Genie Chains: Exploring Outer Bounds on the Degrees of Freedom of MIMO Interference Networks

Chenwei Wang, *Member, IEEE*, Hua Sun, *Student Member, IEEE*, and Syed Ali Jafar, *Fellow, IEEE*

***Abstract*—In this paper, we propose a novel "genie chains" approach to obtain information theoretic degrees of freedom (DoF) outer bounds for MIMO wireless interference networks. This new approach creates a chain of mappings from genie signals provided to a receiver to the exposed signal spaces at that receiver, and then the exposed signal spaces serve as the genie signals for the next receiver in the chain subject to certain linear independence requirements. Our approach essentially converts an information theoretic DoF outer bound problem into a linear algebra problem. Several applications of the genie chains approach are presented.**

***Index Terms*—Capacity, degrees of freedom (DoF), MIMO, interference networks, outer bounds.**

I. Introduction

RECENTLY, Wang *et al.* characterized the spatially normalized degrees of freedom (DoF) for the $K = 3$ user $M_T \times M_R$ interference channel in [2] where each transmitter is equipped with M_T antennas, each receiver with M_R antennas, and M_T, M_R can take *arbitrary* positive integer values.[1] The DoF characterization is comprised of a piece-wise linear mapping with infinitely many linear intervals over the range of the parameter $\gamma = M/N$ where $M = \min(M_T, M_R)$, $N = \max(M_T, M_R)$, shedding light on several interesting elements such as redundant antenna dimensions, decomposability, subspace alignment chains and the feasibility of linear interference alignment. However, existing insights do not suffice beyond the 3-user $M_T \times M_R$ interference channel. In particular, finding good DoF outer bounds for K-user MIMO wireless interference networks continues to be a challenge. It is this challenge of finding good DoF outer bounds that motivates this work.

Manuscript received March 1, 2015; revised February 29, 2016; accepted July 25, 2016. Date of publication August 11, 2016; date of current version September 13, 2016. This work was supported in part by NSF under Grant CCF-1143982, Grant CCF-0830809, and Grant CCF-1317351, in part by ONR under Grant N00014-12-1-0067 and Grant N00014-16-1-2629, and in part by ARO under Grant W911NF-16-1-0215. This paper was presented in part at the 2012 IEEE International Symposium on Information Theory.

C. Wang is with the Mobile Network Technology Group, DOCOMO Innovations Inc., Palo Alto, CA 94304 USA (email: chenweiw@uci.edu).

H. Sun and S. A. Jafar are with the Center for Pervasive Communications and Computing, Department of Electrical Engineering and Computer Science, University of California Irvine, Irvine, CA 92697 USA (email: huas2@uci.edu; syed@uci.edu).

Communicated by T. Liu, Associate Editor for Shannon Theory.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2016.2599515

[1] A strictly weaker set of DoF results for the 3 user $M_T \times M_R$ wireless interference channel, restricted to linear precoding schemes without symbol extensions, is obtained independently by Bresler *et al.* in [3] in parallel work. The information theoretic outer bounds of Wang *et al.* in [2] match the linear outer bounds of Bresler *et al.* in [3], and the achievability in both [2] and [3] is based on linear schemes. Since information theoretic outer bounds imply linear outer bounds (but not vice versa), the results of Bresler *et al.* are strictly contained in the results of Wang *et al.*.

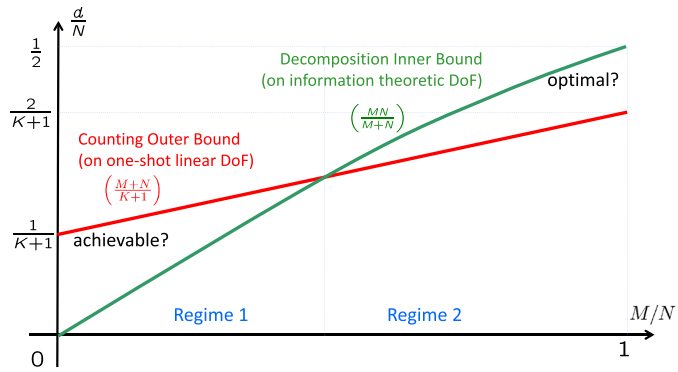

Fig. 1. The DoF counting outer bound (on linear DoF with no symbol extensions) [4] and the decomposition inner bound (on information theoretic DoF) [7] of the K-user $M \times N$ MIMO interference channel.

In order to clarify what we expect from good DoF outer bounds, it is worthwhile to summarize our expectation of the DoF results of MIMO wireless interference networks. This is simply our projection based on all previously known results, re-affirmed by our results in this work, and may be seen as a weak conjecture for the general results that remain elusive. We will focus on the K-user $M_T \times M_R$ wireless interference network and use the Figure 1 as an illustration. In this figure, the horizonal axis denotes the ratio $\gamma = M/N$, and the vertical axis denotes the DoF per user normalized by N. As in the 3-user setting, we use the notation $M = \min(M_T, M_R)$, $N = \max(M_T, M_R)$. There are two curves in the figure. The red straight line, which we label as the "counting" *outer* bound, plots the value $d = \frac{M+N}{K+1}$, and the green curve, which we label as the "decomposition" *inner* bound, plots the value $d = \frac{MN}{M+N}$. An understanding of these two curves is essential to the understanding the DoF of the K-user $M_T \times M_R$ interference channel.

A dichotomy is evident in the existing DoF results for K-user MIMO wireless interference networks. On the one hand, we have the question of linear DoF, i.e., the DoF achievable (almost surely) by linear precoding *without* symbol

extensions in time/frequency. Spatial extension, i.e., scaling of antennas at every node by the same factor, is allowed in this setting. The key distinction between spatial extensions and time/frequency extensions is that the former can only produce generic (structureless) channels whereas the latter give rise to structured (block-diagonal) channel matrices. The linear schemes studied along this research avenue are designed mainly for unstructured generic channels, so they do not benefit from the channel structure, but they *may* be hurt by it if the channel structure causes an overlap of desired and interfering signals. The key to the linear DoF question is the distinction of proper versus improper systems, introduced by Yetis *et al.* in [4] through the counting bound. A system is proper if $d \leq \frac{M+N}{K+1}$ and improper otherwise. The counting bound is obtained simply by counting the number of alignment constraints and comparing it to the number of design variables. If the number of constraints exceeds the number of variables the system is labeled improper. It is labeled proper otherwise. Yetis *et al.* conjecture that improper systems are infeasible (when restricted to linear schemes over unstructured channels), whereas proper systems are feasible (through linear schemes over unstructured channels) *provided* they are information theoretically feasible, i.e., that they satisfy the information theoretic DoF bounds. The first conjecture of Yetis *et al.* is proved by Bresler *et al.* in [3] and by Razaviyayn *et al.* in [5]. The second conjecture of Yetis *et al.* is consistent with all DoF results known so far, including the 3-user case, but has not been proved in general.

On the other hand, we have the question of information theoretic DoF, i.e., DoF achievable (almost surely) by linear and non-linear schemes, with no constraints on symbol extensions. It has been observed, and indeed it has been conjectured by Jafar in [6] that linear schemes over arbitrarily long symbol extensions are still sufficient to achieve the optimal DoF, if generic time-variations are allowed. In the absence of time-variations, more sophisticated schemes, e.g., those based on rational alignments, may be involved. As far as spatial extensions are concerned, there is the spatial scale invariance conjecture by Jafar in [2] and [6] that claims that if the number of antennas at every node is scaled by a certain factor, then the information theoretic DoF will scale by the same factor. The spatial scale invariance conjecture is consistent with all known results but has not been proved in general. This is in part because few good information theoretic outer bounds are known. However, the most important aspect of this discussion is the achievability result by [7], that shows that in a K-user $M_T \times M_R$ wireless interference channel, each user is able to achieve $\frac{MN}{M+N}$ DoF by first decomposing multiple antenna nodes into multiple single antenna nodes, and then using the asymptotic alignment scheme of Cadambe and Jafar [8] (the CJ scheme) over the resulting SISO network, precoding over linear vector space dimensions if channels are time-varying, and over rational scalar dimensions if the channels are constant.

The counting bound is an outer bound on the linear DoF, thus restricted to linear precoding schemes with no symbol extensions. The decomposition bound is an inner bound on information theoretic DoF, thus with no restrictions on the type of coding scheme or the use of symbol extensions. At first sight, the two have little to do with each other. And yet, the two seem to play an important joint role as we explain next. First, note that there are two distinct regimes, labeled Regime 1 and Regime 2 in Figure 1, where the counting bound dominates the decomposition bound and the decomposition bound dominates the counting bound, respectively. Regime 1 is relatively well understood, especially because of the recent insights from the DoF characterization of the 3-user $M_T \times M_R$ MIMO interference channel by Wang *et al.* [2]. Note that the 3-user setting contains only Regime 1. This is easily seen because when $K = 3$, the counting bound $\frac{M+N}{K+1} = \frac{M+N}{4}$ is always greater than or equal to the decomposition bound $\frac{MN}{M+N}$. That is,

$$\frac{M+N}{4} - \frac{MN}{M+N} = \frac{(N-M)^2}{4(M+N)} \geq 0. \tag{1}$$

As we will see in this work, the insights from the 3-user case generalize in a relatively straightforward manner to *most* of Regime 1 of the K user setting: in both cases the optimal DoF curve (for both information theoretic DoF and linear DoF) is piecewise linear, with the linear segments bouncing between the counting bound and the decomposition bound, as they do in the 3-user interference channel.

For this work, it is Regime 2 that is most intriguing. Some interesting observations can be made here. First, note that because the decomposition bound dominates the counting bound, the second conjecture of Yetis *et al.* would suggest that proper systems in this regime should be feasible with linear precoding and no symbol extensions. Because improper systems are already known to be infeasible, if the conjecture holds, it would settle the linear feasibility question for all systems in Regime 2, i.e., the counting bound would be optimal for linear DoF. This is indeed an interesting observation. However, the main question that interests us in this work has to do with the information theoretic DoF, and the information theoretic optimality of the decomposition bound in Regime 2. To test such a hypothesis, we need better information theoretic DoF outer bounds. So we will develop a novel "genie chains" approach that will give us an information theoretic outer bound in terms of a linear algebra problem, specifically requiring the computation of the ranks of certain matrices. The downside is that these matrices become large as the MIMO dimensions M_T, M_R increase, so that we face computational bottlenecks. The upside, however, is that for most practically reasonable values of M_T, M_R, as well as for certain sub-regimes of Regime 2, we are able to compute the outer bound, and indeed verify that it matches the decomposition inner bound. We summarize these observations in a loosely stated conjecture, that the decomposition bound is DoF optimal in *most* of Regime 2.

II. System Model

Consider a fully connected K-user MIMO interference channel where there are M_T and M_R antennas at each transmitter (TX) and receiver (RX), respectively, and each TX has one independent message, intended for its corresponding RX. Denote by $\mathbf{H}^{[ji]}$ the $M_R \times M_T$ channel matrix from

TX i to RX j where $i, j \in \mathcal{K} \triangleq \{1, \cdots, K\}$. For simplicity, we assume that the channel coefficients are independently drawn from a continuous distribution. While we will assume that the channels are constant for simplicity, we note that it is straightforward to extend our DoF outer bounds to the setting where the channel coefficients are varying in time/frequency. Global channel knowledge is assumed to be available at all nodes. For codebooks spanning n channel uses, at time index $t \in \{1, 2, \cdots, n\}$, TX i sends a complex-valued $M_T \times 1$ signal vector $\mathbf{X}^{[i]}(t)$, which satisfies an average power constraint $\frac{1}{n}\sum_{t=1}^{n} \mathbb{E}[\|\mathbf{X}^{[i]}(t)\|^2] \leq \rho$. At the RX side, RX j observes an $M_R \times 1$ signal vector $\bar{\mathbf{Y}}^{[j]}(t)$ at time index t, which is given by:

$$\bar{\mathbf{Y}}^{[j]}(t) = \underbrace{\sum_{i=1}^{K} \mathbf{H}^{[ji]}\mathbf{X}^{[i]}(t)}_{\triangleq \mathbf{Y}^{[j]}(t)} + \mathbf{Z}^{[j]}(t) \tag{2}$$

where $\mathbf{Z}^{[j]}(t)$ is an $M_R \times 1$ column vector representing the i.i.d. circularly symmetric complex additive white Gaussian noise (AWGN) at RX j, each entry of which is an i.i.d. Gaussian random variable with zero-mean and unit-variance.

As a function of the signal-to-noise ratio (SNR) parameter ρ, let $R_k(\rho) = R(\rho)$ denote the symmetric capacity, i.e., the highest rate simultaneously achievable by each user. We define $d(K, M_T, M_R) \triangleq \lim_{\rho\to\infty} R(\rho)/\log\rho$ as the symmetric DoF per user. Here, the user index k is interpreted *modulo* K so that, e.g., User 1 is the same as User $K+1$, etc. The dependence on K, M_T, M_R may be dropped for compact notation when no ambiguity would be caused. Moreover, we use $o(x)$ to represent any function $f(x)$ such that $\lim_{x\to\infty} f(x)/x = 0$. Furthermore, we define $M = \min(M_T, M_R)$, $N = \max(M_T, M_R)$.

III. A Vector Space Perspective

In this section, we introduce a vector space perspective, and its associated notation, terminology and basic properties, that we will later use for information theoretic DoF outer bounds.

Consider a TX with M_T antennas, which transmits the $M_T \times 1$ vector $\mathbf{X}(i)$ over the ith channel use, and satisfies an average transmit power constraint $\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[\|\mathbf{X}(i)\|^2] \leq \rho$ across n channel uses. We will denote by $\mathbf{X}^n = \{\mathbf{X}(1), \mathbf{X}(2), \cdots, \mathbf{X}(n)\}$, the n vectors sent over the n channel uses. When referring to the vector transmitted over a single channel use, we will suppress the channel use index for brevity (whenever the particular channel use index is not significant) and simply refer to it as the $M_T \times 1$ vector $\mathbf{X} = [X_1, X_2, \cdots, X_{M_T}]^T$.

The vector $\mathbf{X}$ lies in the M_T-dimensional vector space spanned by the columns of the $M_T \times M_T$ identity matrix. We are interested in the projections of $\mathbf{X}$ into vector subspaces, and the differential entropies of the projections under additive Gaussian noise. The notation and the underlying concepts are best explained through examples. Suppose $M_T = 3$, i.e., we are operating in a 3-dimensional vector space, and let us consider the following 2-dimensional vector subspace:

$$\mathbf{L} = \text{column span}\left(\begin{bmatrix} 1 & 2 \\ 1 & 0 \\ 0 & 3 \end{bmatrix}\right). \tag{3}$$

Choosing a basis for this subspace, such as the one shown in (3), let us project $\mathbf{X}$ into this basis, say $B_1(\mathbf{L})$, giving us:

$$B_1(\mathbf{L})^T\mathbf{X} = \begin{bmatrix} 1 & 2 \\ 1 & 0 \\ 0 & 3 \end{bmatrix}^T \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} X_1 + X_2 \\ 2X_1 + 3X_3 \end{bmatrix}. \tag{4}$$

Note that a different choice of basis for the same subspace, say $B_2(\mathbf{L})^T = A_{2\times 2}B_1(\mathbf{L})^T$, where $A_{2\times 2}$ is an arbitrary 2×2 full rank matrix, will give us a different projected vector, such as:

$$\begin{aligned} A_{2\times 2}B_1(\mathbf{L})^T\mathbf{X} &= \begin{bmatrix} 1 & 1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 0 \\ 0 & 3 \end{bmatrix}^T \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \\ &= \begin{bmatrix} 3 & 3 \\ 1 & -1 \\ 3 & 6 \end{bmatrix}^T \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} 3X_1 + X_2 + 3X_3 \\ 3X_1 - X_2 + 6X_3 \end{bmatrix}. \end{aligned} \tag{5}$$

However, as we will soon establish, since we are interested only in DoF, the choice of basis is not important for our purpose. Only the span of the space itself is significant.

Next, let us also bring in additive noise into the picture. Given any vector of random variables $\mathbf{U} = [U_1, U_2, \cdots, U_k]^T$, let us define the differential entropy of its noisy version as

$$\hbar(\mathbf{U}) \triangleq h(\mathbf{U} + \mathbf{Z}) = h(U_1 + Z_1, \cdots, U_k + Z_k) \tag{6}$$

where $h(\cdot)$ is the standard differential entropy function, $\mathbf{Z} = [Z_1, Z_2, \cdots, Z_k]^T$ is a circularly symmetrically additive white Gaussian noise vector that is independent of $\mathbf{U}$ and $\mathbf{Z} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$. Similar definitions are used for joint and conditional differential entropies, i.e.,

$$\begin{aligned} &\hbar(\mathbf{U}, \mathbf{V}) \\ &\quad \triangleq h(\mathbf{U} + \mathbf{Z}, \mathbf{V} + \mathbf{Z}') \\ &\quad = h(U_1 + Z_1, \cdots, U_k + Z_k, V_1 + Z_1', \cdots, V_k + Z_k') \\ &\hbar(\mathbf{U}|\mathbf{V}) \\ &\quad \triangleq \hbar(\mathbf{U}, \mathbf{V}) - \hbar(\mathbf{V}) \\ &\quad = h(\mathbf{U} + \mathbf{Z}, \mathbf{V} + \mathbf{Z}') - h(\mathbf{V} + \mathbf{Z}') \\ &\quad = h(\mathbf{U} + \mathbf{Z}|\mathbf{V} + \mathbf{Z}') \\ &\quad = h(U_1 + Z_1, \cdots, U_k + Z_k | V_1 + Z_1', \cdots, V_k + Z_k'), \end{aligned}$$

where $\mathbf{V} = [V_1, V_2, \cdots, V_k]^T$ is another vector of random variables and $\mathbf{Z}' = [Z_1', Z_2', \cdots, Z_k']^T \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$, which is independent of $\mathbf{U}, \mathbf{V}$ and $\mathbf{Z}$.

Lemma 1: Consider an arbitrary subspace $\mathbf{L}$ of the M_T-dimensional vector space $\mathbb{C}^{M_T}$ and let $B_i(\mathbf{L})$, $B_j(\mathbf{L})$ be two arbitrary choices for the basis of $\mathbf{L}$. We have

$$\hbar(B_i(\mathbf{L})^T\mathbf{X}) = h(B_j(\mathbf{L})^T\mathbf{X} + \tilde{\mathbf{Z}}) + o(\log\rho), \tag{7}$$

where $\tilde{\mathbf{Z}} \sim \mathcal{CN}(\mathbf{0}, \tilde{\mathbf{K}})$, and $\tilde{\mathbf{K}}$ is a non-singular covariance matrix. We require that $\mathbf{L}, B_i(\mathbf{L}), B_j(\mathbf{L}), \tilde{\mathbf{K}}$ are held fixed as $\rho \to \infty$.

