58).

Thus, by (57) and (59), we have

$$\begin{aligned}
p_{Y^{(r,\Phi)}}(\mathbf{Y}) &= \sum_i p_{\mathrm{rk}(X),Y^{(r,\Phi)}}(i,\mathbf{Y}) \\
&= \sum_i p_{\mathrm{rk}(X),Y^{(r-1)}}(i,\mathbf{Y}) \\
&= p_{Y^{(r-1)}}(\mathbf{Y}),
\end{aligned}$$

and hence

$$\mathcal{H}(Y^{(r,\Phi)}) = \mathcal{H}(Y^{(r-1)}). \qquad (60)$$

Further, for $\mathbf{X}$ with $\mathrm{rk}(\mathbf{X}) \neq r$, since $P_{Y^{(r,\Phi)}|X}(\mathbf{Y}|\mathbf{X}) = P_{Y^{(r-1)}|X}(\mathbf{Y}|\mathbf{X})$, we have

$$\mathcal{H}(Y^{(r,\Phi)}|X = \mathbf{X}) = \mathcal{H}(Y^{(r-1)}|X = \mathbf{X}). \qquad (61)$$

On the other hand, for $\mathbf{X}$ with $\mathrm{rk}(\mathbf{X}) = r$, by substituting $\mathbf{X} = \mathbf{B}\mathbf{D}_{\langle \mathbf{X}^\top \rangle}$, we have

$$\begin{aligned}
&\mathcal{H}(Y^{(r,\Phi)}|X = \mathbf{X}) \\
&= \sum_{\mathbf{Y}:\langle \mathbf{Y} \rangle \leq \langle \mathbf{B} \rangle} \Pr\{\mathbf{D}_{\langle \mathbf{X}^\top \rangle} H = \Phi(\mathbf{Y}/\mathbf{B})\} \\
&\qquad \times \log \frac{1}{\Pr\{\mathbf{D}_{\langle \mathbf{X}^\top \rangle} H = \Phi(\mathbf{Y}/\mathbf{B})\}} \qquad (62) \\
&= \sum_{\mathbf{Y}:\langle \mathbf{Y} \rangle \leq \langle \mathbf{B} \rangle} \Pr\{\mathbf{D}_{\langle \mathbf{X}^\top \rangle} H = \mathbf{Y}/(\mathbf{B}\Phi^{-1})\} \\
&\qquad \times \log \frac{1}{\Pr\{\mathbf{D}_{\langle \mathbf{X}^\top \rangle} H = \mathbf{Y}/(\mathbf{B}\Phi^{-1})\}} \qquad (63) \\
&= \sum_{\mathbf{Y}:\langle \mathbf{Y} \rangle \leq \langle \mathbf{B} \rangle} P_{Y|X}(Y|\mathbf{B}\Phi^{-1}\mathbf{D}_{\langle \mathbf{X}^\top \rangle}) \\
&\qquad \times \log \frac{1}{P_{Y|X}(Y|\mathbf{B}\Phi^{-1}\mathbf{D}_{\langle \mathbf{X}^\top \rangle})} \qquad (64) \\
&= \mathcal{H}(Y|X = \mathbf{B}\Phi^{-1}\mathbf{D}_{\langle \mathbf{X}^\top \rangle}) \qquad (65)
\end{aligned}$$

where (62) follows from $\langle \mathbf{X} \rangle = \langle \mathbf{B} \rangle$ and the definition of $P_{Y^{(r,\Phi)}|X}(\mathbf{Y}|\mathbf{X})$ for $\mathrm{rk}(\mathbf{X}) = r$; (63) follows by $\Phi(\mathbf{Y}/\mathbf{B}) = \mathbf{Y}/(\mathbf{B}\Phi^{-1})$; (64) follows from Lemma 2; and (65) is obtained by $\langle \mathbf{B}\Phi^{-1}\mathbf{D}_{\langle \mathbf{X}^\top \rangle} \rangle = \langle \mathbf{B} \rangle$.

Hence

$$\begin{aligned}
&\sum_{\mathbf{X}:\mathrm{rk}(\mathbf{X})=r} \mathcal{H}(Y^{(r,\Phi)}|X = \mathbf{X}) p_X(\mathbf{X}) \\
&= \sum_{U \in \mathrm{Gr}(r,\mathbb{F}^M)} \sum_{\mathbf{B} \in \mathrm{Fr}(\mathbb{F}^{T\times r})} \mathcal{H}(Y^{(r,\Phi)}|X = \mathbf{B}\mathbf{D}_U) p_X(\mathbf{B}\mathbf{D}_U) \\
&= \sum_{U \in \mathrm{Gr}(r,\mathbb{F}^M)} \sum_{\mathbf{B} \in \mathrm{Fr}(\mathbb{F}^{T\times r})} \mathcal{H}(Y|X = \mathbf{B}\Phi^{-1}\mathbf{D}_U) p_X(\mathbf{B}\mathbf{D}_U) \\
&= \sum_{U \in \mathrm{Gr}(r,\mathbb{F}^M)} \sum_{\mathbf{B}' \in \mathrm{Fr}(\mathbb{F}^{T\times r})} \mathcal{H}(Y|X = \mathbf{B}'\mathbf{D}_U) p_X(\mathbf{B}'\mathbf{D}_U)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{\mathbf{X}:\mathrm{rk}(\mathbf{X})=r} \mathcal{H}(Y|X=\mathbf{X})p_X(\mathbf{X}) \\
&= \sum_{\mathbf{X}:\mathrm{rk}(\mathbf{X})=r} \mathcal{H}(Y^{(r-1)}|X=\mathbf{X})p_X(\mathbf{X}), \qquad (66)
\end{aligned}
$$

where the second equality follows from (65); the third equality follows by substituting $\mathbf{B}\Phi^{-1}$ by $\mathbf{B}' \in \mathrm{Fr}(\mathbb{F}^{T\times r})$ and the fact that p_X is uniform-given-row-space; and (66) follows from (54).

Therefore,

$$
\begin{aligned}
\mathcal{H}(Y^{(r,\Phi)}|X) &= \sum_{\mathbf{X}:\mathrm{rk}(\mathbf{X})\neq r} \mathcal{H}(Y^{(r,\Phi)}|X=\mathbf{X})p_X(\mathbf{X}) \\
&\quad + \sum_{\mathbf{X}:\mathrm{rk}(\mathbf{X})=r} \mathcal{H}(Y^{(r,\Phi)}|X=\mathbf{X})p_X(\mathbf{X}) \\
&= \sum_{\mathbf{X}:\mathrm{rk}(\mathbf{X})\neq r} \mathcal{H}(Y^{(r-1)}|X=\mathbf{X})p_X(\mathbf{X}) \\
&\quad + \sum_{\mathbf{X}:\mathrm{rk}(\mathbf{X})=r} \mathcal{H}(Y^{(r-1)}|X=\mathbf{X})p_X(\mathbf{X}) \qquad (67) \\
&= \mathcal{H}(Y^{(r-1)}|X), \qquad (68)
\end{aligned}
$$

where (67) follows from (61) and (66). Lastly, the equality in (56) is proved by (60) and (68). ■

Proof of Claim 2: By the definition of $P_{Y^*|X}$,

$$
\begin{aligned}
&\mathcal{H}(Y^*|X) \\
&= \sum_{\mathbf{X}} p_X(\mathbf{X}) \sum_{\mathbf{Y}:\langle\mathbf{Y}\rangle\le\langle\mathbf{X}\rangle} P_{Y^*|X}(\mathbf{Y}|\mathbf{X}) \log \frac{1}{P_{Y^*|X}(\mathbf{Y}|\mathbf{X})} \\
&= \sum_r \sum_{U\in\mathrm{Gr}(r,\mathbb{F}^M)} \sum_{\mathbf{X}:\langle\mathbf{X}^\top\rangle=U} \frac{p_{\langle X^\top\rangle}(U)}{\chi_r^T} \sum_s \sum_{V\in\mathrm{Gr}(s,\mathbb{F}^N)} \\
&\quad \times \sum_{\mathbf{Y}:\langle\mathbf{Y}^\top\rangle=V,\langle\mathbf{Y}\rangle\le\langle\mathbf{X}\rangle} \\
&\quad \times \frac{P_{\langle Y^\top\rangle|\langle X^\top\rangle}(V|U)}{\chi_s^r} \log \frac{\chi_s^r}{P_{\langle Y^\top\rangle|\langle X^\top\rangle}(V|U)} \\
&= \sum_r \sum_{U\in\mathrm{Gr}(r,\mathbb{F}^M)} p_{\langle X^\top\rangle}(U) \sum_s \sum_{V\in\mathrm{Gr}(s,\mathbb{F}^N)} \\
&\quad \times P_{\langle Y^\top\rangle|\langle X^\top\rangle}(V|U) \log \frac{\chi_s^r}{P_{\langle Y^\top\rangle|\langle X^\top\rangle}(V|U)} \\
&= \sum_{s\le r} p_{\mathrm{rk}(X)\,\mathrm{rk}(Y)}(r,s) \log \chi_s^r + \mathcal{H}(\langle Y^\top\rangle|\langle X^\top\rangle),
\end{aligned}
$$

which proves the first equality in the claim.