Proof: We defer the proof to Appendix A. ■

According to Lemma 1, as long as the subspace $\mathbf{L}$, its basis representation $B(\mathbf{L})$ and the additive noise terms $\tilde{\mathbf{Z}}$ do not depend on the SNR, ρ, and the noise in the projected subspace is non-singular, then all that matters is the subspace $\mathbf{L}$ within which $\mathbf{X}$ is projected. Neither the particular choice of basis representation, nor the specific form of the noise covariance matrix is relevant.

In light of this observation, we will henceforth simplify our notation by referring to $\hbar(B(\mathbf{L})^T\mathbf{X})$ as $\hbar(\mathbf{L}\circ\mathbf{X})$ instead, where the symbol "$\circ$" denotes the projection operation, with the understanding that the given representation of $\mathbf{L}$ is equivalent to any other basis representation of the same space for our purpose.

Lemma 1 extends easily to joint and conditional differential entropies as well, for which still only the space matters, not the specific basis representation chosen. For two subspaces $\mathbf{L}^{[1]}, \mathbf{L}^{[2]}$ of $\mathbb{C}^M$, we define $\hbar(\mathbf{L}^{[1]}\circ\mathbf{X}, \mathbf{L}^{[2]}\circ\mathbf{X})$ and $\hbar(\mathbf{L}^{[1]}\circ\mathbf{X}|\mathbf{L}^{[2]}\circ\mathbf{X})$ in a similar way to refer to the joint and conditional differential entropies of $\mathbf{X}$ projected in corresponding spaces, respectively.

It is useful to further familiarize ourselves with the vector space representations, for instance, with unions and intersection operations. Once again, we illustrate these with a simple example. Consider the following subspaces:

$$\mathbf{L}_1^{[1]} = \text{span}([1 \quad 1 \quad 0]^T), \tag{8}$$
$$\mathbf{L}_2^{[1]} = \text{span}([2 \quad 0 \quad 3]^T), \tag{9}$$
$$\mathbf{L}_1^{[2]} = \text{span}([2 \quad -1 \quad 4]^T), \tag{10}$$
$$\mathbf{L}_2^{[2]} = \text{span}([-2 \quad -3 \quad 1]^T), \tag{11}$$

and let $\mathbf{L}^{[1]}, \mathbf{L}^{[2]}$ be defined as the vector spaces spanned by the unions:

$$\mathbf{L}^{[1]} = \{\mathbf{L}_1^{[1]}, \mathbf{L}_2^{[1]}\}, \tag{12}$$
$$\mathbf{L}^{[2]} = \{\mathbf{L}_1^{[2]}, \mathbf{L}_2^{[2]}\}. \tag{13}$$

Note that since the union of vector spaces is not generally a vector space, what is meant here is that $\mathbf{L}^{[i]}$ is the vector space spanned by the union of the vector subspaces $\mathbf{L}_1^{[i]}, \mathbf{L}_2^{[i]}$.

Next let us consider the intersection of $\mathbf{L}^{[1]}$ and $\mathbf{L}^{[2]}$. Note that given $\mathbf{L}^{[i]}$, we can compute $\mathbf{L}^{[i]^c}$ which is the subspace orthogonal to the span of $(\mathbf{L}_1^{[i]}, \mathbf{L}_2^{[i]})$. That is,

$$\mathbf{L}^{[1]^c} = \text{span}([3 \quad -3 \quad -2]^T), \tag{14}$$
$$\mathbf{L}^{[2]^c} = \text{span}([-5.5 \quad 5 \quad 4]^T). \tag{15}$$

Thus, the intersection $\mathbf{L}^{[1]}\cap\mathbf{L}^{[2]}$ can be obtained by computing the subspace orthogonal to both $\mathbf{L}^{[1]^c}$ and $\mathbf{L}^{[2]^c}$, and thus it can be written as:

$$\mathbf{L}^{[1]} \cap \mathbf{L}^{[2]} = ([\mathbf{L}^{[1]^c} \ \mathbf{L}^{[2]^c}])^c = \text{span}([4 \quad 2 \quad 3]^T). \tag{16}$$

Similarly, we define $\mathbf{L}^{[1]} \setminus \mathbf{L}^{[2]}$ to be the subspace of $\mathbf{L}^{[1]}$, which is orthogonal to $\mathbf{L}^{[1]} \cap \mathbf{L}^{[2]}$, i.e.,

$$\begin{aligned}\mathbf{L}^{[1]} \setminus \mathbf{L}^{[2]} &= \mathbf{L}^{[1]} \setminus (\mathbf{L}^{[1]} \cap \mathbf{L}^{[2]}) = \mathbf{L}^{[1]} \cap (\mathbf{L}^{[1]} \cap \mathbf{L}^{[2]})^c \\ &= \text{span}([5 \quad 17 \quad -18]^T).\end{aligned} \tag{17}$$

With this definition, we can also write $\mathbf{L}^{[1]}$ as

$$\mathbf{L}^{[1]} = \{\mathbf{L}^{[1]} \cap \mathbf{L}^{[2]}, \ \mathbf{L}^{[1]} \setminus \mathbf{L}^{[2]}\}. \tag{18}$$

A set of $M_T \times 1$ vectors is generic if and only if any m of them are linearly independent whenever $m \leq M_T$. A generic subspace is the subspace spanned by the column vectors of a matrix, where the column vectors are generic.

In this paper, because we are primarily interested in the notion of DoF, we will use the notations $x(\rho, n) =: y(\rho, n)$, $x(\rho, n) \leq: y(\rho, n)$, $x(\rho, n) \geq: y(\rho, n)$ to represent $x(\rho, n) = y(\rho, n) + n\ o(\log\rho)$, $x(\rho, n) \leq y(\rho, n) + n\ o(\log\rho)$, $x(\rho, n) \geq y(\rho, n) + n\ o(\log\rho)$, respectively. Next we summarize the basic properties associated with the vector subspace representations. The properties are stated in the multi-letter form, which is used in the information theoretic proofs. As such, we extend the vector space terminologies introduced above to their corresponding multi-letter forms.

$\mathbf{L}^n \triangleq \mathbf{L}(1) \times \mathbf{L}(2) \times \cdots \times \mathbf{L}(n)$ is used to represent the collection of n subspaces $\mathbf{L}(1), \mathbf{L}(2), \cdots, \mathbf{L}(n)$ of the M_T-dimensional vector space $\mathbb{C}^{M_T}$. If the dimension of the n subspaces $\mathbf{L}(t), t \in \{1, 2, \ldots, n\}$ is the same, we will denote it as $|\mathbf{L}|$. The basis representation $B(\mathbf{L}^n)$ of $\mathbf{L}^n$ is the collection of the basis representations of each subspace, i.e., $B(\mathbf{L}^n) \triangleq B(\mathbf{L}(1)) \times \cdots \times B(\mathbf{L}(n))$. Also, $\mathbb{C}^{M_T^n} \triangleq \mathbb{C}^{M_T} \times \cdots \times \mathbb{C}^{M_T}$. For two multi-letter subspaces $\mathbf{L}^{[1]^n}$ and $\mathbf{L}^{[2]^n}$, their intersection $\mathbf{L}^{[1]^n} \cap \mathbf{L}^{[2]^n}$ is defined as

$$\mathbf{L}^{[1]^n} \cap \mathbf{L}^{[2]^n} \triangleq \mathbf{L}^{[1]}(1) \cap \mathbf{L}^{[2]}(1) \times \cdots \times \mathbf{L}^{[1]}(n) \cap \mathbf{L}^{[2]}(n). \tag{19}$$

Similar definitions are employed for $\mathbf{L}^{[1]^n} \setminus \mathbf{L}^{[2]^n}$, $\mathbf{L}^n \circ \mathbf{X}^n$ and $B(\mathbf{L}^n)^T\mathbf{X}^n$,

$$\begin{aligned}\mathbf{L}^{[1]^n} \setminus \mathbf{L}^{[2]^n} &\triangleq \mathbf{L}^{[1]}(1) \setminus \mathbf{L}^{[2]}(1) \times \cdots \times \mathbf{L}^{[1]}(n) \setminus \mathbf{L}^{[2]}(n), \\ \mathbf{L}^n \circ \mathbf{X}^n &\triangleq \mathbf{L}(1) \circ \mathbf{X}(1) \times \cdots \times \mathbf{L}(n) \circ \mathbf{X}(n), \\ B(\mathbf{L}^n)^T\mathbf{X}^n &\triangleq B(\mathbf{L}(1))^T\mathbf{X}(1) \times \cdots \times B(\mathbf{L}(n))^T\mathbf{X}(n).\end{aligned}$$

Equipped with these definitions, following (18), we can write

$$\mathbf{L}^{[1]^n} = \{\mathbf{L}^{[1]^n} \cap \mathbf{L}^{[2]^n}, \ \mathbf{L}^{[1]^n} \setminus \mathbf{L}^{[2]^n}\}. \tag{20}$$

Next, we proceed to the statement of the properties.

Lemma 2: We have the following properties:

(P1) $\hbar(B_i(\mathbf{L}^n)^T\mathbf{X}^n) =: \hbar(B_j(\mathbf{L}^n)^T\mathbf{X}^n)$ for any basis representations $B_i(\mathbf{L}^n)$, $B_j(\mathbf{L}^n)$ of $\mathbf{L}^n$.

Justified by this property, we will write $\hbar(B(\mathbf{L}^n)^T\mathbf{X}^n)$ simply as $\hbar(\mathbf{L}^n \circ \mathbf{X}^n)$.

(P2) $\hbar(\mathbf{L}^n \circ \mathbf{X}^n) \leq: n|\mathbf{L}|\log\rho$.

(P3) For generic subspaces $\mathbf{L}^{[1]}(t), \mathbf{L}^{[2]}(t), t \in \{1, 2, \cdots, n\}$ of $\mathbb{C}^{M_T}$ with $|\mathbf{L}^{[1]}| + |\mathbf{L}^{[2]}| \geq M_T$, we have:

P3a) $\hbar(\mathbf{L}^{[1]^n} \circ \mathbf{X}^n, \mathbf{L}^{[2]^n} \circ \mathbf{X}^n) =: \hbar(\mathbf{X}^n)$.

P3b) $\hbar(\mathbf{L}^{[1]^n} \circ \mathbf{X}^n | \mathbf{L}^{[2]^n} \circ \mathbf{X}^n) =: \hbar((\mathbb{C}^{M_T^n} \setminus \mathbf{L}^{[2]^n}) \circ \mathbf{X}^n | \mathbf{L}^{[2]^n} \circ \mathbf{X}^n)$.

P3c) $\min(\hbar(\mathbf{L}^{[1]^n} \circ \mathbf{X}^n | \mathbf{L}^{[2]^n} \circ \mathbf{X}^n), \hbar(\mathbf{L}^{[2]^n} \circ \mathbf{X}^n | \mathbf{L}^{[1]^n} \circ \mathbf{X}^n)) \leq: \frac{1}{2}\hbar(\mathbf{X}^n)$.

Proof: We will show the proofs for each property sequentially.

Property (P1): This property is the multi-letter version of Lemma 1, whose proof follows directly.

Property (P2):

$$\begin{aligned}
&\hbar(\mathbf{L}^n \circ \mathbf{X}^n)\\
&= \sum_{t=1}^{n} \hbar(\mathbf{L}(t) \circ \mathbf{X}(t) | \mathbf{L}(1) \circ \mathbf{X}(1), \cdots, \mathbf{L}(t-1) \circ \mathbf{X}(t-1))\\
&\leq \sum_{t=1}^{n} \hbar(\mathbf{L}(t) \circ \mathbf{X}(t)) \qquad (21)\\
&= \sum_{t=1}^{n} \sum_{i=1}^{|\mathbf{L}|} \hbar(\mathbf{L}_i(t) \circ \mathbf{X}(t) | \mathbf{L}_{i-1}(t) \circ \mathbf{X}(t), \cdots, \mathbf{L}_1(t) \circ \mathbf{X}(t)) \qquad (22)\\
&\leq \sum_{t=1}^{n} \sum_{i=1}^{|\mathbf{L}|} \hbar(\mathbf{L}_i(t) \circ \mathbf{X}(t)) \qquad (23)\\
&\leq: n|\mathbf{L}| \log \rho \qquad (24)
\end{aligned}$$

where (21) follows from the fact that removing conditional terms does not decrease the differential entropy; (22) is obtained due to the chain rule and $\mathbf{L}_i(t)$ denotes the space spanned by the i-th basis vector of $\mathbf{L}(t)$; (24) is obtained because one dimension can contribute upto one $\log \rho + o(\log \rho)$ term.

In addition, incorporating (P1), we can also see that if $|\mathbf{L}| = M_T$, then $\hbar(\mathbf{X}^n) = \hbar(\mathbb{C}^{M^n} \circ \mathbf{X}^n) =: \hbar(\mathbf{L}^n \circ \mathbf{X}^n) \leq: nM_T \log \rho$.

Property (P3a):

$$\begin{aligned}
&\hbar(\mathbf{L}^{[1]^n} \circ \mathbf{X}^n, \mathbf{L}^{[2]^n} \circ \mathbf{X}^n)\\
&= \hbar(\{\mathbf{L}^{[1]^n} \cap \mathbf{L}^{[2]^n}, \mathbf{L}^{[1]^n} \setminus \mathbf{L}^{[2]^n}\} \circ \mathbf{X}^n, \mathbf{L}^{[2]^n} \circ \mathbf{X}^n) \qquad (25)\\
&= \hbar((\mathbf{L}^{[1]^n} \cap \mathbf{L}^{[2]^n}) \circ \mathbf{X}^n, (\mathbf{L}^{[1]^n} \setminus \mathbf{L}^{[2]^n}) \circ \mathbf{X}^n, \mathbf{L}^{[2]^n} \circ \mathbf{X}^n) \qquad (26)\\
&= \hbar((\mathbf{L}^{[1]^n} \cap \mathbf{L}^{[2]^n}) \circ \mathbf{X}^n, \{\mathbf{L}^{[1]^n} \setminus \mathbf{L}^{[2]^n}, \mathbf{L}^{[2]^n}\} \circ \mathbf{X}^n) \qquad (27)\\
&=: \hbar((\mathbf{L}^{[1]^n} \cap \mathbf{L}^{[2]^n}) \circ \mathbf{X}^n, \mathbb{C}^{M_T^n} \circ \mathbf{X}^n) \qquad (28)\\
&= \hbar((\mathbf{L}^{[1]^n} \cap \mathbf{L}^{[2]^n}) \circ \mathbf{X}^n, \mathbf{X}^n) \qquad (29)\\
&= \hbar(\mathbf{X}^n) + \hbar((\mathbf{L}^{[1]^n} \cap \mathbf{L}^{[2]^n}) \circ \mathbf{X}^n | \mathbf{X}^n) \qquad (30)\\
&= \hbar(\mathbf{X}^n) + n\ o(\log \rho) \qquad (31)\\
&=: \hbar(\mathbf{X}^n) \qquad (32)
\end{aligned}$$

where (25) follows directly from (20); both (26) and (27) are obtained due to the fact that the two subspaces participating the splitting or the union operations are orthogonal to each each other; (28) follows from the assumption $|\mathbf{L}^{[1]}| + |\mathbf{L}^{[2]}| \geq M_T$ and Property (P1); and (31) is obtained because the subspace $\mathbf{L}^{[1]^n} \cap \mathbf{L}^{[2]^n}$ is contained in $\mathbb{C}^{M_T^n}$. Basically, it implies that the M_T variables comprising the vector $\mathbf{X}$ can be used to construct any linear combination of $\mathbf{X}$ subject to the bounded noise distortion.

Property (P3b):

$$\begin{aligned}
&\hbar(\mathbf{L}^{[1]^n} \circ \mathbf{X}^n | \mathbf{L}^{[2]^n} \circ \mathbf{X}^n)\\
&= \hbar((\mathbf{L}^{[1]^n} \setminus \mathbf{L}^{[2]^n}) \circ \mathbf{X}^n, (\mathbf{L}^{[1]^n} \cap \mathbf{L}^{[2]^n}) \circ \mathbf{X}^n | \mathbf{L}^{[2]^n} \circ \mathbf{X}^n)\\
&= \hbar(\{\mathbf{L}^{[1]^n} \setminus \mathbf{L}^{[2]^n}, \mathbf{L}^{[2]^n}\} \circ \mathbf{X}^n, (\mathbf{L}^{[1]^n} \cap \mathbf{L}^{[2]^n}) \circ \mathbf{X}^n | \cdots\\
&\qquad \cdots \mathbf{L}^{[2]^n} \circ \mathbf{X}^n)\\
&=: \hbar(\mathbf{X}^n, (\mathbf{L}^{[1]^n} \cap \mathbf{L}^{[2]^n}) \circ \mathbf{X}^n | \mathbf{L}^{[2]^n} \circ \mathbf{X}^n)\\
&= \hbar(\mathbf{X}^n | \mathbf{L}^{[2]^n} \circ \mathbf{X}^n)\\
&\quad + h((\mathbf{L}^{[1]^n} \cap \mathbf{L}^{[2]^n}) \circ \mathbf{X}^n | \mathbf{L}^{[2]^n} \circ \mathbf{X}^n, \mathbf{X}^n)\\
&= \hbar((\mathbb{C}^{M_T^n} \setminus \mathbf{L}^{[2]^n}) \circ \mathbf{X}^n | \mathbf{L}^{[2]^n} \circ \mathbf{X}^n) + n\ o(\log \rho)\\
&=: \hbar((\mathbb{C}^{M_T^n} \setminus \mathbf{L}^{[2]^n}) \circ \mathbf{X}^n | \mathbf{L}^{[2]^n} \circ \mathbf{X}^n).
\end{aligned}$$

The intuition of this property is that adding $\hbar(\mathbf{L}^{[2]^n} \circ \mathbf{X}^n)$ to both sides of the equation produces the $\hbar(\mathbf{X}^n)$ term on both sides.

Property (P3c):

$$\begin{aligned}
&\min(\hbar(\mathbf{L}^{[1]^n} \circ \mathbf{X}^n | \mathbf{L}^{[2]^n} \circ \mathbf{X}^n), \hbar(\mathbf{L}^{[2]^n} \circ \mathbf{X}^n | \mathbf{L}^{[1]^n} \circ \mathbf{X}^n))\\
&\leq \frac{1}{2}\left[\hbar(\mathbf{L}^{[1]^n} \circ \mathbf{X}^n | \mathbf{L}^{[2]^n} \circ \mathbf{X}^n) + \hbar(\mathbf{L}^{[2]^n} \circ \mathbf{X}^n | \mathbf{L}^{[1]^n} \circ \mathbf{X}^n)\right]\\
&\leq \frac{1}{2}\left[\hbar(\mathbf{L}^{[1]^n} \circ \mathbf{X}^n | \mathbf{L}^{[2]^n} \circ \mathbf{X}^n) + \hbar(\mathbf{L}^{[2]^n} \circ \mathbf{X}^n)\right] \qquad (33)\\
&= \frac{1}{2}\hbar(\mathbf{L}^{[1]^n} \circ \mathbf{X}^n, \mathbf{L}^{[2]^n} \circ \mathbf{X}^n) \qquad (34)\\
&=: \frac{1}{2}\hbar(\mathbf{X}^n) \qquad (35)
\end{aligned}$$

where (35) is obtained owing to Property (P3a). ■

A. Multiple Subspaces of the Vector Space

In this section, we introduce important properties associated with vector subspaces.

Suppose we have Kn subspaces $\mathbf{L}^{[k]}(t)$, $t \in \{1, 2, \cdots, n\}$, $k \in \{1, 2, \cdots, K\}$ of the M_T-dimensional vector space $\mathbb{C}^{M_T}$. The dimension of $\mathbf{L}^{[k]}(t)$ is $l_k, \forall t$ and we define $l^* \triangleq \sum_{k=1}^{K} l_k$. Over the tth channel use, we enumerate all the l^* basis vectors contained in the K subspaces, $\mathbf{L}^{[k]}(t)$ and denote the span of these vectors as $\mathbf{L}_1(t), \mathbf{L}_2(t), \cdots, \mathbf{L}_{l^*}(t)$, so that the basis representation of $\mathbf{L}^{[1]}(t)$ is comprised of the first l_1 basis vectors, the basis representation of $\mathbf{L}^{[2]}(t)$ the next l_2 basis vectors and so forth. Repeating such enumeration for all channel uses, we have

$$\begin{aligned}
\mathbf{L}^{[1]^n} &= \{\mathbf{L}_1^n, \mathbf{L}_2^n, \cdots, \mathbf{L}_{l_1}^n\}, \qquad (36)\\
\mathbf{L}^{[2]^n} &= \{\mathbf{L}_{l_1+1}^n, \cdots, \mathbf{L}_{l_1+l_2}^n\}, \qquad (37)\\
&\vdots\\
\mathbf{L}^{[K]^n} &= \left\{\mathbf{L}_{\sum_{i=1}^{K-1} l_i+1}^n, \cdots, \mathbf{L}_{l^*}^n\right\}. \qquad (38)
\end{aligned}$$

Now, let us start sequentially in the order of $\mathbf{L}^{[1]^n}, \mathbf{L}^{[2]^n}, \ldots$ to collect subspaces into a set and go as far as we can without the total number of *linear independent* basis vectors exceeding M_T. There are two possibilities. If we happen to collect exactly M_T independent vectors then we set these aside and start building the next set of vectors, proceeding sequentially again from where the first set terminated. On the other hand, if we fall short of M_T vectors, i.e., we cannot include the next subspace in the set without exceeding a total of M_T independent vectors in the set, then we need to split the next subspace into two parts. This is done by taking the intersection of the next subspace in the sequence with the space spanned by the basis vectors in the current set to form the intersecting space. The intersecting part is separated out as the remainder of the subspace, and the non-intersecting part is incorporated into the set to complete the desired M_T

independent vectors. The process then continues with the remaining subspaces, starting with the remainder of the most recently split subspace. The process is terminated when we run out of basis vectors. The number of complete sets (sets of M_T linearly independent basis vectors) that are generated through this process is denoted as L_Σ. The remaining basis vectors are discarded if they are insufficient to create another complete basis.

We now proceed to the statement of the properties on vector subspaces.

Lemma 3: The following bound on the entropy holds:

$$\sum_{k=1}^{K} \hbar\left(\mathbf{L}^{[k]^n} \circ \mathbf{X}^n\right) \geq: L_\Sigma \hbar(\mathbf{X}^n). \tag{39}$$

Intuitively, Lemma 3 implies that when a collection of l^* linear combinations of the M_T variables comprising $\mathbf{X}$ can reconstruct $\mathbf{X}$ L_Σ times, the equations must carry at least their proportional share of the total entropy of $\mathbf{X}$. Note that when the subspaces are generic, $L_\Sigma = \lfloor \frac{l^*}{M_T} \rfloor$.

We start with two simple cases, and then present the general proof.

Case 1: $K = 1,\ l_1 = M_T$.

In this case, $l^* = l_1 = M_T$ and $L_\Sigma = 1$. Property (P1) of Lemma 2 gives us

$$\hbar\left(\mathbf{L}^{[1]^n} \circ \mathbf{X}^n\right) =: \hbar\left(\mathbb{C}^{M_T^n} \circ \mathbf{X}^n\right) = \hbar(\mathbf{X}^n) \tag{40}$$

which implies (39).

Case 2: $K = M_T + 1$, $M_T > 1$, $l_k = 1$, $\forall k \in \{1, 2, \ldots, K\}$ and all the subspaces are generic.

In this case, we want to create $L_\Sigma = 1$ set with M basis vectors with $M_T + 1$ generic vectors. Then we have

$$\begin{aligned}
&\sum_{k=1}^{K} \hbar\left(\mathbf{L}^{[k]^n} \circ \mathbf{X}^n\right) \\
&= \hbar\left(\mathbf{L}^{[1]^n} \circ \mathbf{X}^n\right) + \cdots + \hbar\left(\mathbf{L}^{[M_T]^n} \circ \mathbf{X}^n\right) \\
&\quad + \hbar\left(\mathbf{L}^{[M_T+1]^n} \circ \mathbf{X}^n\right) && (41)\\
&\geq \hbar\left(\mathbf{L}^{[1]^n} \circ \mathbf{X}^n, \ldots, \mathbf{L}^{[M_T]^n} \circ \mathbf{X}^n\right) + \hbar\left(\mathbf{L}^{[M_T+1]^n} \circ \mathbf{X}^n\right) \\
&= \hbar\left(\{\mathbf{L}^{[1]^n}, \ldots, \mathbf{L}^{[M_T]^n}\} \circ \mathbf{X}^n\right) + \underbrace{\hbar\left(\mathbf{L}^{[M_T+1]^n} \circ \mathbf{X}^n\right)}_{\geq: 0} && (42)\\
&\geq: \hbar(\mathbf{X}^n) && (43)
\end{aligned}$$

where (43) follows from Case 1, i.e., the space spanned by the union of M_T generic vectors, $\{\mathbf{L}^{[1]}, \ldots, \mathbf{L}^{[M_T]}\}$, is the M_T-dimensional vector space $\mathbb{C}^{M_T}$. Note that the second term of (42) contains no less differential entropy than the noise therein and the differential entropy of noise normalized by $n \log \rho$ is non-negative.

Now we present the proof for the general setting of Lemma 3.

Proof: The collection of vectors and the splitting of the subspaces are consistent with the chain rule of entropy, so that the same direction of inequalities is obtained. In the end, we have collected L_Σ sets, each with M_T basis vectors and the projection of $\mathbf{X}^n$ to the space spanned by the vectors in each set would contribute entropy $\hbar(\mathbf{X}^n)$, which is guaranteed by Property (P1) of Lemma 2. Finally, the entropy of the discarded equations is no less than the entropy of the noise contained thus its normalization by $n \log \rho$ is non-negative. This completes the proof. ■

We illustrate this lemma with the following example.

Example: $M_T = 3$, $K = 6$, $(l_1, l_2, l_3, l_4, l_5, l_6) = (1, 2, 1, 1, 3, 2)$. We assume $n = 1$ and the subspaces are given by:

$$\mathbf{L}^{[1]} = \text{span}\{[1 \ \ 1 \ \ 1]^T\}, \tag{44a}$$
$$\mathbf{L}^{[2]} = \text{span}\{[0 \ \ 2 \ \ 3]^T, [0 \ \ 1 \ \ -1]^T\}, \tag{44b}$$
$$\mathbf{L}^{[3]} = \text{span}\{[1 \ \ -1 \ \ 0]^T\}, \tag{44c}$$
$$\mathbf{L}^{[4]} = \text{span}\{[1 \ \ 0 \ \ 1]^T\}, \tag{44d}$$
$$\mathbf{L}^{[5]} = \text{span}\{[1 \ \ -1 \ \ 3]^T, [1 \ \ 0 \ \ 0]^T, [0 \ \ 1 \ \ 0]^T\}, \tag{44e}$$
$$\mathbf{L}^{[6]} = \text{span}\{[0 \ \ 0 \ \ 1]^T, [1 \ \ 2 \ \ -4]^T\}. \tag{44f}$$

Note that $l^* = \sum_{k=1}^{6} l_k = 10$. It turns out that we can build 3 sets whose vector-elements collectively build full rank matrices. We start the proof by building the first set up to $\{\mathbf{L}^{[1]}, \mathbf{L}^{[2]}\}$. At this stage we have collected exactly $l_1 + l_2 = 3 = M_T$ independent vectors. So we terminate this set and start building the next set. Now we can go up to $\{\mathbf{L}^{[3]}, \mathbf{L}^{[4]}\}$ which contains $l_3 + l_4 = 2$ independent vectors, i.e., short of $M_T = 3$, but we cannot include $\mathbf{L}^{[5]}$ entirely because $l_3 + l_4 + l_5 = 5$ will exceed $M_T = 3$. So we will split $\mathbf{L}^{[5]}$ into a part, $\mathbf{L}_a^{[5]}$ that overlaps with $\{\mathbf{L}^{[3]}, \mathbf{L}^{[4]}\}$ and the remainder that does not overlap with $\{\mathbf{L}^{[3]}, \mathbf{L}^{[4]}\}$. Specifically,

$$|\mathbf{L}^{[5]}| = l_5 = 3, \tag{45}$$
$$|\{\mathbf{L}^{[3]}, \mathbf{L}^{[4]}\}| = l_3 + l_4 = 2, \tag{46}$$
$$M_T = 3, \tag{47}$$
$$|\mathbf{L}_a^{[5]}| = |\mathbf{L}^{[5]} \cap \{\mathbf{L}^{[3]}, \mathbf{L}^{[4]}\}| = 3 + 2 - M_T = 2, \tag{48}$$
$$|\mathbf{L}^{[5]} \setminus \mathbf{L}_a^{[5]}| = 3 - 2 = 1, \tag{49}$$
$$|\{\mathbf{L}^{[3]}, \mathbf{L}^{[4]}, \mathbf{L}^{[5]} \setminus \mathbf{L}_a^{[5]}\}| = M_T, \tag{50}$$
$$\mathbf{L}^{[5]} = \{\mathbf{L}_a^{[5]}, \mathbf{L}^{[5]} \setminus \mathbf{L}_a^{[5]}\}. \tag{51}$$

Thus, we obtain

$$\begin{aligned}
\mathbf{L}_a^{[5]} &= \text{span}\left\{\begin{bmatrix} 1 & 1 \\ -1 & 0 \\ 0 & 1 \end{bmatrix}\right\} \cap \text{span}\left\{\begin{bmatrix} 1 & 1 & 0 \\ -1 & 0 & 1 \\ 3 & 0 & 0 \end{bmatrix}\right\} \\
&= \text{span}\left\{\begin{bmatrix} 1 & 1 \\ -1 & 0 \\ 0 & 1 \end{bmatrix}\right\}, \\
\mathbf{L}^{[5]} \setminus \mathbf{L}_a^{[5]} &= \text{span}\{[1 \ \ 1 \ \ -1]^T\}.
\end{aligned}$$

The union of these three vectors spans $\mathbb{C}^3$, i.e., $\{\mathbf{L}^{[3]}, \mathbf{L}^{[4]}, \mathbf{L}^{[5]} \setminus \mathbf{L}_a^{[5]}\} = \mathbb{C}^3$. Thus, our second set becomes $\{\mathbf{L}^{[3]}, \mathbf{L}^{[4]}, \mathbf{L}^{[5]} \setminus \mathbf{L}_a^{[5]}\}$ which contains $M_T = 3$ linearly independent vectors. Finally, we start to build the third set starting with $\mathbf{L}_a^{[5]}$ and continuing on to $\mathbf{L}^{[6]}$. Again, since $|\mathbf{L}_a^{[5]}| < 3$ and $|\mathbf{L}_a^{[5]}| + l_6 > 3$, we need to split $\mathbf{L}^{[6]}$ into two parts, one $\mathbf{L}_a^{[6]}$ that overlaps with $\mathbf{L}_a^{[5]}$ and the other that does not overlap with $\mathbf{L}_a^{[5]}$. Therefore, our final set becomes $\{\mathbf{L}_a^{[5]}, \mathbf{L}^{[6]} \setminus \mathbf{L}_a^{[6]}\}$ with dimension $M_T = 3$, where

$\mathbf{L}_a^{[6]} = \mathbf{L}^{[6]} \cap \mathbf{L}_a^{[5]}$ has dimension $2 + 2 - M_T = 1$, and is given by

$$\begin{aligned}\mathbf{L}_a^{[6]} &= \text{span}\left\{\begin{bmatrix} 1 & 1 \\ -1 & 0 \\ 0 & 1 \end{bmatrix}\right\} \cap \text{span}\left\{\begin{bmatrix} 0 & 1 \\ 0 & 2 \\ 1 & -4 \end{bmatrix}\right\} \\ &= \text{span}\left\{\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}\right\}, \\ \mathbf{L}^{[6]} \setminus \mathbf{L}_a^{[6]} &= \text{span}\{[3 \quad 6 \quad -5]^T\}.\end{aligned}$$

This construction can be translated to the following information theoretical proof:

$$\sum_{k=1}^{6} \hbar(\mathbf{L}^{[k]} \circ \mathbf{X}) \tag{52}$$

$$\geq \hbar(\mathbf{L}^{[1]} \circ \mathbf{X}, \mathbf{L}^{[2]} \circ \mathbf{X}) + \sum_{k=3}^{6} \hbar(\mathbf{L}^{[k]} \circ \mathbf{X}) \tag{53}$$

$$=: \hbar(\mathbf{X}) + \sum_{k=3}^{6} \hbar(\mathbf{L}^{[k]} \circ \mathbf{X}) \tag{54}$$

$$\begin{aligned}&\geq \hbar(\mathbf{X}) + \hbar(\mathbf{L}^{[3]} \circ \mathbf{X}, \mathbf{L}^{[4]} \circ \mathbf{X}) + \hbar(\mathbf{L}^{[5]} \circ \mathbf{X}) + \hbar(\mathbf{L}^{[6]} \circ \mathbf{X}) \\ &= \hbar(\mathbf{X}) + \hbar(\mathbf{L}^{[3]} \circ \mathbf{X}, \mathbf{L}^{[4]} \circ \mathbf{X}) \\ &\quad + \hbar(\mathbf{L}_a^{[5]} \circ \mathbf{X}, (\mathbf{L}^{[5]} \setminus \mathbf{L}_a^{[5]}) \circ \mathbf{X}) + \hbar(\mathbf{L}^{[6]} \circ \mathbf{X}) \end{aligned} \tag{55}$$

$$\begin{aligned}&= \hbar(\mathbf{X}) + \hbar(\mathbf{L}^{[3]} \circ \mathbf{X}, \mathbf{L}^{[4]} \circ \mathbf{X}) + \hbar(\mathbf{L}_a^{[5]} \circ \mathbf{X}) \\ &\quad + \hbar((\mathbf{L}^{[5]} \setminus \mathbf{L}_a^{[5]}) \circ \mathbf{X} | \mathbf{L}_a^{[5]} \circ \mathbf{X}) + \hbar(\mathbf{L}^{[6]} \circ \mathbf{X}) \end{aligned} \tag{56}$$

$$\begin{aligned}&\geq \hbar(\mathbf{X}) + \hbar(\mathbf{L}^{[3]} \circ \mathbf{X}, \mathbf{L}^{[4]} \circ \mathbf{X}) + \hbar(\mathbf{L}_a^{[5]} \circ \mathbf{X}) \\ &\quad + \hbar((\mathbf{L}^{[5]} \setminus \mathbf{L}_a^{[5]}) \circ \mathbf{X} | \mathbf{L}_a^{[5]} \circ \mathbf{X}, \mathbf{L}^{[3]} \circ \mathbf{X}, \mathbf{L}^{[4]} \circ \mathbf{X}) \\ &\quad + \hbar(\mathbf{L}^{[6]} \circ \mathbf{X}) \end{aligned} \tag{57}$$

$$\begin{aligned}&= \hbar(\mathbf{X}) + \hbar(\mathbf{L}^{[3]} \circ \mathbf{X}, \mathbf{L}^{[4]} \circ \mathbf{X}) + \hbar(\mathbf{L}_a^{[5]} \circ \mathbf{X}) \\ &\quad + \hbar((\mathbf{L}^{[5]} \setminus \mathbf{L}_a^{[5]}) \circ \mathbf{X} | \mathbf{L}^{[3]} \circ \mathbf{X}, \mathbf{L}^{[4]} \circ \mathbf{X}) \\ &\quad + \hbar(\mathbf{L}^{[6]} \circ \mathbf{X}) \end{aligned} \tag{58}$$

$$\begin{aligned}&= \hbar(\mathbf{X}) + \hbar(\{\mathbf{L}^{[3]}, \mathbf{L}^{[4]}, \mathbf{L}^{[5]} \setminus \mathbf{L}_a^{[5]}\} \circ \mathbf{X}) + \hbar(\mathbf{L}_a^{[5]} \circ \mathbf{X}) \\ &\quad + \hbar(\mathbf{L}^{[6]} \circ \mathbf{X}) \end{aligned} \tag{59}$$

$$=: 2\hbar(\mathbf{X}) + \hbar(\mathbf{L}_a^{[5]} \circ \mathbf{X}) + \hbar(\mathbf{L}^{[6]} \circ \mathbf{X}) \tag{60}$$

$$= 2\hbar(\mathbf{X}) + \hbar(\mathbf{L}_a^{[5]} \circ \mathbf{X}) + \hbar(\mathbf{L}_a^{[6]} \circ \mathbf{X}, (\mathbf{L}^{[6]} \setminus \mathbf{L}_a^{[6]}) \circ \mathbf{X}) \tag{61}$$

$$\begin{aligned}&\geq 2\hbar(\mathbf{X}) + \hbar(\mathbf{L}_a^{[5]} \circ \mathbf{X}) \\ &\quad + \hbar((\mathbf{L}^{[6]} \setminus \mathbf{L}_a^{[6]}) \circ \mathbf{X} | \mathbf{L}_a^{[6]} \circ \mathbf{X}, \mathbf{L}_a^{[5]} \circ \mathbf{X}) + \hbar(\mathbf{L}_a^{[6]} \circ \mathbf{X}) \end{aligned} \tag{62}$$

$$\begin{aligned}&= 2\hbar(\mathbf{X}) + \hbar(\mathbf{L}_a^{[5]} \circ \mathbf{X}) \\ &\quad + \hbar((\mathbf{L}^{[6]} \setminus \mathbf{L}_a^{[6]}) \circ \mathbf{X} | \mathbf{L}_a^{[5]} \circ \mathbf{X}) + \hbar(\mathbf{L}_a^{[6]} \circ \mathbf{X}) \end{aligned} \tag{63}$$

$$= 2\hbar(\mathbf{X}) + \hbar(\{\mathbf{L}_a^{[5]}, \mathbf{L}^{[6]} \setminus \mathbf{L}_a^{[6]}\} \circ \mathbf{X}) + \hbar(\mathbf{L}_a^{[6]} \circ \mathbf{X}) \tag{64}$$

$$\geq: 3\hbar(\mathbf{X}) \tag{65}$$

where (57) follows from the property that adding conditioning terms does not increase the entropy; (58) is obtained because $\mathbf{L}_a^{[6]}$ is the intersection of $\mathbf{L}^{[5]}$ and $\{\mathbf{L}^{[3]}, \mathbf{L}^{[4]}\}$, thus it is also contained in $\{\mathbf{L}^{[3]}, \mathbf{L}^{[4]}\}$. This means $\mathbf{L}_a^{[5]} \circ \mathbf{X}$ can be reconstructed from $\{\mathbf{L}^{[3]} \circ \mathbf{X}, \mathbf{L}^{[4]} \circ \mathbf{X}\}$, within bounded noise distortion. In (54), (60) and (65), we use the fact that the construction produces each set with M_T independent vectors and the argument that their differential entropy is no less than $\hbar(\mathbf{X})$ follows from Case 1. Thus we have the desired result. Note that the derivations from (60) to (65) also follow from Property (P3) of Lemma 2. ■

When each subspace $\mathbf{L}^{[k]}(t)$ is generic, we have the following corollary.