For $\mathbf{Y}$ with $\langle\mathbf{Y}^\top\rangle = V$ and $\mathrm{rk}(\mathbf{Y}^\top) = s$, we have

$$
\begin{aligned}
p_{Y^*}(\mathbf{Y}) &= \sum_{\mathbf{X}:\langle\mathbf{Y}\rangle\le\langle\mathbf{X}\rangle} P_{Y^*|X}(\mathbf{Y}|\mathbf{X})p_X(\mathbf{X}) \\
&= \sum_r \sum_{U\in\mathrm{Gr}(r,\mathbb{F}^M)} \sum_{\mathbf{X}:\langle\mathbf{X}^\top\rangle=U,\langle\mathbf{Y}\rangle\le\langle\mathbf{X}\rangle} \frac{1}{\chi_s^r} \\
&\quad \times P_{\langle Y^\top\rangle|\langle X^\top\rangle}(V|U) \frac{1}{\chi_r^T} p_{\langle X^\top\rangle}(U) \\
&= \sum_r \sum_{U\in\mathrm{Gr}(r,\mathbb{F}^M)} \frac{1}{\chi_s^T} P_{\langle Y^\top\rangle|\langle X^\top\rangle}(V|U) p_{\langle X^\top\rangle}(U) \\
&= \frac{1}{\chi_s^T} p_{\langle Y^\top\rangle}(V)
\end{aligned}
$$

where the third equality follows from

$$
\begin{aligned}
&|\{\mathbf{X} : \langle\mathbf{X}^\top\rangle = U, \langle\mathbf{Y}\rangle \le \langle\mathbf{X}\rangle\}| \\
&\quad = \sum_{\tilde{U}\in\mathrm{Gr}(r,\mathbb{F}^T):\langle\mathbf{Y}\rangle\le\tilde{U}} |\{\mathbf{X} : \langle\mathbf{X}^\top\rangle = U, \langle\mathbf{X}\rangle = \tilde{U}\}| \\
&\quad = \begin{bmatrix} T \\ r \end{bmatrix} \frac{\chi_s^r}{\chi_s^T} \chi_r^r.
\end{aligned}
$$

Here, $\{\tilde{U} \in \mathrm{Gr}(r,\mathbb{F}^T) : \langle\mathbf{Y}\rangle \le \tilde{U}\}$ is calculated in Lemma 1. Hence,

$$
\begin{aligned}
\mathcal{H}(Y^*) &= \sum_{\mathbf{Y}} p_{Y^*}(\mathbf{Y}) \log \frac{1}{p_{Y^*}(\mathbf{Y})} \\
&= \sum_s \sum_{V\in\mathrm{Gr}(s,\mathbb{F}^N)} \sum_{\mathbf{Y}:\langle\mathbf{Y}^\top\rangle=V} \frac{p_{\langle Y^\top\rangle}(V)}{\chi_s^T} \log \frac{\chi_s^T}{p_{\langle Y^\top\rangle}(V)} \\
&= \sum_s \sum_{V\in\mathrm{Gr}(s,\mathbb{F}^N)} p_{\langle Y^\top\rangle}(V) \log \frac{\chi_s^T}{p_{\langle Y^\top\rangle}(V)} \\
&= \sum_s p_{\mathrm{rk}(Y)}(s) \log \chi_s^T + \mathcal{H}(\langle Y^\top\rangle).
\end{aligned}
$$

■

APPENDIX C

A TECHNICAL LEMMA

The following lemma gives a lower bound on the difference $R(\mathbb{F}^M) - R(V)$ for $V \in \mathrm{Pj}(\mathbb{F}^M)$. The intuition behind the bound is that if the input rank is larger, the output rank also tends to be larger.

Lemma 14: Consider LOC(H,T) with $T \ge M$. Fix a uniform-given-row-space input. For $V \in \mathrm{Pj}(\mathbb{F}^M)$ with $\dim(V) = r < \mathrm{rk}^*(H)$,

$$R(\mathbb{F}^M) - R(V) > \Theta(T,r,H)\log q,$$

where

$$
\begin{aligned}
\Theta(T,r,H) \triangleq (T-M) &\sum_{k:k>r} \Pr\{\mathrm{rk}(H)\ge k\} \\
&-r(M-r) + \log_q \zeta_r^r.
\end{aligned}
$$

Proof: Let $\tilde{U} = \mathbb{F}^M$. Since $V \le \tilde{U}$, there exists a full rank $M \times M$ matrix