Corollary 1: For generic subspaces $\mathbf{L}^{[k]}(t)$ of $\mathbb{C}^{M_T}$, we have

$$\sum_{k=1}^{K} \hbar\left(\mathbf{L}^{[k]^n} \circ \mathbf{X}^n\right) \geq: \lfloor \frac{l^*}{M_T} \rfloor \hbar(\mathbf{X}^n). \tag{66}$$

Proof: According to Lemma 3, with $l^* = \sum_{k=1}^{K} l_k$ generic vectors, we can build $\lfloor \frac{l^*}{M_T} \rfloor$ sets, each with M_T basis vectors. ■

In case that one may be also interested in Lemma 3 with conditional terms, we have the following corollary.

Corollary 2: For an arbitrary random variable Q, we have

$$\sum_{k=1}^{K} \hbar\left(\mathbf{L}^{[k]^n} \circ \mathbf{X}^n | Q\right) \geq: L_\Sigma \hbar(\mathbf{X}^n | Q). \tag{67}$$

Proof: The proof follows along the same lines as Lemma 3 and thus we omit it here. ■

IV. Four Ideas Comprising the Genie Chains Approach

To keep the presentation complete and as intuitive as possible, we need additional terminologies, particularly the notion of the *exposed subspace*, some notations for the *generic subspace* and the *interference subspace* available to a RX after decoding and removing the signal carrying its desired message. Here, we remind the reader that although they are called subspaces for convenience, in fact they represent the linear combinations of signals projected to those subspaces.

- **Exposed Subspace:** An exposed subspace, e.g., from TX 1 to RX 2, denoted as $\bar{\mathbf{X}}^{[1\sim 2]}$, refers to the linear combinations involving only $\mathbf{X}^{[1]}$ variables that are obtained at RX 2 after subtracting the signal carrying its desired message (for RX 2 this would be $\mathbf{X}^{[2]}$) and zero forcing (i.e., projecting into the null space, or simply using Gaussian elimination to remove) the other interference (in this case $\mathbf{X}^{[3]}, \mathbf{X}^{[4]}$). For example, consider the exposed subspace $\bar{\mathbf{X}}^{[1\sim 2]}$ in the $(M_T, M_R) = (2, 5)$ setting. At RX 2, after removing desired signal $\mathbf{X}^{[2]}$, we have 5 equations involving 6 variables $\mathbf{X}^{[1]}, \mathbf{X}^{[3]}, \mathbf{X}^{[4]}$ (Since $M_T = 2$, each $\mathbf{X}^{[k]}$ represents 2 variables). Eliminating 4 variables, $\mathbf{X}^{[3]}, \mathbf{X}^{[4]}$, leaves only one equation involving the two variables $\mathbf{X}^{[1]}$. This remaining linear combination, involving $\mathbf{X}^{[1]}$ only is the exposed subspace at RX 2 from TX 1. The dimensionality of the exposed space is indicated with a subscript, e.g., $\bar{\mathbf{X}}_1^{[1\sim 2]}$ in this example. As the AWGN terms are always presented in the received signal, the exposed subspace is always noisy. When we refer to the noise-free exposed subspace, we omit the bar notation on the top, e.g., $\mathbf{X}^{[1\sim 2]}$ is the noise-free version of $\bar{\mathbf{X}}^{[1\sim 2]}$.
- **Generic Subspace:** We use $\mathbf{X}_{(m)}^{[k]}$ to denote m generic linear combinations of the M variables in $\mathbf{X}^{[k]}$,

where the coefficients of the linear combinations are drawn from a continuous distribution. When the linear combinations are added with bounded variance independent noise, we denote them as $\bar{\mathbf{X}}^{[k]}_{(m)}$.

- **Interference Subspace:** We use the notation $\bar{\mathbf{S}}^{[k]}$ to refer to the received signal at RX k, after the *desired* variables $\mathbf{X}^{[k]}$ are set to zero. This is meaningful because the RX is always guaranteed to be able to reliably decode, and therefore subtract out, its desired signals, leaving it with a view of only the interference subspace from which it may attempt to resolve undesired signal dimensions. Similarly, the noise-free interference space is denoted as $\mathbf{S}^{[k]}$.

Note that the extension of these subspaces from the single-letter to the multi-letter form is immediate.

A. Ideas Illustrated Through the 4-User MIMO Interference Channel

The starting point of our outer bound is the common principle of providing a RX enough additional linear combinations of transmitted symbols to allow it to resolve all of the interferers, so that subject to noise distortion (which is inconsequential for DoF), it can decode all the messages. In general, because we are proving a converse, which means that we start with a reliable coding scheme, a RX is already guaranteed to reliably decode its desired message, which also allows the RX to subtract its desired symbols from its received signal. Now, the question remains whether the RX can decode *all* messages. For the $K = 4$ user MIMO interference channel, with M_R receive antennas, if $3M_T > M_R$, then we have fewer equations and more unknowns, so that resolution of interfering symbols is not guaranteed.[2] So, we provide $3M_T - M_R$ genie dimensions, i.e., $|\bar{\mathbf{G}}| = 3M_T - M_R$ linearly independent combinations of interference symbols where $\bar{\mathbf{G}}$ represents the genie symbols set. $\mathbf{G}$ denotes the noise-free version of $\bar{\mathbf{G}}$. This provides the RX enough equations to resolve all transmitted symbols. Equivalently, the undesired signal vectors $\mathbf{X}^{[i]},\ i \in \mathcal{K} \setminus \{k\}$ are now invertible (within noise distortion) from the RX's own observations combined with the genie dimensions. Since noise distortion is irrelevant for DoF arguments, the ability to resolve all symbols is equivalent to the ability to decode all symbols for DoF purposes. This forms the general basis for the outer bound, and is so far not a novel concept at all.

The challenging aspect, and where the novelty of our approach comes in, is to determine *which genie dimensions* to provide so that a useful DoF outer bound results. We propose a series of steps where we continue to cycle through various receivers in a chain of genie aided outer bounds containing entropies of various subspace equations introduced above, following four basic principles, that lead us to a cancelation of successive entropy terms, producing the desired outer bound. The four basic principles of the "genie chains" approach are highlighted next through simple examples.

[2]The receiver with M_R antennas sees M_R linear combination equations of the interfering transmit symbols. There are 3 interfering transmitters and each has M_T antennas. So we have $3M_T$ unknown variables in the M_R equations. When $3M_T > M_R$, we have fewer equations than unknowns.

Idea 1: Use the exposed space from one RX as a genie for the next.

Example 1: $(M_T, M_R) = (2, 5) \Rightarrow d \leq 10/7$

In this example, $|\bar{\mathbf{G}}| = 3M_T - M_R = 1$, so we need to provide a one-dimensional genie. Suppose we start with the generic subspace $\bar{\mathbf{G}}_1 = \bar{\mathbf{X}}^{[1]}_{(1)}$ and give it as genie to RX 2. Since this genie allows RX 2 to decode all the messages subject to the noise distortion, we have

$$
\begin{aligned}
&n(R_1 + R_2 + R_3 + R_4) - n\epsilon_n \\
&\quad \leq: I(W_1, W_2, W_3, W_4; \bar{\mathbf{Y}}^{[2]^n}, \bar{\mathbf{G}}_1^n) && (68)\\
&\quad = \hbar(\mathbf{Y}^{[2]^n}, \mathbf{X}^{[1]^n}_{(1)}) - \hbar(\mathbf{Y}^{[2]^n}, \mathbf{X}^{[1]^n}_{(1)} | W_1, W_2, W_3, W_4) \\
&\quad =: \hbar(\mathbf{Y}^{[2]^n}) + \hbar(\mathbf{X}^{[1]^n}_{(1)} | \mathbf{Y}^{[2]^n}, W_2) && (69)\\
&\quad = \hbar(\mathbf{Y}^{[2]^n}) + \hbar(\mathbf{X}^{[1]^n}_{(1)} | \mathbf{S}^{[2]^n}) && (70)\\
&\quad = \hbar(\mathbf{Y}^{[2]^n}) + \hbar(\mathbf{X}^{[1]^n}_{(1)} | [\mathbf{H}^{[23]^n}\ \mathbf{H}^{[24]^n}]^T \mathbf{S}^{[2]^n}, \cdots \\
&\qquad \cdots ([\mathbf{H}^{[23]^n}\ \mathbf{H}^{[24]^n}]^c)^T \mathbf{S}^{[2]^n}) && (71)\\
&\quad \leq \hbar(\mathbf{Y}^{[2]^n}) + \hbar(\mathbf{X}^{[1]^n}_{(1)} | ([\mathbf{H}^{[23]^n}\ \mathbf{H}^{[24]^n}]^c)^T \mathbf{S}^{[2]^n}) && (72)\\
&\quad \leq \hbar(\mathbf{Y}^{[2]^n}) + \hbar(\mathbf{X}^{[1]^n}_{(1)} | \mathbf{X}_1^{[1\sim 2]^n}) && (73)\\
&\quad = \hbar(\mathbf{Y}^{[2]^n}) + \hbar(\mathbf{X}^{[1]^n}_{(1)}, \mathbf{X}_1^{[1\sim 2]^n}) - \hbar(\mathbf{X}_1^{[1\sim 2]^n}) && (74)\\
&\quad \leq: 5n \log \rho + nR - \hbar(\mathbf{X}_1^{[1\sim 2]^n}). && (75)
\end{aligned}
$$

In the derivations above, (68) follows from Fano's inequality. (69) is obtained because from all the four messages, one can reconstruct the received signal vector and the genie signal subject to bounded noise distortion. Since RX 2 can decode its own message W_2, it can subtract the signal contributed by $\mathbf{X}^{[2]}$ from $\bar{\mathbf{Y}}^{[2]}$, and then produce the interference space $\bar{\mathbf{S}}^{[2]}$. Note that the subtracted part is a linear function of $\mathbf{X}^{[2]}$ and thus is independent of $\mathbf{X}^{[1]}_{(1)}$, as shown in (70). (71) follows from the fact that we can separate the space $\bar{\mathbf{S}}^{[2]}$ into two orthogonal subspaces, a one-dimensional projection that is orthogonal to the channels from TX 3 and TX 4 to RX 2, and the remaining four-dimensional subspace. Thus, RX 2 obtains the one-dimensional exposed subspace $\bar{\mathbf{X}}_1^{[1\sim 2]}$ in (73). Finally, the first term in (75) follows from Property (P2) of Lemma 2 and the second term in (75) is obtained because of the fact that we can use the two dimensional observations $\{\bar{\mathbf{X}}^{[1]^n}_{(1)}, \bar{\mathbf{X}}_1^{[1\sim 2]^n}\}$ to recover the transmitted signal vector $\mathbf{X}^{[1]}$ within noise distortion, thus contributing the term nR subject to noise distortion, as proved in Property (P3a) of Lemma 2. Because of the symmetry of the problem, there is no loss of generality in focusing on symmetric rates $R_1 = R_2 = R_3 = R_4 = R$, which allows us to re-write (75) as follows,

$$4nR - n\epsilon_n \leq: 5n \log \rho + nR - \hbar(\mathbf{X}_1^{[1\sim 2]^n}). \qquad (76)$$

Note that the exposed space has appeared as a negative entropy term. As a rule, in our approach, the negative entropy terms will become the genie signals for the subsequent bounds, leading to their eventual cancellation. Also, we will attempt to obtain a total of M_T useful bounds. In this case, $M_T = 2$, so we move to our final bound, and to the next RX, RX 3. The genie, as just mentioned, will be in the previous negative entropy term $\bar{\mathbf{G}}_2 = \bar{\mathbf{X}}_1^{[1\sim 2]}$. As $\bar{\mathbf{G}}_2$ is the exposed subspace at RX 2, which is independent of the channels associated

with RX 3, almost surely $\bar{\mathbf{G}}_2$ is independent of $\bar{\mathbf{Y}}_3$ such that RX 3 can now decode all the messages with this genie. The resulting bound is the following

$$\begin{aligned} n(R_1+R_2+R_3+R_4) - n\epsilon_n & \\ \leq: \; & I(W_1, W_2, W_3, W_4; \bar{\mathbf{Y}}^{[3]^n}, \bar{\mathbf{G}}_2^n) && (77)\\ \leq: \; & \hbar(\mathbf{Y}^{[3]^n}, \mathbf{X}_1^{[1\sim 2]^n}) && (78)\\ \leq \; & \hbar(\mathbf{Y}^{[3]^n}) + \hbar(\mathbf{X}_1^{[1\sim 2]^n}) && (79)\\ \Longrightarrow \; & 4nR - n\epsilon_n \leq: 5n\log\rho + \hbar(\mathbf{X}_1^{[1\sim 2]^n}) && (80) \end{aligned}$$

where (79) follows from chain rule and the fact that removing the conditional terms does not decrease the differential entropy.

Adding up the two inequalities (76) and (80), we obtain

$$\begin{aligned} 8nR - 2n\epsilon_n \; & \leq: \; 10n\log\rho + nR && (81)\\ \Longrightarrow 7nR \; & \leq \; 10n\log\rho + n\; o(\log\rho) + 2n\epsilon_n. && (82) \end{aligned}$$

By letting $n \to \infty$ first and then $\rho \to \infty$, we obtain the desired outer bound on DoF per user:

$$d \leq 10/7. \tag{83}$$

■

In order for the reader to have a more intuitive idea about the associated subspaces in each step, we provide a numerical example in the following.

`Initialization:` $M_T = 2$, $M_R = 5$, randomly generate 5×2 channel matrices from each TX to each RX. For example, we randomly generate the following associated channel realizations that are relevant in the proof:

$$\mathbf{H}^{[21]} = \begin{bmatrix} 0.5888 & -0.3927 \\ 1.0095 & -1.5730 \\ -0.4297 & -1.3400 \\ 0.3536 & 0.4674 \\ -1.4046 & 0.6240 \end{bmatrix},$$

$$\mathbf{H}^{[23]} = \begin{bmatrix} -2.4617 & 0.1171 \\ 1.9378 & 1.5657 \\ 0.8237 & 0.5253 \\ -0.8099 & 1.5186 \\ 0.4344 & -0.6581 \end{bmatrix},$$

$$\mathbf{H}^{[24]} = \begin{bmatrix} -0.5819 & -1.4890 \\ 0.2349 & 0.1483 \\ -0.0988 & 0.9539 \\ -0.1352 & 2.2932 \\ -1.8865 & -0.1452 \end{bmatrix},$$

$$\mathbf{H}^{[31]} = \begin{bmatrix} 0.0720 & -1.9399 \\ 0.7140 & 2.4346 \\ 1.2446 & 0.3470 \\ 0.4961 & -0.9756 \\ 0.5580 & 0.4654 \end{bmatrix},$$

$$\mathbf{H}^{[32]} = \begin{bmatrix} -0.0999 & -0.9784 \\ -0.2805 & -1.1571 \\ 0.4136 & -0.0548 \\ 0.2967 & 1.1387 \\ 1.1556 & 0.7722 \end{bmatrix},$$

$$\mathbf{H}^{[34]} = \begin{bmatrix} 0.6760 & 0.0171 \\ -0.8062 & -0.3684 \\ 0.0049 & -0.3526 \\ 0.8783 & 0.3086 \\ -0.9020 & 0.3290 \end{bmatrix}.$$

`Step 1:` We randomly generate a vector rand(2, 1) which captures the direction of the genie signal $\bar{\mathbf{G}}_1$. For example, the vector is $[0.6109 \;\; 0.0712]^T$ and thus $\bar{\mathbf{G}}_1 = 0.6109X_1^{[1]} + 0.0712X_2^{[1]} + Z_1$, where Z_1 is an independent noise with bounded variance. Then by zero forcing the interference from TX 3 and TX 4, RX 2 obtains a one-dimensional observation, say $\mathcal{O}_2$, of the transmitted signals from TX 1. That is,

$$\begin{aligned} \mathbf{X}_1^{[1\sim 2]} = \mathcal{O}_2 &= \mathbf{X}^{[1]^T}\mathbf{H}^{[21]^T}([\mathbf{H}^{[23]} \;\; \mathbf{H}^{[24]}])^c \\ &= 0.3227X_1^{[1]} + 1.2639X_2^{[1]}, \end{aligned}$$

which is linearly independent of $\mathbf{G}_1$ because the two linear combinations are not collinear.

`Step 2:` We provide $\bar{\mathbf{G}}_2 = \mathcal{O}_2 + Z_2 = 0.3227X_1^{[1]} + 1.2639X_2^{[1]} + Z_2$ as genie to RX 3, where Z_2 is another independent noise. Then by zero forcing the interference from TX 2 and TX 4, RX 3 also obtains a one-dimensional observation, say $\mathcal{O}_3$, of the transmitted signals from TX 1. That is,

$$\begin{aligned} \mathcal{O}_3 &= \mathbf{X}^{[1]^T}\mathbf{H}^{[31]^T}([\mathbf{H}^{[32]} \;\; \mathbf{H}^{[34]}])^c \\ &= 0.7366X_1^{[1]} + 1.0464X_2^{[1]}, \end{aligned}$$

which is also linearly independent of $\mathbf{G}_2$ so that we can provide $\bar{\mathbf{G}}_2$ as genie to ensure RX 3 can decode all the messages. ■

Remark: The $(M_T, M_R) = (2, 5)$ example is perhaps a bit serendipitous because the size of the exposed space exactly matches the required size of the genie at the next RX. In general, the two will not be the same, and we need to create either a bigger or a smaller genie. How to achieve a larger or smaller genie is the subject of the remaining three ideas.

Remark: Notice that for each sum rate inequality, we always start from Fano's inequality by providing enough dimensional genie signals $\bar{\mathbf{G}}$ to RX k such that it can decode all the messages subject to noise distortion. RX k can obtain the interference space $\bar{\mathbf{S}}^{[k]}$ by decoding W_k first and then subtracting out the signal carrying W_k from the observations $\bar{\mathbf{Y}}^{[k]}$. Thus, we will always omit the derivations from (68) to (70), and start directly from

$$KnR - n\epsilon_n \leq: \hbar(\mathbf{Y}^{[k]^n}) + \hbar(\mathbf{G}^n|\mathbf{S}^{[k]^n})$$

in the remainder of this paper.

Idea 2: Obtain a larger genie by exposing more dimensions.

Example 2: $(M_T, M_R) = (3, 7) \Rightarrow d \leq 21/10$.

In this example, $|\bar{\mathbf{G}}| = 3M_T - M_R = 2$ so we need a two-dimensional genie. However, $M_R - 2M_T = 1$, so the exposed space, e.g., $\bar{\mathbf{X}}^{[1\sim 2]}$ is only one-dimensional. Similar to Example 1, we start with a generic genie $\bar{\mathbf{G}}_1 = \bar{\mathbf{X}}_{(2)}^{[1]}$ at RX 2, which is linearly independent of $\bar{\mathbf{S}}^{[2]}$. Thus, RX 2 can

decode all the messages and we have

$$4nR - n\epsilon_n \leq: \hbar(\mathbf{Y}^{[2]^n}) + \hbar(\mathbf{G}_1^n|\mathbf{S}^{[2]^n}) \quad (84)$$
$$\leq: 7n\log\rho + \hbar(\mathbf{X}_{(2)}^{[1]^n}|\mathbf{S}^{[2]^n}) \quad (85)$$
$$\leq: 7n\log\rho + nR - \hbar(\mathbf{X}_1^{[1\sim2]^n}). \quad (86)$$

For the next bound, we move to RX 3. We will use the genie corresponding to the previous negative entropy term, $\bar{\mathbf{X}}_1^{[1\sim2]}$, but since this is only one-dimensional and we need 2 genie dimensions, we will complement it with a generic dimension from the next TX, $\bar{\mathbf{X}}_{(1)}^{[2]}$. That is, $\bar{\mathbf{G}}_2 = \{\bar{\mathbf{X}}_1^{[1\sim2]}, \bar{\mathbf{X}}_{(1)}^{[2]}\}$. The most important element here is how a new dimension gets exposed. RX 3 originally has one exposed dimension from TX 2. However, when the genie provides $\bar{\mathbf{X}}_1^{[1\sim2]}$, it exposes one additional dimension from TX 2, so that the new exposed space from TX 2 is denoted as $\bar{\mathbf{X}}_2^{[2\sim3]}$. The resulting bound is given by:

$$4nR - n\epsilon_n$$
$$\leq: \hbar(\mathbf{Y}^{[3]^n}) + \hbar(\mathbf{G}_2^n|\mathbf{S}^{[3]^n}) \quad (87)$$
$$\leq: 7n\log\rho + \hbar(\mathbf{X}_1^{[1\sim2]^n}, \mathbf{X}_{(1)}^{[2]^n}|\mathbf{S}^{[3]^n}) \quad (88)$$
$$= 7n\log\rho + \hbar(\mathbf{X}_1^{[1\sim2]^n}) + \hbar(\mathbf{X}_{(1)}^{[2]^n}|\mathbf{S}^{[3]^n}, \mathbf{X}_1^{[1\sim2]^n})$$
$$\leq: 7n\log\rho + \hbar(\mathbf{X}_1^{[1\sim2]^n}) + \hbar(\mathbf{X}_{(1)}^{[2]^n}|\mathbf{X}_2^{[2\sim3]^n}) \quad (89)$$
$$\leq: 7n\log\rho + \hbar(\mathbf{X}_1^{[1\sim2]^n}) + nR - \hbar(\mathbf{X}_2^{[2\sim3]^n}). \quad (90)$$

Now, with the additional exposed dimension, the exposed space $\bar{\mathbf{X}}_2^{[2\sim3]}$ is two-dimensional and matches the desired size of the genie. This gives us our third, and final, bound as we cyclically move on to the next RX, RX 4, with the genie $\bar{\mathbf{G}}_3 = \bar{\mathbf{X}}_2^{[2\sim3]}$. Since the channel coefficients associated with RX 4 are generic, the one-dimensional observation available at RX 4 from TX 2 is linearly independent of $\bar{\mathbf{G}}_3$ almost surely. Thus, RX 4 can decode all the messages subject to noise distortion, and we have

$$4nR - n\epsilon_n \leq: \hbar(\mathbf{Y}^{[4]^n}) + \hbar(\mathbf{G}_3^n|\mathbf{S}^{[4]^n}) \quad (91)$$
$$\leq: 7n\log\rho + \hbar(\mathbf{X}_2^{[2\sim3]^n}). \quad (92)$$

Adding up the inequalities (86), (90), (92), we have

$$12nR - 3n\epsilon_n \leq 21n\log\rho + 2nR + n\ o(\log\rho) \quad (93)$$

which produces the desired outer bound

$$d \leq 21/10. \quad (94)$$

■

In the following we provide an *alternative* proof for this example to shed light on the following idea.

Idea 3: Combine exposed subspaces from multiple receivers to create a larger genie.

After obtaining (86) at RX 2, similarly if a genie provides to RX 3 two random linear combinations of $\mathbf{X}^{[1]}$, i.e., $\bar{\mathbf{G}}_2' = \bar{\mathbf{X}}_{(2)}^{[1]'}$, we have another inequality at RX 3

$$4nR - n\epsilon_n \leq: 7n\log\rho + nR - \hbar(\mathbf{X}_1^{[1\sim3]^n}) \quad (95)$$

where $\bar{\mathbf{X}}_1^{[1\sim3]}$ is the exposed one dimensional observation available at RX 3 projecting from TX 1.

Finally, a genie provides $\bar{\mathbf{G}}_3' = \{\bar{\mathbf{X}}_1^{[1\sim2]}, \bar{\mathbf{X}}_1^{[1\sim3]}\}$ to RX 4 where $\bar{\mathbf{G}}_3'$ is linearly independent of the 7-dimensional $\mathbf{S}^{[4]^n}$ space. Thus, RX 4 can decode all the messages as well. So we have

$$4nR - n\epsilon_n \leq: \hbar(\mathbf{Y}^{[4]^n}) + \hbar(\mathbf{G}_3^{n'}|\mathbf{S}^{[4]^n}) \quad (96)$$
$$\leq: 7n\log\rho + \hbar(\mathbf{X}_1^{[1\sim2]^n}) + \hbar(\mathbf{X}_1^{[1\sim3]^n}). \quad (97)$$

Adding (86), (95) and (97), we again obtain the desired outer bound

$$12nR - 3n\epsilon_n \leq: 21n\log\rho + 2nR \Longrightarrow d \leq 21/10. \quad (98)$$

■

Remark: Idea 3 is especially useful in the $M_T > M_R$ settings. This is because here the dimension of the exposed subspace at the receiver is smaller than the number of transmit antennas, so that we may need to combine several exposed subspaces to form a proper genie.

Idea 4: Obtain a smaller size genie by intersections.

Example 3: $(M_T, M_R) = (3, 8) \Rightarrow d \leq 24/11$

Starting with a generic genie $\bar{\mathbf{G}}_1 = \bar{\mathbf{X}}_{(1)}^{[1]}$ at RX 2, we have the first inequality

$$4nR - n\epsilon_n \leq: \hbar(\mathbf{Y}^{[2]^n}) + \hbar(\mathbf{G}_1^n|\mathbf{S}^{[2]^n}) \quad (99)$$
$$\leq: 8n\log\rho + \hbar(\mathbf{X}_{(1)}^{[1]^n}|\mathbf{S}^{[2]^n}) \quad (100)$$
$$\leq: 8n\log\rho + nR - \hbar(\mathbf{X}_2^{[1\sim2]^n}). \quad (101)$$

Now, the required size of the genie is $|\bar{\mathbf{G}}| = 3M - N = 1$ while exposed spaces have size 2. How to create a smaller genie? We will do that by creating multiple exposed spaces, each of which may be too big to be an acceptable genie, but their intersection will turn out to be an acceptable genie. A genie provides to RX 3 another random linear combination of $\mathbf{X}^{[1]}$, i.e., $\bar{\mathbf{G}}_2 = \bar{\mathbf{X}}_{(1)}^{[1]'}$, so that

$$4nR - n\epsilon_n \leq: 8n\log\rho + nR - \hbar(\mathbf{X}_2^{[1\sim3]^n}) \quad (102)$$

where $\mathcal{O}_3 = \mathbf{X}_2^{[1\sim3]}$ is the two-dimensional exposed space of TX 1 at RX 3. Since the construction of $\mathcal{O}_2 = \mathbf{X}_2^{[1\sim2]}$ and $\mathcal{O}_3$ only involve the channel coefficients associated with their own receivers, they are generic and have $2 + 2 - 3 = 1$ dimensional intersection, denoted as $\mathcal{I} = \mathcal{O}_2 \cap \mathcal{O}_3$. Thus, we can rewrite (102) as

$$4nR - n\epsilon_n$$
$$\leq: 8n\log\rho + nR - \hbar(\mathbf{X}_2^{[1\sim3]^n}) - \hbar(\mathbf{X}_2^{[1\sim2]^n}) + \hbar(\mathbf{X}_2^{[1\sim2]^n}) \quad (103)$$
$$= 8n\log\rho + \hbar(\mathbf{X}_2^{[1\sim2]^n}) + nR - \hbar(\mathcal{O}_3^n \setminus \mathcal{I}^n, \mathcal{I}^n) - \hbar(\mathcal{O}_2^n) \quad (104)$$
$$= 8n\log\rho + \hbar(\mathbf{X}_2^{[1\sim2]^n}) + nR - \hbar(\mathcal{I}^n) - \hbar(\mathcal{O}_3^n \setminus \mathcal{I}^n|\mathcal{I}^n) - \hbar(\mathcal{O}_2^n) \quad (105)$$
$$\leq 8n\log\rho + \hbar(\mathbf{X}_2^{[1\sim2]^n}) + nR - \hbar(\mathcal{I}^n) - \hbar(\mathcal{O}_3^n \setminus \mathcal{I}^n|\mathcal{I}^n, \mathcal{O}_2^n) - \hbar(\mathcal{O}_2^n) \quad (106)$$
$$=: 8n\log\rho + \hbar(\mathbf{X}_2^{[1\sim2]^n}) + nR - \hbar(\mathcal{I}^n) - \hbar(\mathcal{O}_2^n) - \hbar(\mathcal{O}_3^n \setminus \mathcal{I}^n|\mathcal{O}_2^n) \quad (107)$$
$$= 8n\log\rho + \hbar(\mathbf{X}_2^{[1\sim2]^n}) + nR - \hbar(\mathcal{I}^n) - \hbar(\mathcal{O}_2^n, \mathcal{O}_3^n \setminus \mathcal{I}^n)$$
$$=: 8n\log\rho + \hbar(\mathbf{X}_2^{[1\sim2]^n}) - \hbar(\mathcal{I}^n) \quad (108)$$

where (107) is obtained because $\mathcal{I}$ is included in $\mathcal{O}_2$, and (108) follows from the property that $\{\mathcal{O}_2, \mathcal{O}_3 \setminus \mathcal{I}\}$ are three linear independent equations in $\mathbf{X}^{[1]}$ and we can use Property (P3) of Lemma 2. We call the inequality (108) the "intermediate bound" which is constructed by intersecting two subspaces at different receivers. The derivations above are the same as that in Lemma 3.

Finally, we should provide the observations we obtain in the last step as the genie to RX 4, i.e., $\bar{\mathbf{G}}_3 = \mathcal{I} + Z$, where Z is an independent noise. Since $\bar{\mathbf{G}}_3$ only involves the channel coefficients associated with RX 2 and 3, it is linearly independent of the original two dimensional observations from TX 1 at RX 4. Thus, RX 4 can decode all the messages, and we have the last inequality

$$4nR - n\epsilon_n \leq: \hbar(\mathbf{Y}^{[4]^n}) + \hbar(\mathbf{G}_3^n|\mathbf{S}^{[4]^n}) \tag{109}$$

$$\leq 8n\log\rho + \hbar(\mathcal{I}^n). \tag{110}$$

Adding up the inequalities (101), (108) and (110), we have

$$12nR - 3n\epsilon_n \leq: 3NnR + nR \tag{111}$$

$$\Longrightarrow d \leq 3N/11 = 24/11. \tag{112}$$

■

The three examples above show that our goal is to use a chain of arguments, where we start with the exposed spaces and continue to build new genies with more dimensions by peeling off overlaps, or less dimensions by taking intersections, until we have the genie of the correct size, which requires exactly M_T steps, and produces the bound $d \leq \frac{M_T M_R}{M_T+M_R}$, if all genies in this process are acceptable, i.e., linearly independent of the space already available to the receivers.

This is summarized in the following observation in the context of K-user interference channel.

Observation 1: For the K user $M_T \times M_R$ MIMO interference channel where each TX has M_T and each RX has M_R antennas, if we can create a genie chain with M_T genie signal sets and each genie signal (with the appropriate size of $(K-1)M_T - M_R$) is linearly independent of the exposed subspace at each corresponding receiver, then the DoF value per user is given by $d = \frac{M_T M_R}{M_T+M_R}$.

Note that the genie chain technique is applicable to arbitrary channel realizations, to the extent that the genie signals remain linearly independent of previously exposed spaces. Thus this technique can be used to test arbitrary settings, although in this paper, we focus exclusively on deriving DoF results that hold almost surely for generic channels.

Now we have a general result for the K-user MIMO interference channel, then building genie signals with appropriate sizes and testing the linear independence condition are all that remain. This is the problem that we will address for various cases, and leave open for others.

Also, it should be noted that when the genie signals start becoming linearly dependent, one can terminate the chain by simply replacing the entropy term by its maximum signal dimension. The bound may be loose but it is still the best bound we can get through the genie chain approach, and likely better than any other existing approach.

V. Application: $K = 4$ User MIMO Interference Channel

In this section, we apply the genie chains approach to investigate the DoF characterization for the $K = 4$ user $M_T \times M_R$ MIMO interference channel. For brevity, let $M = \min(M_T, M_R)$ and $N = \max(M_T, M_R)$. Since the DoF results and corresponding proofs when $M/N \leq 3/8$ follow from the $K = 3$ user case [2] but require much more complicated analysis, we will consider this regime later in Section VII.1. In this section, we only consider the setting $M_T/M_R > 3/8$. The main result for this regime is presented in the theorem.

Theorem 1: For the $K = 4$ user $M_T \times M_R$ MIMO interference channel where each TX has M_T and each RX has M_R antennas, the DoF value per user is given by $d = \frac{M_T M_R}{M_T+M_R}$ for every (M_T, M_R) where $\frac{M_T}{M_R} \in \mathcal{P}_1 \cup \mathcal{P}_2 \cup \mathcal{P}_3 \triangleq \mathcal{P} \subset (3/8, 1]$, $\mathcal{P}_1 = \{\frac{M_T}{M_R} | \frac{1}{2} \leq \frac{M_T}{M_R} < 1, M_T, M_R \in \mathbb{Z}^+, M_R \leq 20\}$, $\mathcal{P}_2 = \{\frac{M_T}{M_R} | \frac{2}{5} \leq \frac{M_T}{M_R} < \frac{1}{2}, M_T, M_R \in \mathbb{Z}^+, M_R \leq 100\}$ and $\mathcal{P}_3 = \{\frac{8}{21}\} \cup \{\frac{M_T}{M_R} = \frac{2c-1}{5c-2} | c \in \mathbb{Z}^+, c \geq 2, M_T, M_R \in \mathbb{Z}^+, M_R \leq 100\}$.

As reported in [7], $\frac{MN}{M+N}$ DoF per user are achievable almost surely by using the rational alignment framework. To establish the DoF result implied by Theorem 1, it suffices to show that $d = \frac{MN}{M+N}$ is also the information theoretic DoF outer bound per user. In the remainder of this section, we will use the systematic "genie chains" approach, based on four central ideas that we show in Section IV. We will provide proofs through specific algorithms for M_T/M_R belonging to $\mathcal{P}_1, \mathcal{P}_2, \mathcal{P}_3$ sequentially. In addition, note that we only consider the $M_T < M_R$ setting in this section, which means that $(M_T, M_R) = (M, N)$. Further discussion on the results will be presented in Section VII.

Remark: Note that in the three regimes $\mathcal{P}_1, \mathcal{P}_3, \mathcal{P}_3$ in Theorem 1, although we impose the constraints $M_R \leq 20$, $M_R \leq 100$ and $M_R \leq 100$, respectively, it does not mean that the DoF converse proofs do not hold without these constraints. The reason of imposing constraints in $\mathcal{P}_1$ is the difficulty of computing the determinant of certain matrices whose elements are generated randomly to determine those matrices have full rank or not, and when those matrices become large as the MIMO dimensions M_T, M_R increase, we face computational bottlenecks. The reason of imposing constraints in $\mathcal{P}_2$ and $\mathcal{P}_3$ is that although we get rid of numerical uncertainty by choosing specific channel matrices (so we are now computing the ranks of deterministic matrices whose elements are explicitly specified instead of the ranks of random matrices whose elements are generated randomly as in $\mathcal{P}_1$), it is still very challenging, if not impossible, to derive closed-form expressions of the ranks of certain matrices. Therefore, the constraints in Theorem 1 mean that we have tested the cases that satisfy the constraints and we face the issues mentioned above for cases that go beyond the constraints.

A. $M/N \in [1/2, 1)$ Case

Since $N - 2M \leq 0$, each RX cannot directly obtain exposed subspaces from any interferer by zero forcing the signals from

the other two interferers. For brevity we let $M_0 = N - M$ where M_0 is a positive integer. Also, note that the random linear combinations provided by a genie in each step are generic, although the notations may be the same.

Proof: The general proof for this setting is given by the following algorithm.