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}_0 \\ \mathbf{D}_1 \end{bmatrix}$$

such that $\langle\mathbf{D}^\top\rangle = \tilde{U}$ and $\langle\mathbf{D}_1^\top\rangle = V$. By Lemma 4,

$$\sum_{s\ge k} P_{\mathrm{rk}(Y)|\langle X^\top\rangle}(s|V) = \Pr\{\mathrm{rk}(\mathbf{D_1}H) \ge k\},$$

and

$$
\begin{aligned}
P_{\mathrm{rk}(Y)|\langle X^\top\rangle}(s|\tilde{U}) &= \Pr\{\mathrm{rk}(\mathbf{D}H) = s\} \\
&= \Pr\{\mathrm{rk}(H) = s\}. \qquad (69)
\end{aligned}
$$

We know $\Pr\{\mathrm{rk}(H) \ge s\} \ge \Pr\{\mathrm{rk}(\mathbf{D_1}H) \ge s\}$. So

$$\sum_{s\ge k} P_{\mathrm{rk}(Y)|\langle X^\top\rangle}(s|\tilde{U}) \ge \sum_{s\ge k} P_{\mathrm{rk}(Y)|\langle X^\top\rangle}(s|V). \tag{70}$$

Moreover, for k such that $r < k \le \mathrm{rk}^*(H)$,

$$\sum_{s:s\ge k} P_{\mathrm{rk}(Y)|\langle X^\top\rangle}(s|V) = 0. \tag{71}$$

Thus,

$$\begin{aligned}
&\sum_s s(P_{\mathrm{rk}(Y)|\langle X^\top\rangle}(s|\tilde{U}) - P_{\mathrm{rk}(Y)|\langle X^\top\rangle}(s|V)) \\
&\quad = \sum_k \sum_{s:s\geq k} (P_{\mathrm{rk}(Y)|\langle X^\top\rangle}(s|\tilde{U}) - P_{\mathrm{rk}(Y)|\langle X^\top\rangle}(s|V)) \\
&\quad \geq \sum_{k:\mathrm{rk}^*(H)\geq k>r} \sum_{s:s\geq k} P_{\mathrm{rk}(Y)|\langle X^\top\rangle}(s|\tilde{U}) && (72) \\
&\quad \geq \sum_{k:\mathrm{rk}^*(H)\geq k>r} \Pr\{\mathrm{rk}(H)\geq k\} && (73) \\
&\quad \triangleq \mathrm{E}[H,r], && (74)
\end{aligned}$$

where (72) is obtained by (70) and (71); (73) follows from (69).

By the definition of $R(U)$ in (26),

$$\begin{aligned}
&\frac{R(\tilde{U}) - R(V)}{\log q} \\
&\quad = \sum_s P_{\mathrm{rk}(Y)|\langle X^\top\rangle}(s|\tilde{U}) \left((T-M)s + \log_q \frac{\zeta_s^T}{\zeta_s^M}\right) \\
&\qquad - \sum_s P_{\mathrm{rk}(Y)|\langle X^\top\rangle}(s|\tilde{V}) \left((T-r)s + \log_q \frac{\zeta_s^T}{\zeta_s^r}\right) \\
&\quad = (T-M)\sum_s s(P_{\mathrm{rk}(Y)|\langle X^\top\rangle}(s|\tilde{U}) - P_{\mathrm{rk}(Y)|\langle X^\top\rangle}(s|V)) \\
&\qquad -(M-r)\sum_s s P_{\mathrm{rk}(Y)|\langle X^\top\rangle}(s|V) \\
&\qquad + \sum_s P_{\mathrm{rk}(Y)|\langle X^\top\rangle}(s|\tilde{U}) \log_q \frac{\zeta_s^T}{\zeta_s^M} \\
&\qquad - \sum_s P_{\mathrm{rk}(Y)|\langle X^\top\rangle}(s|V) \log_q \frac{\zeta_s^T}{\zeta_s^r} \\
&\quad > (T-M)\,\mathrm{E}[H,r] - r(M-r) + \log_q \zeta_r^r,
\end{aligned}$$

where the last inequality follows from (74),

$$(M-r)\sum_s s P_{\mathrm{rk}(Y)|\langle X^\top\rangle}(s|V) \leq r(M-r),$$

$$\sum_s P_{\mathrm{rk}(Y)|\langle X^\top\rangle}(s|\tilde{U}) \log_q \frac{\zeta_s^T}{\zeta_s^M} \geq 0,$$

and

$$\begin{aligned}
\sum_s P_{\mathrm{rk}(Y)|\langle X^\top\rangle}(s|V) \log_q \frac{\zeta_s^T}{\zeta_s^r} &< \sum_s P_{\mathrm{rk}(Y)|\langle X^\top\rangle}(s|V) \log_q \frac{1}{\zeta_s^r} \\
&\leq \log_q \frac{1}{\zeta_r^r}. \qquad \blacksquare
\end{aligned}$$