`Algorithm 1` ($M/N \in [1/2, 1)$)

- `Step 1:`
  Start from RX $k = 2$. A genie provides RX k the signal set $\bar{\mathbf{G}} = \{\bar{\mathbf{X}}^{[k+1]}, \bar{\mathbf{X}}^{[k-1]}_{(M-M_0)}\}$, where $\mathbf{X}^{[k-1]}_{(M-M_0)}$ are $M - M_0$ random linear combinations of the transmit signals from TX $k-1$. In the absence of interference from TX $k+1$, RX k has M_0 dimensional observations of the transmit signals from TX $k-1$ by zero forcing the signals from TX $k+2$. We denote by $\mathcal{O}$ the M_0 dimensional observations. This process produces the first sum rate inequality:

$$\begin{aligned} 4nR - n\epsilon_n &\leq: \hbar(\mathbf{Y}^{[k]^n}) + \hbar(\mathbf{G}^n|\mathbf{S}^{[k]^n}) \\ &\leq: Nn\log\rho + \hbar(\mathbf{X}^{[k+1]^n}) \\ &\quad + \hbar(\mathbf{X}^{[k-1]^n}_{(M-M_0)}|\mathbf{S}^{[k]}, \mathbf{X}^{[k+1]^n}) \\ &\leq: Nn\log\rho + nR + \hbar(\mathbf{X}^{[k-1]^n}_{(M-M_0)}|\mathcal{O}^n) \\ &\leq: Nn\log\rho + 2nR - \hbar(\mathcal{O}^n). \end{aligned}$$

- `Step 2:`
  If $|\mathcal{O}| = |\mathbf{G}| - M = M - M_0$, go to Step 3.
  If $|\mathcal{O}| < |\mathbf{G}| - M = M - M_0$, go to Step 4.
  If $|\mathcal{O}| > |\mathbf{G}| - M = M - M_0$, go to Step 5.
- `Step 3:`
  A genie provides RX $k+1$ the set $\bar{\mathbf{G}} = \{\bar{\mathbf{X}}^{[k+2]}, \mathcal{O}+\mathbf{Z}\}$. Note that in the absence of interference from TX $k+2$, RX $k+1$ originally has M_0 dimensional observations of TX $k-1$ after zero forcing the interference from TX k, which is denoted as $\mathcal{O}'$. As $\mathcal{O}$ and $\mathcal{O}'$ are observations at different receivers, they are independent almost surely. So from $\{\mathcal{O}, \mathcal{O}'\}$, RX $k+1$ is able to recover the transmit signal from TX $k-1$ subject to noise distortion. This process produces one sum DoF inequality

$$\begin{aligned} 4nR - n\epsilon_n &\leq: \hbar(\mathbf{Y}^{[k+1]^n}) + \hbar(\mathbf{G}^n|\mathbf{S}^{[k+1]^n}) \\ &\leq: Nn\log\rho + \hbar(\mathbf{X}^{[k+2]^n}, \mathcal{O}^n|\mathbf{S}^{[k+1]^n}) \\ &\leq: Nn\log\rho + nR + \hbar(\mathcal{O}^n). \end{aligned}$$

  Adding up all inequalities that we have so far produces the inequality

$$4MnR - Mn\epsilon_n \leq: MNnR + (3M - N)nR$$

  which leads to our desired DoF outer bound $d \leq \frac{MN}{M+N}$. Then we stop.
- `Step 4:`
  A genie provides $\bar{\mathbf{G}} = \{\bar{\mathbf{X}}^{[k+2]}, \mathcal{O} + \mathbf{Z}, \bar{\mathbf{X}}^{[k]}_{(M-M_0-|\mathcal{O}|)}\}$ to RX $k+1$. In the absence of interference from TX $k+2$, RX $k+1$ originally has M_0 dimensional observations of the transmit signals from TX k after zero forcing the interference from TX $k-1$, which is denoted as $\mathcal{O}'$. Since providing $\mathcal{O}$ to RX $k+1$ releases another M_0 dimensional observations of the transmit signals from TX k, which is denoted as $\tilde{\mathcal{O}}$, RX $k+1$ has a total of $|\tilde{\mathcal{O}}| + M_0$ dimensional observations of $\mathbf{X}^{[k]^n}$. This process produces the sum rate inequality

$$\begin{aligned} &4nR - n\epsilon_n \\ &\leq: \hbar(\mathbf{Y}^{[k+1]^n}) + \hbar(\mathbf{G}^n|\mathbf{S}^{[k+1]^n}) \\ &\leq: Nn\log\rho + \hbar(\mathbf{X}^{[k+2]^n}, \mathcal{O}^n, \mathbf{X}^{[k]^n}_{(M-M_0-|\mathcal{O}|)}|\mathbf{S}^{[k+1]^n}) \\ &\leq: Nn\log\rho + \hbar(\mathbf{X}^{[k+2]^n}) + \hbar(\mathcal{O}^n) \\ &\quad + \hbar(\mathbf{X}^{[k]^n}_{(M-M_0-|\mathcal{O}|)}|\tilde{\mathcal{O}}^n, \mathcal{O}'^n) \\ &\leq: Nn\log\rho + nR + \hbar(\mathcal{O}^n) + nR - \hbar(\tilde{\mathcal{O}}^n, \mathcal{O}'^n). \end{aligned}$$

  Now we update $\mathcal{O} = \{\tilde{\mathcal{O}}, \mathcal{O}'\}$ and $k = k+1$. Go back to Step 2.
- `Step 5:`
  A genie provides $\bar{\mathbf{G}} = \{\mathcal{O} + \mathbf{Z}, \bar{\mathbf{X}}^{[k]}_{(2M-M_0-|\mathcal{O}|)}\}$ to RX $k+1$. In the N dimensional observation $\mathbf{S}^{[k+1]^n}$, after zero forcing the interference from TX $k-2$, we still have $N - M = M_0$ observations of the interference from TX $k-1$ and TX k. After providing $\mathcal{O}$ to RX $k+1$, now we have a total of $M_0 + |\mathcal{O}| > M$ dimensions of the interference from TX $k-1$ and TX k. Therefore, we continue to zero force the interference from TX $k-1$, thus only leaving $M_0 + |\mathcal{O}| - M$ dimensional observations of $\mathbf{X}^{[k]}$, denoted as $\mathcal{O}'$. Note that $\mathcal{O}'$ is linearly independent of $\bar{\mathbf{X}}^{[k]^n}_{(2M-M_0-|\mathcal{O}|)}$ provided by the genie, and from them together RX $k+1$ is able to recover the transmit signal from TX k subject to noise distortion. This process produces the inequality

$$\begin{aligned} 4nR - n\epsilon_n &\leq: \hbar(\mathbf{Y}^{[k+1]^n}) + \hbar(\mathbf{G}^n|\mathbf{S}^{[k+1]^n}) \\ &\leq: Nn\log\rho \\ &\quad + \hbar(\mathcal{O}^n, \mathbf{X}^{[k]^n}_{(2M-M_0-|\mathcal{O}|)}|\mathbf{S}^{[k+1]^n}) \\ &\leq: Nn\log\rho + \hbar(\mathcal{O}^n) + nR - \hbar(\mathcal{O}'^n). \end{aligned}$$

  Now we update $\mathcal{O} = \mathcal{O}'$ and $k = k+1$. Go back to Step 2. ■

Remark: Note that we do not need the intermediate DoF outer bound in this case. In contrast, for $M/N \in [2/5, 1/2)$ or $[3/8, 2/5)$ cases the intermediate bound is necessary, as we have shown in Example 3 in Section IV.

In this algorithm, the genie signal $\bar{\mathbf{G}}$ always contains $3M - N$ dimensions in each step. If we want to recover all the interference symbols from $\{\bar{\mathbf{G}}, \bar{\mathbf{S}}^{[k]}\}$ subject to noise distortion, it suffices to show that the $\bar{\mathbf{G}}$ is linearly independent of $\bar{\mathbf{S}}^{[k]}$, which is a linear algebra problem now. We are able to verify the linear independence through numerical tests when $M_R \leq 20$. This completes the proof for this regime.

B. $M/N \in$ [2/5,1/2) Case

Since $N - 2M > 0$, each RX obtains a fixed $N - 2M$ dimensional clean observations from each interferer, by simply zero forcing the signals from the other two interferers. For brevity we let $M_0 = N - 2M$ where M_0 is a positive integer.

Proof: The general proof for this setting is given by the following algorithm.

`Algorithm 2` ($M/N \in [2/5, 1/2)$)

- `Step 1:`
  Start from RX $k = 2$. A genie provides signals $\bar{\mathbf{G}} = \bar{\mathbf{X}}^{[k-1]}_{(M-M_0)}$ to RX k which originally has M_0 dimensional observations of transmit signals from TX $k-1$, after it zero forces the interference of the other two users. We denote by $\mathcal{O}$ these M_0 dimensional observations. This process produces the first sum rate inequality

$$\begin{aligned} 4nR - n\epsilon_n &\leq: \hbar(\mathbf{Y}^{[k]^n}) + \hbar(\mathbf{G}^n|\mathbf{S}^{[k]^n}) \\ &\leq: Nn\log\rho + \hbar(\mathbf{X}^{[k-1]^n}_{(M-M_0)}|\mathcal{O}^n) \\ &\leq: Nn\log\rho + nR - \hbar(\mathcal{O}^n). \end{aligned}$$

- `Step 2:`
  If $|\mathcal{O}| = |\mathbf{G}| = M - M_0$, go to Step 3.
  If $|\mathcal{O}| < |\mathbf{G}| = M - M_0$, go to Step 4.
  If $|\mathcal{O}| > |\mathbf{G}| = M - M_0$, go to Step 5.
- `Step 3:`
  A genie provides $\bar{\mathbf{G}} = \mathcal{O} + \mathbf{Z}$ to RX $k+1$. RX $k+1$ originally has M_0 dimensional observations of the transmit signals from TX k after it zero forces the interference from the other two users. We denote by $\mathcal{O}'$ these M_0 dimensional observations. Combined with the $|\mathcal{O}|$ dimensional observations of transmit signals from TX k opened up by $\mathcal{O}$, they allow RX $k+1$ to recover signals from TX k subject to noise distortion. This process produces the inequality

$$\begin{aligned} 4nR - n\epsilon_n &\leq: \hbar(\mathbf{Y}^{[k+1]^n}) + \hbar(\mathbf{G}^n|\mathbf{S}^{[k+1]^n}) \\ &\leq: Nn\log\rho + \hbar(\mathcal{O}^n). \end{aligned}$$

  Adding up the inequalities we have so far produces the inequality

$$4MnR - Mn\epsilon_n \leq: MNnR + (3M - N)nR$$

  which leads to our desired DoF outer bound $d \leq \frac{MN}{M+N}$. Then we stop.
- `Step 4:`
  A genie provides the set $\bar{\mathbf{G}} = \{\mathcal{O} + \mathbf{Z}, \bar{\mathbf{X}}^{[k]}_{(M-M_0-|\mathcal{O}|)}\}$ to RX $k+1$. RX $k+1$ originally has M_0 dimensional observations of transmit signals from TX k after zero forcing the interference from the other two suers. Denote these M_0 dimensional observations as $\mathcal{O}'$. Since providing $\mathcal{O}$ to RX $k+1$ releases another M_0 observations of transmit signals from TX k, denoted as $\tilde{\mathcal{O}}$, RX $k+1$ has a total of $|\tilde{\mathcal{O}}| + M_0$ dimensional observations of $\mathbf{X}^{[k]}$. This process produces the inequality

$$\begin{aligned} &4nR - n\epsilon_n \\ &\leq: \hbar(\mathbf{Y}^{[k+1]^n}) + \hbar(\mathbf{G}^n|\mathbf{S}^{[k+1]^n}) \\ &\leq: Nn\log\rho + \hbar(\mathcal{O}^n, \mathbf{X}^{[k]^n}_{(M-M_0-|\mathcal{O}|)}|\mathbf{S}^{[k+1]^n}) \\ &\leq: Nn\log\rho + \hbar(\mathcal{O}^n) + \hbar(\mathbf{X}^{[k]^n}_{(M-M_0-|\mathcal{O}|)}|\tilde{\mathcal{O}}^n, \mathcal{O}'^n) \\ &\leq: Nn\log\rho + \hbar(\mathcal{O}^n) + nR - \hbar(\tilde{\mathcal{O}}^n, \mathcal{O}'^n). \end{aligned}$$

  Now we update $\mathcal{O} = \{\tilde{\mathcal{O}}, \mathcal{O}'\}$ and $k = k+1$. Go back to Step 2.
- `Step 5:`
  A genie provides $\bar{\mathbf{G}} = \bar{\mathbf{X}}^{[k-1]}_{(M-M_0)}$ to RX $k+1$, which originally has M_0 dimensional observations of transmit signals from TX $k-1$ after zero forcing the interference from the other two users. Denote these M_0 dimensional observations as $\mathcal{O}'$. Because $|\mathcal{O}| + |\mathcal{O}'| > M$, the subspaces $\mathcal{O}$ and $\mathcal{O}'$ have an $|\mathcal{O}| + M_0 - M$ dimensional intersection. We denote this intersection by $\mathcal{I}$. Note that $\{\mathcal{O}, \mathcal{O}' \setminus \mathcal{I}\}$ is already the whole M dimensional space, thus contributing $R + o(\log\rho)$ differential entropy. Note that we still have the remaining $|\mathcal{O}| + M_0 - M$ dimensional observations of $\mathbf{X}^{[k-1]}$, i.e., the intersection $\mathcal{I}$. This process produces the intermediate bound

$$\begin{aligned} &4nR - n\epsilon_n \\ &\leq: \hbar(\mathbf{Y}^{[k+1]^n}) + \hbar(\mathbf{G}^n|\mathbf{S}^{[k+1]^n}) \\ &\leq: Nn\log\rho + \hbar(\mathbf{X}^{[k-1]^n}_{(M-M_0)}|\mathcal{O}'^n) \\ &\leq: Nn\log\rho + nR - \hbar(\mathcal{O}'^n) - \hbar(\mathcal{O}^n) + \hbar(\mathcal{O}^n) \\ &\leq: Nn\log\rho + \hbar(\mathcal{O}^n) + nR \\ &\quad - \hbar(\mathcal{O}'^n \setminus \mathcal{I}^n, \mathcal{I}^n) - \hbar(\mathcal{O}^n) \\ &\leq: Nn\log\rho + \hbar(\mathcal{O}^n) - \hbar(\mathcal{I}^n) + nR \\ &\quad - \hbar(\mathcal{O}'^n \setminus \mathcal{I}^n|\mathcal{I}^n) - \hbar(\mathcal{O}^n) \\ &\leq: Nn\log\rho + \hbar(\mathcal{O}^n) - \hbar(\mathcal{I}^n) + nR \\ &\quad - \hbar(\mathcal{O}'^n \setminus \mathcal{I}^n|\mathcal{I}^n, \mathcal{O}^n) - \hbar(\mathcal{O}^n) \\ &\leq: Nn\log\rho + \hbar(\mathcal{O}^n) - \hbar(\mathcal{I}^n) + nR \\ &\quad - \hbar(\mathcal{O}'^n \setminus \mathcal{I}^n|\mathcal{O}^n) - \hbar(\mathcal{O}^n) \\ &\leq: Nn\log\rho + \hbar(\mathcal{O}^n) - \hbar(\mathcal{I}^n) + nR \\ &\quad - \hbar(\mathcal{O}'^n \setminus \mathcal{I}^n, \mathcal{O}^n) \\ &\leq: Nn\log\rho + \hbar(\mathcal{O}^n) - \hbar(\mathcal{I}^n). \end{aligned}$$

  Now we update $\mathcal{O} = \mathcal{I}$, $k = k+1$. Go back to Step 2. ■

Remark: The phrase "intermediate" implies that the observations associated with the negative entropy term are obtained by intersecting the two subspaces available at successive two receivers looking at the same interferer. Moreover, in this case we only need intermediate bounds (with respect to two receivers) by intersecting at most once.

Similar to Algorithm 1, in this algorithm, we also need to show that the $3M - N$ dimensional genie signal $\bar{\mathbf{G}}$ is linearly independent of $\bar{\mathbf{S}}^{[k]}$. Equivalently, since each RX has a clean $N - 2M$ dimensional subspace of each interferer, we need to guarantee that $\bar{\mathbf{G}}$ is linearly independent of these subspaces. The detailed proof of this claim is deferred to Appendix C-A.

C. $M/N \in [3/8, 2/5)$ Case

When M/N falls into $[3/8, 2/5)$ regime, we show the proofs when $M/N = (2c-1)/(5c-2)$, $c \in \mathbb{Z}^+ \setminus \{1\}$ and $M/N = 8/21$.

Proof: Let us first consider $M/N = (2c-1)/(5c-2)$, $c \in \mathbb{Z}^+ \setminus \{1\}$ cases. It can be checked that for these cases we only need *one successive intermediate bound*. Thus, we can still use Algorithm 2 to produce the information theoretic outer bound proofs for these cases. Similarly, what remains to be

shown is that at each step the clean $N-2M$ dimensional observations of the associated TX are linearly independent of the $|\bar{\mathbf{G}}| = 3M - N$ dimensional observations of that TX opened up by the provided genie $\bar{\mathbf{G}}$, i.e., we will show that the resulting $M \times M$ square matrix has full rank. The proof in detail is deferred to Appendix C-B.

Next, we prove the case where $M/N = 8/21$. Note that this does not fall into the category that $M/N = (2c-1)/(5c-2)$. What is special for this case is that we need *two successive intermediate bounds*. Suppose $(M, N) = (8a, 21a),\ a \in \mathbb{Z}^+$, then the proof is shown through the following eight steps.

- Step 1: Start from RX 2 and TX 1. A genie provides $\bar{\mathbf{G}}_1 = \bar{\mathbf{X}}^{[1]}_{(3a)}$ to RX 2 such that it can decode all the messages subject to noise distortion. After zero forcing the interference from TX 3 and TX 4, RX 2 originally has $5a$ dimensional observations of the signals sent from TX 1, which is denoted as $\mathcal{O}$ where $|\mathcal{O}| = 5a$. This process produces the first sum rate inequality
$$\begin{aligned} 4nR - n\epsilon_n &\leq: \hbar(\mathbf{Y}^{[2]^n}) + \hbar(\mathbf{G}_1^n|\mathbf{S}^{[2]^n}) \\ &\leq: Nn\log\rho + \hbar(\mathbf{G}_1^n|\mathcal{O}^n) \\ &\leq: Nn\log\rho + nR - \hbar(\mathcal{O}^n). \end{aligned} \tag{113}$$
- Step 2: Since $|\mathcal{O}| + 5a > M$, we go to RX 3 looking at TX 1 for an intermediate bound. A genie provides $\bar{\mathbf{G}}_2 = \bar{\mathbf{X}}'^{[1]}_{(3a)}$ to RX 3, which originally has $5a$ dimensional observations of signals sent from TX 1 (denoted as $\mathcal{O}'$), after zero forcing these $5a$ dimensional observations. Because $|\mathcal{O}|+|\mathcal{O}'| = 10a > 8a = M$, they have a $2a$-dimensional intersection. We denote this intersection by $\mathcal{I}$. This process produces the intermediate bound
$$\begin{aligned} &4nR - n\epsilon_n \\ &\quad\leq: \hbar(\mathbf{Y}^{[3]^n}) + \hbar(\mathbf{G}_2^n|\mathbf{S}^{[3]^n}) \\ &\quad\leq: Nn\log\rho + \hbar(\mathcal{O}^n|\mathcal{O}'^n) \\ &\quad\leq: Nn\log\rho + nR - \hbar(\mathcal{O}'^n) - \hbar(\mathcal{O}^n) + \hbar(\mathcal{O}^n) \\ &\quad\leq: Nn\log\rho + \hbar(\mathcal{O}^n) - \hbar(\mathcal{I}^n). \end{aligned} \tag{114}$$
Then we update $\mathcal{O} = \mathcal{I}$ and $|\mathcal{O}| = 2a$.
- Step 3: Since $|\mathcal{I}| + 5a = 7a < M$, we do not need an intermediate bound here. We go to RX 4 looking at TX 3. A genie provides $\bar{\mathbf{G}}_3 = \{\mathcal{O}+\mathbf{Z}, \bar{\mathbf{X}}^{[3]}_{(a)}\}$ to RX 4, which originally has $5a$ dimensional observations of signals sent from TX 3, denoted as $\mathcal{O}'$. Since providing $\mathcal{O}$, which is associated with User 1, to RX 4 releases another $|\mathcal{O}| = 2a$ observations of signals sent from TX 3, denoted as $\tilde{\mathcal{O}}$, RX 4 has a total of $|\tilde{\mathcal{O}}| + M_0 = 7a$ dimensional observations of $\mathbf{X}^{[3]}$. This process produces
$$\begin{aligned} &4nR - n\epsilon_n \\ &\quad\leq: \hbar(\mathbf{Y}^{[4]^n}) + \hbar(\mathbf{G}_3^n|\mathbf{S}^{[4]^n}) \\ &\quad\leq: Nn\log\rho + \hbar(\mathcal{O}^n, \mathbf{X}^{[3]^n}_{(a)}|\mathbf{S}^{[4]^n}) \\ &\quad\leq\ Nn\log\rho + \hbar(\mathcal{O}^n) + \hbar(\mathbf{X}^{[3]^n}_{(a)}|\mathbf{S}^{[4]^n}, \mathcal{O}^n, \mathcal{O}'^n) \\ &\quad\leq\ Nn\log\rho + \hbar(\mathcal{O}^n) + \hbar(\mathbf{X}^{[3]^n}_{(a)}|\tilde{\mathcal{O}}^n, \mathcal{O}'^n) \\ &\quad\leq: Nn\log\rho + \hbar(\mathcal{O}^n) + nR - \hbar(\tilde{\mathcal{O}}^n, \mathcal{O}'^n). \end{aligned} \tag{115}$$
Now we update $\mathcal{O} = \{\tilde{\mathcal{O}}, \mathcal{O}'\}$ and $|\mathcal{O}| = 7a$.
- Step 4: Now since $|\mathcal{O}| + 5a > M$, we need an intermediate bound. So we go to RX 1 still looking at TX 3. A genie provides $\bar{\mathbf{G}}_4 = \bar{\mathbf{X}}^{[3]}_{(3a)}$ to RX 1, which originally has $5a$ dimensional observations of signals sent from TX 1, denoted as $\mathcal{O}'$ and $|\mathcal{O}'| = 5a$. Because $|\mathcal{O}| + |\mathcal{O}'| = 12a > M$, they have a $4a$-dimensional intersection. We denote this intersection by $\mathcal{I}$. This process produces the intermediate bound as follows:
$$\begin{aligned} &4nR - n\epsilon_n \\ &\quad\leq: \hbar(\mathbf{Y}^{[1]^n}) + \hbar(\mathbf{G}_4^n|\mathbf{S}^{[1]^n}) \\ &\quad\leq: Nn\log\rho + \hbar(\mathbf{X}^{[3]^n}_{(3a)}|\mathcal{O}'^n) \\ &\quad\leq: Nn\log\rho + nR - \hbar(\mathcal{O}'^n) - \hbar(\mathcal{O}^n) + \hbar(\mathcal{O}^n) \\ &\quad\leq: Nn\log\rho + \hbar(\mathcal{O}^n) - \hbar(\mathcal{I}^n). \end{aligned} \tag{116}$$
Then we update $\mathcal{O} = \mathcal{I}$ and $|\mathcal{O}| = 4a$.
- Step 5: Because we have again $|\mathcal{O}| + 5a > M$, we need to resort to the next RX, RX 2, looking at TX 3 for another intermediate bound. Now a genie provides the set $\bar{\mathbf{G}}_3 = \bar{\mathbf{X}}'^{[3]^n}_{(3a)}$ to RX 2, which also has $5a$ dimensional observations of $\mathbf{X}^{[3]}$, denoted as $\mathcal{O}'$. We denote by $\mathcal{I}$ the intersection of $\mathcal{O}$ and $\mathcal{O}'$, and $|\mathcal{I}| = |\mathcal{O}| + 5a - 8a = a$. This process produces the intermediate bound
$$\begin{aligned} &4nR - n\epsilon_n \\ &\quad\leq: \hbar(\mathbf{Y}^{[2]^n}) + \hbar(\mathbf{G}_5^n|\mathbf{S}^{[2]^n}) \\ &\quad\leq: Nn\log\rho + \hbar(\mathbf{X}'^{[3]^n}_{(3a)}|\mathcal{O}'^n) \\ &\quad\leq: Nn\log\rho + nR - \hbar(\mathcal{O}'^n) - \hbar(\mathcal{O}^n) + \hbar(\mathcal{O}^n) \\ &\quad\leq: Nn\log\rho + \hbar(\mathcal{O}^n) - \hbar(\mathcal{I}^n). \end{aligned} \tag{117}$$
Then we update $\mathcal{O} = \mathcal{I}$ and $|\mathcal{O}| = a$.
- Step 6: Since $|\mathcal{O}| + 5a = 6a < M$, we do not need an intermediate bound here. Now let us recall how we obtain the observations (subspaces) $\mathcal{O}$ here. We start from RX 4 in Step 3 where we have $7a$ dimensional observations of $\mathbf{X}^{[3]}$, $5a$ dimensions that RX 4 originally has and the other $2a$ dimensions opened up by the genie. Then we intersect these $7a$ dimensional observations with the $5a$ dimensional observations at RX 1 and RX 2 in Step 4 and Step 5, respectively, to produce $\mathcal{O}$. That is to say, the a dimensional observations $\mathcal{O}$ are already contained in the clean observations at RX 1 and RX 2. Therefore, we cannot provide $\mathcal{O}$ as a genie to those two receivers. Also, $\mathcal{O}$ is the observations of $\mathbf{X}^{[3]}$ and thus cannot be a genie provided to RX 3. Thus, we can only provide it as a genie to RX 4. Moreover, we want to use $\mathcal{O}$ as a genie to open up the dimensions of signals sent from other TXs, i.e., not TX 3 or TX 4. Suppose a genie provides $\bar{\mathbf{G}}_4 = \{\mathcal{O}+\mathbf{Z}, \bar{\mathbf{X}}^{[2]}_{(2a)}\}$ to RX 2, which originally has $5a$ dimensional observations of signals sent from TX 2, denoted as $\mathcal{O}'$. As we have described, providing $\mathcal{O}$ to RX 4 releases another $|\mathcal{O}| = a$ observations of signals sent from TX 2, denoted as $\tilde{\mathcal{O}}$, RX 4 has a total of $|\tilde{\mathcal{O}}| + M_0 = 6a$ dimensional observations of $\mathbf{X}^{[2]}$.

This process produces the following inequality:

$$\begin{aligned}
&4nR - n\epsilon_n \\
&\quad \leq: \hbar(\mathbf{Y}^{[4]^n}) + \hbar(\mathbf{G}_6^n|\mathbf{S}^{[4]^n}) \\
&\quad \leq: Nn\log\rho + \hbar(\mathcal{O}^n, \mathbf{X}_{(2a)}^{[2]^n}|\mathbf{S}^{[4]^n}) \\
&\quad \leq Nn\log\rho + \hbar(\mathcal{O}^n) + \hbar(\mathbf{X}_{(2a)}^{[2]^n}|\mathbf{S}^{[4]^n}, \mathcal{O}^n, \mathcal{O}'^n) \\
&\quad \leq Nn\log\rho + \hbar(\mathcal{O}^n) + \hbar(\mathbf{X}_{(2a)}^{[2]^n}|\tilde{\mathcal{O}}^n, \mathcal{O}'^n) \\
&\quad \leq: Nn\log\rho + \hbar(\mathcal{O}^n) + nR - \hbar(\tilde{\mathcal{O}}^n, \mathcal{O}'^n). \qquad (118)
\end{aligned}$$

Now we update $\mathcal{O} = \{\tilde{\mathcal{O}}, \mathcal{O}'\}$ and $|\mathcal{O}| = 6a$.

- Step 7: Since $|\mathcal{O}| + 5a > M$, we again need an intermediate bound. Consider RX 1 looking at TX 2. A genie provides $\bar{\mathbf{G}} = \bar{\mathbf{X}}_{(3a)}^{[2]}$ to RX 1, which originally has $5a$ dimensional observations of signals sent from TX 2, denoted as $\mathcal{O}'$ and $|\mathcal{O}'| = 5a$. Because $|\mathcal{O}| + |\mathcal{O}'| = 11a > M$, they have a $3a$-dimensional intersection. We denote this intersection by $\mathcal{I}$. Then this process produces the intermediate bound

$$\begin{aligned}
&4nR - n\epsilon_n \\
&\quad \leq: \hbar(\mathbf{Y}^{[1]^n}) + \hbar(\mathbf{G}_7^n|\mathbf{S}^{[1]^n}) \\
&\quad \leq: Nn\log\rho + \hbar(\mathbf{X}_{(3a)}^{[2]^n}|\mathcal{O}'^n) \\
&\quad \leq: Nn\log\rho + nR - \hbar(\mathcal{O}'^n) - \hbar(\mathcal{O}^n) + \hbar(\mathcal{O}^n) \\
&\quad \leq: Nn\log\rho + \hbar(\mathcal{O}^n) - \hbar(\mathcal{I}^n). \qquad (119)
\end{aligned}$$

Then we update $\mathcal{O} = \mathcal{I}$ and $|\mathcal{O}| = 3a$.

- Step 8: Finally, consider RX 3 looking at TX 2. A genie provides $\bar{\mathbf{G}}_8 = \mathcal{O} + \mathbf{Z}$ to RX 3, which originally has $5a$ dimensional observations of TX 2, denoted as $\mathcal{O}'$, which combined with the $|\mathcal{O}| = 3a$ dimensional genie signals from TX 2 allows RX 3 to recover $\mathbf{X}^{[2]}$ subject to noise distortion. This process produces the inequality

$$\begin{aligned}
4nR &\leq: \hbar(\mathbf{Y}^{[3]^n}) + \hbar(\mathbf{G}_8^n|\mathbf{S}^{[3]^n}) \qquad (120) \\
&\leq Nn\log\rho + \hbar(\mathcal{O}^n). \qquad (121)
\end{aligned}$$

Notice that in each of the eight inequalities from (113) to (121), the differential entropy term with the negative sign always appears with the positive sign in the next inequality. Thus, adding up all the eight sum rate inequalities from (113) to (121), all the negative terms on the right-hand side are fully canceled out, thus producing the following inequality:

$$\begin{aligned}
32nR - 8n\epsilon_n &\leq: 8Nn\log\rho + 3nR + n\ o(\log\rho) \\
\Rightarrow d &\leq \frac{8N}{29} = \frac{8\times 21}{29}.
\end{aligned}$$

Note that in each step we still need to ensure at each step the clean $N-2M$ dimensional observations of the associated TX are linearly independent of the $|\bar{\mathbf{G}}| = 3M-N$ dimensional observations of that TX opened up by the provided genie $\bar{\mathbf{G}}$. For this case of $M/N = 8/21$, we rely on the numerical test by randomly generating the channel matrices to show that this independence claim is true. ■

Remark: For any (M, N) pair where $M/N \in [3/8, 2/5)$, after running the algorithm we obtain a series of inequalities, in which the intermediate bounds can appear successively for at most twice. In addition, we cannot derive more than two *successive* intermediate bounds. The reason is the following. At any step, if we provide $\mathcal{O}$ as a genie where $|\mathcal{O}| + (N-2M) > M$, then we need an intermediate bound. After deriving that inequality, we provide a $|\mathcal{O}|+(N-2M)-M$ dimensional genie to the RX we consider next. Again, if $|\mathcal{O}|+(N-2M)-M+(N-2M) > M$, we need immediately another intermediate bound. With the same analysis, suppose we need a third successive intermediate bound, we have to have:

$$|\mathcal{O}| + (N-2M) - M + (N-2M) - M + (N-2M) > M$$

which, due to $|\mathcal{O}| < M$, implies that $\frac{M}{N} < \frac{3}{8}$, which is contradictive. Intuitively, this conclusion implies that the three $N-2M$ dimensional clean observations of one interferer at all undesired receivers have only null intersection in common after projecting the clean subspaces back to that TX. Furthermore, for the K user $M\times N$ MIMO interference channel, the clean observations of one interferer at all undesired receivers, after we project them back to that transmit space, will have a common intersection with $[(K-1)(N-(K-2)M) - (K-2)M]^+$ dimension. This intersection would be the null space as long as

$$\begin{aligned}
(K-1)(N-(K-2)M) &\leq (K-2)M \\
\Longrightarrow \frac{M}{N} &\geq \frac{K-1}{K(K-2)}.
\end{aligned}$$

In general, in the K-user case, we may have up to $K-2$ successive intermediate bounds.

VI. Examples of Applications of Genie Chains

Aside from the application of the genie chains in $K = 4$ user semi-symmetric $M_T \times M_R$ MIMO interference channel where $M_T < M_R$, the tool of genie chains can also be applied to the reciprocal $M_T > M_R$ setting and many other wireless networks to produce the desired information theoretic DoF outer bound. In this section, we will provide four specific examples to show the application of genie chains. Note that the DoF of all these four examples were open before the introduction of the 'genie chains' approach.

A. DoF of the $K = 4$ User Reciprocal Setting

We consider $(M_T, M_R) = (8, 3) = (N, M)$ MIMO interference channel, as an example of the reciprocal $M_T > M_R$ setting in this section. The channel model and associated definitions and notations are identical to that in Section II. We are going to show that each user in this channel has 24/11 DoF. Since the achievability has already been shown in [7], we focus on the information theoretic DoF outer bound.

Proof: As we have shown in previous examples, intuitively, we need a total of $M_T = 8$ sum rate bounds, which can be produced through the following eight steps. Note that in each step, the genie should have at least $|\bar{\mathbf{G}}| = 3N - M = 21$ dimensions.

- Step 1: Consider RX 2 and TX 1. After decoding its desired message W_2, RX 2 can reconstruct $\mathbf{X}^{[2]}$

and subtract it from its received signal vector. Thus, RX 2 has 3 linear combinations of $8 \times 3 = 24$ interference symbols from TX1, TX 3 and TX 4. A genie provides $\bar{\mathbf{G}}_1 = \{\bar{\mathbf{X}}^{[1]}_{(5)}, \bar{\mathbf{X}}^{[3]}, \bar{\mathbf{X}}^{[4]}\}$ to RX 2. From provided $\bar{\mathbf{X}}^{[3]}, \bar{\mathbf{X}}^{[4]}$, RX 2 can decode W_3, W_4. After RX 2 subtracts $\mathbf{X}^{[2]}, \mathbf{X}^{[3]}, \mathbf{X}^{[4]}$, it has three dimensional observations of interference from TX 1, which are linearly independent of the provided genie signals $\bar{\mathbf{X}}^{[1]}_{(5)}$. By inverting the channel matrix associated with TX 1, RX 2 can decode W_1 as well subject to noise distortion. Hence, we obtain the first sum rate inequality:

$$\begin{aligned}
&4nR - n\epsilon_n \\
&\quad \leq: \hbar(\mathbf{Y}^{[2]^n}) + \hbar(\mathbf{G}_1^n|\mathbf{S}^{[2]^n}) && (122)\\
&\quad \leq: nM_R \log\rho + \hbar(\mathbf{X}^{[1]^n}_{(5)}, \mathbf{X}^{[3]^n}, \mathbf{X}^{[4]^n}|\mathbf{Y}^{[2]^n}) \\
&\quad \leq: nM_R \log\rho + \hbar(\mathbf{X}^{[3]^n}, \mathbf{X}^{[4]^n}) \\
&\qquad + \hbar(\mathbf{X}^{[1]^n}_{(5)}|\mathbf{Y}^{[2]^n}, \mathbf{X}^{[3]^n}, \mathbf{X}^{[4]^n}) && (123)\\
&\quad \leq: nM_R \log\rho + 2nR + \hbar(\mathbf{X}^{[1]^n}_{(5)}|\mathbf{X}_3^{[1\sim 2]^n}) && (124)\\
&\quad \leq: nM_R \log\rho + 2nR + \hbar(\mathbf{X}^{[1]^n}_{(5)}, \mathbf{X}_3^{[1\sim 2]^n}) \\
&\qquad - \hbar(\mathbf{X}_3^{[1\sim 2]^n}) && (125)\\
&\quad \leq: nM_R \log\rho + 3nR - \hbar(\mathbf{X}_3^{[1\sim 2]^n}). && (126)
\end{aligned}$$

- Step 2: Consider RX 3 and TX 1. Similar to Step 1, a genie provides $\bar{\mathbf{G}}_2 = \{\bar{\mathbf{X}}^{[1]}_{(5)}, \bar{\mathbf{X}}^{[2]}, \bar{\mathbf{X}}^{[4]}\}$ to RX 2, such that it can decode all the messages as well subject to noise distortion. This step produces the second inequality as follows:

$$\begin{aligned}
4nR - n\epsilon_n &\leq: \hbar(\mathbf{Y}^{[3]^n}) + \hbar(\mathbf{G}_2^n|\mathbf{S}^{[3]^n}) \\
&\leq: nM_R \log\rho + 3nR - \hbar(\mathbf{X}_3^{[1\sim 3]^n}).
\end{aligned}$$