APPENDIX D
PROOF OF LEMMA 12

We will show that for a row-space-symmetric LOC,

$$I(\langle X\rangle; \langle Y\rangle) \leq J(\mathrm{rk}(X); \mathrm{rk}(Y)) + I(\langle X\rangle; \mathrm{rk}(Y)) \qquad (75)$$

where equality holds if and only if $p_{\langle Y\rangle}(U) = p_{\langle Y\rangle}(U')$ for all U, U' with $\dim(U) = \dim(U')$. For convenience, we call an input distribution β-type if for any $U \in \mathbb{F}^T$, there exists $\mathbf{X}_U$ with $\langle \mathbf{X}_U\rangle = U$ such that $p_X(\mathbf{X}_U) = p_{\langle X\rangle}(U)$. By Lemma 7, there must exist a β-type input distribution achieving C_{SS}. When the input distribution is β-type,

$$\begin{aligned}
I(\langle X\rangle; \mathrm{rk}(Y)) &= I(X; \mathrm{rk}(Y)) \\
&= I(X\langle X^\top\rangle; \mathrm{rk}(Y)) \\
&= I(\langle X^\top\rangle; \mathrm{rk}(Y)), \qquad (76)
\end{aligned}$$

where the first equality is due to the fact that X is β-type, and the last equality follows from the Markov chain $X \to \langle X^\top\rangle \to \mathrm{rk}(Y)$ implied by Lemma 4. Then, for a row-space-symmetric LOC,

$$\begin{aligned}
C_{\mathrm{SS}} &= \max_{p_X:\beta\text{-type}} I(\langle X\rangle; \langle Y\rangle) && (77) \\
&\leq \max_{p_X:\beta\text{-type}} [J(\mathrm{rk}(X); \mathrm{rk}(Y)) + I(\langle X\rangle; \mathrm{rk}(Y))] && (78) \\
&= \max_{p_X:\beta\text{-type}} \left[J(\mathrm{rk}(X); \mathrm{rk}(Y)) + I(\langle X^\top\rangle; \mathrm{rk}(Y))\right] && (79) \\
&\leq \max_{p_{\langle X^\top\rangle}} \left[J(\mathrm{rk}(X); \mathrm{rk}(Y)) + I(\langle X^\top\rangle; \mathrm{rk}(Y))\right], && (80)
\end{aligned}$$

where (77) follows Lemma 7, (78) is obtained by applying (75) for row-space-symmetric LOCs, (79) follows from (76), and (80) follows that $J(\mathrm{rk}(X); \mathrm{rk}(Y))$ and $I(\langle X^\top\rangle; \mathrm{rk}(Y))$ are related to p_X only through $p_{\langle X^\top\rangle}$.

To prove (75), fix a row-space-symmetric LOC. Let $\mathbf{X}$ be an input matrix with rank r. Consider two subspaces V' and V of $\langle \mathbf{X}\rangle$ with dimension s. There exists a full rank matrix Φ such that $\Phi V = V'$. Then, by the property of row-space-symmetric LOCs,

$$\begin{aligned}
P_{\langle Y\rangle|X}(V'|\mathbf{X}) &= \sum_{\mathbf{Y}:\langle \mathbf{Y}\rangle = V'} P_{Y|X}(\mathbf{Y}|\mathbf{X}) \\
&= \sum_{\mathbf{Y}:\langle \mathbf{Y}\rangle = V} P_{Y|X}(\Phi\mathbf{Y}|\mathbf{X}) \\
&= \sum_{\mathbf{Y}:\langle \mathbf{Y}\rangle = V} \frac{1}{\chi_s^r} P_{\langle Y^\top\rangle|\langle X^\top\rangle}(\langle \mathbf{Y}^\top\Phi^\top\rangle|\langle \mathbf{X}^\top\rangle) \\
&= \sum_{\mathbf{Y}:\langle \mathbf{Y}\rangle = V} \frac{1}{\chi_s^r} P_{\langle Y^\top\rangle|\langle X^\top\rangle}(\langle \mathbf{Y}^\top\rangle|\langle \mathbf{X}^\top\rangle) \\
&= P_{\langle Y\rangle|X}(V|\mathbf{X}).
\end{aligned}$$

In other words, for all the subspaces V of $\langle \mathbf{X}\rangle$ with the same dimension, $P_{\langle Y\rangle|X}(V|\mathbf{X})$ are the same. Since by Lemma 4,

$$\begin{aligned}
P_{\mathrm{rk}(Y)|\langle X^\top\rangle}(s|\langle \mathbf{X}^\top\rangle) &= P_{\mathrm{rk}(Y)|X}(s|\mathbf{X}) \\
&= \sum_{V\in\mathrm{Gr}(s,\langle \mathbf{X}\rangle)} P_{\langle Y\rangle|X}(V|\mathbf{X}),
\end{aligned}$$

we have for any $V \in \mathrm{Gr}(s, \langle \mathbf{X}\rangle)$,

$$P_{\langle Y\rangle|X}(V|\mathbf{X}) = \frac{1}{\begin{bmatrix} r \\ s \end{bmatrix}} P_{\mathrm{rk}(Y)|\langle X^\top\rangle}(s|\langle \mathbf{X}^\top\rangle).$$

Then we have for $V \le U$ with $\dim(U) = r$ and $\dim(V) = s$,

$$\begin{aligned}
&P_{\langle Y\rangle|\langle X\rangle}(V|U)\\
&= \sum_{\mathbf{X}:\langle\mathbf{X}\rangle = U} P_{\langle Y\rangle|X}(V|\mathbf{X}) P_{X|\langle X\rangle}(\mathbf{X}|U)\\
&= \sum_{\tilde{U}\in \mathrm{Gr}(r,\mathbb{F}^M)} \sum_{\mathbf{X}:\langle\mathbf{X}\rangle = U, \langle\mathbf{X}^\top\rangle = \tilde{U}} \frac{1}{\begin{bmatrix} r\\ s\end{bmatrix}}\\
&\quad \times P_{\mathrm{rk}(Y)|\langle X^\top\rangle}(s|\langle\mathbf{X}^\top\rangle) P_{X|\langle X\rangle}(\mathbf{X}|U)\\
&= \sum_{\tilde{U}\in \mathrm{Gr}(r,\mathbb{F}^M)} \frac{1}{\begin{bmatrix} r\\ s\end{bmatrix}} P_{\mathrm{rk}(Y)|\langle X^\top\rangle}(s|\tilde{U}) P_{\langle X^\top\rangle|\langle X\rangle}(\tilde{U}|U)\\
&= \frac{1}{\begin{bmatrix} r\\ s\end{bmatrix}} P_{\mathrm{rk}(Y)|\langle X\rangle}(s|U). \qquad (81)
\end{aligned}$$

Substituting (81) into the conditional entropy $\mathcal{H}(\langle Y\rangle|\langle X\rangle)$, we obtain

$$\mathcal{H}(\langle Y\rangle|\langle X\rangle) = \sum_{r,s} p_{\mathrm{rk}(X)\,\mathrm{rk}(Y)}(r,s) \log \begin{bmatrix} r\\ s\end{bmatrix} + \mathcal{H}(\mathrm{rk}(Y)|\langle X\rangle). \qquad (82)$$

Further, we have

$$\begin{aligned}
\mathcal{H}(\langle Y\rangle) &= \mathcal{H}(\langle Y\rangle\, \mathrm{rk}(Y))\\
&= \mathcal{H}(\mathrm{rk}(Y)) + \mathcal{H}(\langle Y\rangle|\, \mathrm{rk}(Y))\\
&= \mathcal{H}(\mathrm{rk}(Y)) + \sum_s p_{\mathrm{rk}(Y)}(s)\, \mathcal{H}(\langle Y\rangle|\, \mathrm{rk}(Y) = s)\\
&\le \mathcal{H}(\mathrm{rk}(Y)) + \sum_s p_{\mathrm{rk}(Y)}(s) \log \begin{bmatrix} T\\ s\end{bmatrix} \qquad (83)
\end{aligned}$$

with equality if and only if

$$p_{\langle Y\rangle}(V) = p_{\mathrm{rk}(Y)}(\dim(V)) / \begin{bmatrix} T\\ \dim(V)\end{bmatrix}$$

for all V. Therefore, (75) is proved by (82) and (83).

ACKNOWLEDGEMENT

The authors would like to thank Kenneth W. Shum and Raymond W. Yeung for helpful discussions.

REFERENCES

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

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

[3] P. A. Chou, Y. Wu, and K. Jain, "Practical network coding," in *Proc. Annu. Allerton Conf. Commun. Control Comput.*, Oct. 2003, pp. 1–7.

[4] P. Maymounkov, N. J. A. Harvey, and D. S. Lun, "Methods for efficient network coding," in *Proc. 44th Allerton Conf. Commun., Control, Comput.*, Sep. 2006, pp. 1–10.

[5] T. Ho *et al.*, "A random linear network coding approach to multicast," *IEEE Trans. Inf. Theory*, vol. 52, no. 10, pp. 4413–4430, Oct. 2006.

[6] S. Yang, J. Meng, and E.-H. Yang, "Coding for linear operator channels over finite fields," in *Proc. IEEE ISIT*, Jun. 2010, pp. 2413–2417.

[7] S. Yang and R. W. Yeung, "Coding for a network coded fountain," in *Proc. IEEE ISIT*, Aug. 2011, pp. 2647–2651.

[8] S. Yang and R. W. Yeung, "Batched sparse codes," submitted for publication.

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

[10] D. Silva, F. Kschischang, and R. Koetter, "A rank-metric approach to error control in random network coding," *IEEE Trans. Inf. Theory*, vol. 54, no. 9, pp. 3951–3967, Sep. 2008.

[11] D. Silva and F. Kschischang, "On metrics for error correction in network coding," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5479–5490, Dec. 2009.

[12] M. Gadouleau and Z. Yan, "Packing and covering properties of subspace codes for error control in random linear network coding," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2097–2108, May 2010.

[13] R. W. Nóbrega and B. F. Uchôa-Filho, "Multishot codes for network coding: Bounds and a multilevel construction," in *Proc. IEEE ISIT*, Jul. 2009, pp. 1–5.

[14] R. W. Nóbrega and B. F. Uchôa-Filho, "Multishot codes for network coding using rank-metric codes," in *Proc. IEEE WiNC*, Jun. 2010, pp. 1–5.

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

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

[17] R. W. Nóbrega, B. F. Uchôa-Filho, and D. Silva, "On the capacity of multiplicative finite-field matrix channels," in *Proc. IEEE ISIT*, Aug. 2011, pp. 1–11.

[18] R. W. Nóbrega, D. Silva, and B. F. Uchôa-Filho, "On the capacity of multiplicative finite-field matrix channels," in *Proc. IEEE ISIT*, Aug. 2011, pp. 341–345.

[19] D. S. Lun, M. Médard, R. Koetter, and M. Effros, "On coding for reliable communication over packet networks," *Phys. Commun.*, vol. 1, no. 1, pp. 3–20, 2008.

[20] G. E. Andrews, *The Theory of Partitions* (Encyclopedia of Mathematics and its Applications), vol. 2. Reading, MA, USA: Addison-Wesley, 1976.

[21] E. M. Gabidulin, "Theory of codes with maximum rank distance," *Problems Inform. Trans.*, vol. 21, no. 1, pp. 1–12, 1985.

[22] C. Cooper, "On the distribution of rank of a random matrix over a finite field," *Random Struct. Algorithms*, vol. 17, nos. 3–4, pp. 197–212, 2000.

[23] M. Gadouleau and Z. Yan, "Packing and covering properties of rank metric codes," *IEEE Trans. Inf. Theory*, vol. 54, no. 9, pp. 3873–3883, Sep. 2008.

[24] S. Yang, S.-W. Ho, J. Meng, and E.-H. Yang, "Linear operator channels over finite fields," Arxiv:1002.2293v1, 2010.

[25] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: Wiley, 1968.

[26] R. W. Yeung, *Information Theory and Network Coding*. New York, NY, USA: Springer-Verlag, 2008.

[27] S. Yang, "Superposition coding for linear operator channels over finite fields," in *Proc. IEEE ITW*, Lausanne, Switzerland, Sep. 2012, pp. 1–9.

[28] D. P. Bertsekas, A. Nedic, and A. E. Ozdaglar, *Convex Analysis and Optimization*. Belmont, MA, USA: Athena Scientific, 2003.

[29] M. J. Siavoshani, S. Yang, and R. W. Yeung, "Non-coherent network coding: An arbitrarily varying channel approach," in *Proc. IEEE ISIT*, Cambridge, MA, USA, Jul. 2012, pp. 1–7.

Shenghao Yang (S'07–M'11) received the B.S. degree in Electronics Engineering from Nankai University in 2001, M.Eng. degree in Electronics Engineering from Peking University in 2004, and Ph.D. degree in Information Engineering from The Chinese University of Hong Kong in 2008.

He was a Postdoctoral Fellow in Department of Electrical and Computer Engineering, University of Waterloo from 2008 to 2009, and a Postdoctoral Fellow/Research Associate in Institute of Network Coding, The Chinese University of Hong Kong from 2010 to 2012. He is currently an Assistant Professor with Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing, China.

Siu-Wai Ho (S'05–M'07) received the B.Eng., M.Phil., and Ph.D. degrees in information engineering from The Chinese University of Hong Kong in 2000, 2003, and 2006, respectively.

During 2006–2008, he was a Postdoctoral Research Fellow in the Department of Electrical Engineering, Princeton University, Princeton, NJ. Since 2009, he has been with the Institute for Telecommunications Research (ITR) in University of South Australia (UniSA), Adelaide, Australia, where he is now a senior research fellow. His current research interests include Shannon theory, visible light communications, information-theoretic security, and biometric security systems.

Dr. Ho was a recipient of the Croucher Foundation Fellowship for 2006–2008, the 2008 Young Scientist Award from the Hong Kong Institution of Science, UniSA Research SA Fellowship for 2010–2013, and the Australian Research Council Australian Postdoctoral Fellowship for 2010–2013.

Jin Meng received the B.Eng. degree in information and electronics engineering (2006) from Tsinghua University, Beijing, China, and Ph.D. degree in electrical and computer engineering (2013) from University of Waterloo, Waterloo, Ontario, Canada. He is now a postdoctoral fellow in Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ontario, Canada.

En-Hui Yang (M'97-SM'00-F'08) received the B.S. degree in applied mathematics from Huaqiao University, Quanzhou, China, and Ph.D. degree in mathematics from Nankai University, Tianjin, China, in 1986 and 1991, respectively.

Since June 1997, he has been with the Department of Electrical and Computer Engineering, University of Waterloo, ON, Canada, where he is currently a Professor and Canada Research Chair in information theory and multimedia compression. He held a Visiting Professor position at the Chinese University of Hong Kong, Hong Kong, from September 2003 to June 2004; positions of Research Associate and Visiting Scientist at the University of Minnesota, Minneapolis-St. Paul, the University of Bielefeld, Bielefeld, Germany, and the University of Southern California, Los Angeles, from January 1993 to May 1997; and a faculty position (first as an Assistant Professor and then an Associate Professor) at Nankai University, Tianjin, China, from 1991 to 1992. He is the founding Director of the Leitch-University of Waterloo multimedia communications lab, and a Co-Founder of SlipStream Data Inc. (now a subsidiary of BlackBerry). He currently also serves as an Executive Council Member of China Overseas Exchange Association and an Overseas Advisor for the Overseas Chinese Affairs Office of the City of Shanghai, and is sitting on the Overseas Expert Advisory Committee for the Overseas Chinese Affairs Office of the State Council of China and a Review Panel for the International Council for Science. His current research interests are: multimedia compression, multimedia transmission, digital communications, information theory, source and channel coding, image and video coding, image and video understanding and management, and Big Data analytics.

Dr. Yang is a recipient of several research awards including the 1992 Tianjin Science and Technology Promotion Award for Young Investigators; the 1992 third Science and Technology Promotion Award of Chinese Ministry of Education; the 2000 Ontario Premier's Research Excellence Award, Canada; the 2000 Marsland Award for Research Excellence, University of Waterloo; the 2002 Ontario Distinguished Researcher Award; the prestigious Inaugural Premier's Catalyst Award in 2007 for the Innovator of the Year; the 2007 Ernest C. Manning Award of Distinction, one of the Canada's most prestigious innovation prizes, and the 2013 CPAC Professional Achievement Award. Products based on his inventions and commercialized by SlipStream received the 2006 Ontario Global Traders Provincial Award. With over 200 papers and more than 200 patents/patent applications worldwide, his research work has had an impact on the daily life of hundreds of millions people over 170 countries. He is a Fellow of the Canadian Academy of Engineering and a Fellow of the Royal Society of Canada: the Academies of Arts, Humanities and Sciences of Canada. He served, among many other roles, as a General Co-Chair of the 2008 IEEE International Symposium on Information Theory, an Associate Editor for IEEE TRANSACTIONS ON INFORMATION THEORY, a Technical Program Vice-Chair of the 2006 IEEE International Conference on Multimedia & Expo (ICME), the Chair of the award committee for the 2004 Canadian Award in Telecommunications, a Co-Editor of the 2004 Special Issue of the IEEE TRANSACTIONS ON INFORMATION THEORY, a Co-Chair of the 2003 US National Science Foundation (NSF) workshop on the interface of Information Theory and Computer Science, and a Co-Chair of the 2003 Canadian Workshop on Information Theory.