- Step 3: Consider RX 4 and TX 2. A genie provides $\bar{\mathbf{G}}_3 = \{\bar{\mathbf{X}}_3^{[1\sim 2]}, \bar{\mathbf{X}}_3^{[1\sim 3]}, \bar{\mathbf{X}}^{[2]}_{(7)}, \bar{\mathbf{X}}^{[3]}\}$ to RX 4. Note that RX 4 is able to decode W_4 and reconstruct $\mathbf{X}^{[4]}$, and then subtract it from its received signal vector. Thus, RX 4 has three dimensional observations of the 16 interference symbols from TX 1 and TX 2. With genie $\bar{\mathbf{X}}_3^{[1\sim 2]}, \bar{\mathbf{X}}_3^{[1\sim 3]}, \bar{\mathbf{X}}^{[2]}_{(7)}$, RX 4 can invert the square channel matrix associated with TX 1 and TX 2, and thus decode the other two messages as well subject to noise distortion. Therefore, we have the second inequality as follows:

$$\begin{aligned}
&4nR - n\epsilon_n \\
&\leq: \hbar(\mathbf{Y}^{[4]^n}) + \hbar(\mathbf{G}_3^n|\mathbf{S}^{[4]^n}) \\
&\leq: nM_R \log\rho \\
&\quad + \hbar(\mathbf{X}_3^{[1\sim 2]^n}, \mathbf{X}_3^{[1\sim 3]^n}, \mathbf{X}^{[2]^n}_{(7)}, \mathbf{X}^{[3]^n}|\mathbf{Y}^{[4]^n}) \\
&\leq: nM_R \log\rho + \hbar(\mathbf{X}_3^{[1\sim 2]^n}, \mathbf{X}_3^{[1\sim 3]^n}) \\
&\quad + \hbar(\mathbf{X}^{[3]^n}) + \hbar(\mathbf{X}^{[2]^n}_{(7)}|\mathbf{S}^{[4]^n}, \mathbf{X}_3^{[1\sim 2]^n}, \mathbf{X}_3^{[1\sim 3]^n}, \mathbf{X}^{[3]^n}) \\
&\leq: nM_R \log\rho + \hbar(\mathbf{X}_3^{[1\sim 2]^n}) + \hbar(\mathbf{X}_3^{[1\sim 3]^n}) \\
&\quad + nR + \hbar(\mathbf{X}^{[2]^n}_{(7)}|\mathbf{X}_1^{[2\sim 4]^n}) \\
&\leq: nM_R \log\rho + \hbar(\mathbf{X}_3^{[1\sim 2]^n}) + \hbar(\mathbf{X}_3^{[1\sim 3]^n}) + nR \\
&\quad + \hbar(\mathbf{X}^{[2]^n}_{(7)}, \mathbf{X}_1^{[2\sim 4]^n}) - \hbar(\mathbf{X}_1^{[2\sim 4]^n}) \\
&\leq: nM_R \log\rho + 2nR + \hbar(\mathbf{X}_3^{[1\sim 2]^n}) \\
&\quad + \hbar(\mathbf{X}_3^{[1\sim 3]^n}) - \hbar(\mathbf{X}_1^{[2\sim 4]^n}).
\end{aligned}$$

- Step 4: Consider RX 1 and TX 2. Similar to Step 1, a genie provides $\bar{\mathbf{G}}_4 = \{\bar{\mathbf{X}}^{[2]}_{(5)}, \bar{\mathbf{X}}^{[3]}, \bar{\mathbf{X}}^{[4]}\}$ to RX 1, such that RX 1 can decode all the messages subject to noise distortion. Following the derivations in Step 1, we obtain the fourth sum rate inequality:

$$\begin{aligned}
4nR - n\epsilon_n &\leq: \hbar(\mathbf{Y}^{[1]^n}) + \hbar(\mathbf{G}_4^n|\mathbf{S}^{[1]^n}) \\
&\leq: nM_R \log\rho + 3nR - \hbar(\mathbf{X}_3^{[2\sim 1]^n}).
\end{aligned}$$

- Step 5: Consider RX 3 and TX 2. Similar to Step 2, a genie provides $\bar{\mathbf{G}}_5 = \{\bar{\mathbf{X}}^{[2]}_{(5)}, \bar{\mathbf{X}}^{[1]}, \bar{\mathbf{X}}^{[4]}\}$ to RX 2, such that it can decode all the messages subject to noise distortion. Thus this step produces the fifth inequality as follows:

$$\begin{aligned}
4nR - n\epsilon_n &\leq: \hbar(\mathbf{Y}^{[3]^n}) + \hbar(\mathbf{G}_2^n|\mathbf{S}^{[3]^n}) \\
&\leq: nM_R \log\rho + 3nR - \hbar(\mathbf{X}_3^{[2\sim 3]^n}).
\end{aligned}$$

- Step 6: Consider RX 2 and TX 4. A genie provides $\bar{\mathbf{G}}_6 = \{\bar{\mathbf{X}}^{[4]}_{(5)}, \bar{\mathbf{X}}^{[1]}, \bar{\mathbf{X}}^{[3]}\}$ to RX 2, such that RX 2 is able to decode all the messages subject to noise distortion. The reasoning and derivations of this step are similar to that of Step 1 and Step 4, and thus we have the following inequality:

$$\begin{aligned}
4nR - n\epsilon_n &\leq: \hbar(\mathbf{Y}^{[2]^n}) + \hbar(\mathbf{G}_6^n|\mathbf{S}^{[2]^n}) \\
&\leq: nM_R \log\rho + 3nR - \hbar(\mathbf{X}_3^{[4\sim 2]^n}).
\end{aligned}$$

- Step 7: Consider RX 3 and TX 4. A genie provides $\bar{\mathbf{G}}_7 = \{\bar{\mathbf{X}}^{[4]}_{(5)}, \bar{\mathbf{X}}^{[1]}, \bar{\mathbf{X}}^{[2]}\}$ to RX 3, such that RX 3 can decode all the messages subject to noise distortion. This step is similar to Step 2 and Step 5. This step produces the seventh inequality as follows:

$$\begin{aligned}
4nR - n\epsilon_n &\leq: \hbar(\mathbf{Y}^{[3]^n}) + \hbar(\mathbf{G}_7^n|\mathbf{S}^{[3]^n}) \\
&\leq: nM_R \log\rho + 3nR - \hbar(\mathbf{X}_3^{[4\sim 3]^n}).
\end{aligned}$$

- Step 8: Consider RX 1 and TX 4. A genie provides $\bar{\mathbf{G}}_8 = \{\bar{\mathbf{X}}_3^{[4\sim 2]}, \bar{\mathbf{X}}_3^{[4\sim 3]}, \bar{\mathbf{X}}^{[2]}_{(7)}, \bar{\mathbf{X}}^{[3]}\}$ to RX 1, such that RX 1 is able to decode all the messages subject to noise distortion. This step is similar to Step 3. Therefore, we have the eighth inequality as follows:

$$\begin{aligned}
4nR - n\epsilon_n &\leq: \hbar(\mathbf{Y}^{[1]^n}) + \hbar(\mathbf{G}_8^n|\mathbf{S}^{[1]^n}) \\
&\leq: nM_R \log\rho + 2nR + \hbar(\mathbf{X}_3^{[4\sim 2]^n}) \\
&\quad + \hbar(\mathbf{X}_3^{[4\sim 3]^n}) - \hbar(\mathbf{X}_1^{[2\sim 1]^n}).
\end{aligned}$$

Finally, adding up all the eight sum rate inequalities we have so far, we obtain the following inequality:

$$\begin{aligned}
32nR - 8n\epsilon_n &\leq 8M_R n\log\rho + 22nR - \hbar(\mathbf{X}_1^{[2\sim 4]^n}) \\
&\quad - \hbar(\mathbf{X}_3^{[2\sim 1]^n}) - \hbar(\mathbf{X}_3^{[2\sim 3]^n}) - \hbar(\mathbf{X}_1^{[2\sim 1]^n}) \\
&\leq 8M_R n\log\rho + 22nR - nR.
\end{aligned}$$

where

$$\begin{aligned} &R - \epsilon_n \\ &\quad = \hbar(\mathbf{X}_1^{[2\sim4]}, \mathbf{X}_3^{[2\sim1]}, \mathbf{X}_3^{[2\sim3]}, \mathbf{X}_1^{[2\sim1]}) \\ &\quad \le \hbar(\mathbf{X}_1^{[2\sim4]}) + \hbar(\mathbf{X}_3^{[2\sim1]}) + \hbar(\mathbf{X}_3^{[2\sim3]}) + \hbar(\mathbf{X}_1^{[2\sim1]}). \end{aligned}$$

By letting $\rho \to \infty$ and $n \to \infty$ we obtain the desired outer bound:

$$d \le \frac{8M_R}{11} = \frac{24}{11}. \tag{127}$$

■

B. DoF of the K-User MIMO Interference Channel

In this section, we take one simple example of the MIMO interference channel beyond the $K = 4$ user setting to convey the idea of genie chains. Consider the $(K, M, N) = (5, 4, 15)$ setting, we will show that $d = \frac{MN}{M+N} = \frac{60}{19}$.

Proof: We need $M = 4$ sum rate bounds, which can be produced through the following four steps.

- Step 1: Start from RX 2 and TX 1. After decoding its desired message W_2, RX 2 can reconstruct $\mathbf{X}^{[2]}$ and subtract it from its received signal vector. Thus, RX 2 has 15 linear combinations of $4 \times 4 = 16$ interference symbols from TX 2 to TX 5. A genie provides $\bar{\mathbf{G}}_1 = \bar{\mathbf{X}}_{(1)}^{[1]}$ to RX 2. Since $\mathbf{X}_{(1)}^{[1]}$ is linearly independent of all the other 15 dimensions, RX 2 can invert the 16×16 square matrix to reconstruct the interference vectors sent from all interferers, and thus RX can decode all the messages. Thus, we obtain the first sum rate inequality:

$$\begin{aligned} 5nR - n\epsilon_n &\le: \hbar(\mathbf{Y}^{[2]^n}) + \hbar(\mathbf{G}_1^n|\mathbf{S}^{[2]^n}) && (128)\\ &\le: Nn\log\rho + \hbar(\mathbf{X}_{(1)}^{[1]^n}|\mathcal{O}_2^n) && (129)\\ &=: Nn\log\rho + nR - \hbar(\mathcal{O}_2^n) && (130) \end{aligned}$$

  where $\mathcal{O}_2$ denotes the 3 dimensional observations of TX 1 at RX 2, after zero forcing the interference from TX 3, TX 4 and TX 5.
- Step 2: Since $|\mathcal{O}| + 3 > M$, we go to RX 3 looking at TX 1 for an intermediate bound. Similar to Step 1, a genie provides $\bar{\mathbf{G}}_2 = \bar{\mathbf{X}}_{(1)}^{[1]}$ to RX 3 such that it can decode all messages. Again, RX 3 originally has 3 dimensional observations of TX 1 after it zero forces the interference from TX 2, TX 4 and TX 5. We denote by $\mathcal{O}_3$ the 3 dimensional observations at RX 3. Because $|\mathcal{O}_2| + |\mathcal{O}_3| = 6 > M$, they have a 2 dimensional intersection. We denote this intersection by $\mathcal{I}_3 = \mathcal{O}_2 \cap \mathcal{O}_3$ at RX 3. Thus, we have the intermediate bound as follows:

$$\begin{aligned} &5nR - n\epsilon_n \\ &\quad \le: \hbar(\mathbf{Y}^{[3]^n}) + \hbar(\mathbf{G}_2^n|\mathbf{S}^{[3]^n}) && (131)\\ &\quad \le: Nn\log\rho + \hbar(\mathbf{X}_{(1)}^{[1]^n}|\mathcal{O}_3^n) && (132)\\ &\quad \le: Nn\log\rho + \hbar(\mathbf{X}_{(1)}^{[1]^n}, \mathcal{O}_3^n) - \hbar(\mathcal{O}_3^n) && (133)\\ &\quad \le: Nn\log\rho + nR + \hbar(\mathcal{O}_2^n) - \hbar(\mathcal{I}_3^n, \mathcal{O}_3^n \setminus \mathcal{I}_3^n) \\ &\qquad - \hbar(\mathcal{O}_2^n) && (134)\\ &\quad \le: Nn\log\rho + nR + \hbar(\mathcal{O}_2^n) - \hbar(\mathcal{I}_3^n) \\ &\qquad - \hbar(\mathcal{O}_3^n \setminus \mathcal{I}_3^n|\mathcal{I}_3^n) - \hbar(\mathcal{O}_2^n) && (135)\\ &\quad \le: Nn\log\rho + nR + \hbar(\mathcal{O}_2^n) - \hbar(\mathcal{I}_3^n) \\ &\qquad - \hbar(\mathcal{O}_3^n \setminus \mathcal{I}_3^n|\mathcal{I}_3^n, \mathcal{O}_2^n) - \hbar(\mathcal{O}_2^n) && (136)\\ &\quad \le: Nn\log\rho + nR + \hbar(\mathcal{O}_2^n) - \hbar(\mathcal{I}_3^n) \\ &\qquad - \hbar(\mathcal{O}_3^n \setminus \mathcal{I}_3^n|\mathcal{O}_2^n) - \hbar(\mathcal{O}_2^n) && (137)\\ &\quad \le: Nn\log\rho + nR + \hbar(\mathcal{O}_2^n) - \hbar(\mathcal{I}_3^n) - nR && (138)\\ &\quad \le: Nn\log\rho + \hbar(\mathcal{O}_2^n) - \hbar(\mathcal{I}_3^n). && (139) \end{aligned}$$

- Step 3: Next, because $|\mathcal{I}_3| + 3 > M$ still, we go to RX 4 looking at TX 1 for another intermediate bound. A genie provides $\bar{\mathbf{G}}_3 = \bar{\mathbf{X}}_{(1)}^{[1]}$ to RX 4, which originally also has 3 dimensional observations of TX 1, denoted as $\mathcal{O}_4$. Because $|\mathcal{I}_3| + |\mathcal{O}_4| = 5 > M$ again, they have a one-dimensional intersection. We denote this intersection by $\mathcal{I}_4 = \mathcal{I}_3 \cap \mathcal{O}_4$ at RX 4. Similar to Step 2, this process produces the intermediate bound

$$\begin{aligned} 5nR - n\epsilon_n &\le: \hbar(\mathbf{Y}^{[4]^n}) + \hbar(\mathbf{G}_3^n|\mathbf{S}^{[4]^n}) && (140)\\ &\le: Nn\log\rho + \hbar(\mathcal{I}_3^n) - \hbar(\mathcal{I}_4^n). && (141) \end{aligned}$$

- Step 4: Finally, consider RX 5 looking at RX 1. A genie provides the one-dimensional symbol $\bar{\mathbf{G}}_4 = \mathcal{I}_4 + \mathbf{Z}$ to RX 5, which again originally has 3 dimensional observations of TX 1, denoted as $\mathcal{O}_5$, which combined with $\mathcal{I}_4$ allows RX 5 to recover $\mathbf{X}^{[1]}$ subject to noise distortion. Once again, this process produces the inequality

$$\begin{aligned} 5nR - n\epsilon_n &\le: \hbar(\mathbf{Y}^{[5]^n}) + \hbar(\mathbf{G}_4^n|\mathbf{S}^{[5]^n}) && (142)\\ &\le: Nn\log\rho + \hbar(\mathcal{I}_4^n). && (143) \end{aligned}$$

  Adding up all the four sum rate inequalities we have so far, we obtain the following inequality:

$$20nR - 4n\epsilon_n \le: 4Nn\log\rho + nR. \tag{144}$$

  By letting $\rho \to \infty$ and $n \to \infty$ we obtain the desired bound:

$$d \le \frac{4N}{19} = \frac{60}{19}. \tag{145}$$

■

C. DoF of the Many-to-One MIMO Interference Channel

In this section, we consider an example of Many-to-One MIMO interference channel, where in a K user interference channel, only one RX hears all transmitters while the other receivers can only hear their own desired signals. The Five-to-One MIMO interference channel is shown in Figure 2, where each TX has M antennas and each RX has N antennas.

1) Five-to-One MIMO Interference Channel: Consider the $(M, N) = (2, 5)$ setting in Figure 2, we are interested in the DoF per user of this network. Since RX k, $k = 2, 3, 4, 5$ only hear their own desired signals and $M < N$, they can decode their own messages respectively by the reliable communications requirement. Thus, we need to ensure that at RX 1 all interference is aligned as much as possible. Note that since the DoF counting bound is less than the decomposition

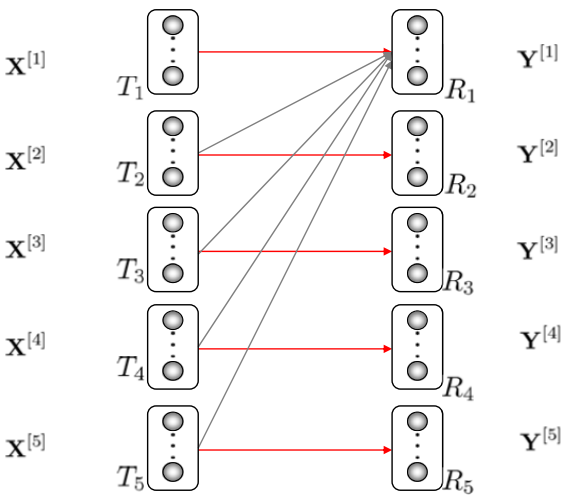

Fig. 2. (M, N) MIMO Five-to-One Interference Channel.

bound, we expect that the decomposition DoF bound, i.e., 10/7 DoF per user, is also the information theoretic DoF outer bound. Next, we show that this intuition is correct.

Proof: We need $M = 2$ sum rate bounds, which can be produced as follows.

- Step 1: A genie provides $\bar{\mathbf{G}}_1 = \{\bar{\mathbf{X}}^{[2]}_{(1)}, \bar{\mathbf{X}}^{[3]}\}$ to RX 1. By the assumption of reliable communications, RX 1 is able to decode its desired message W_1. After decoding W_1, RX 1 can reconstruct the signal vector $\mathbf{X}^{[1]}$ and then subtract it from $\bar{\mathbf{Y}}^{[1]}$. Therefore, RX 1 has 5-dimensional observations of the eight interference symbols from TX 2 to TX 5. Providing $\bar{\mathbf{G}}_1$ to RX 1 allows it to invert the square channel associated with the interferers. Therefore, RX 1 can reconstruct all signal vectors sent from all interferers, and thus decode all the messages. This argument produces the following inequality:

$$\begin{aligned}
&5nR - n\epsilon_n \\
&\quad \leq: \hbar(\mathbf{Y}^{[1]^n}) + \hbar(\mathbf{G}_1^n|\mathbf{S}^{[1]^n}) && (146)\\
&\quad \leq: Nn\log\rho + \hbar(\mathbf{X}^{[2]^n}_{(1)}, \mathbf{X}^{[3]^n}|\mathbf{S}^{[1]^n}) && (147)\\
&\quad \leq: Nn\log\rho + \hbar(\mathbf{X}^{[3]^n}) + \hbar(\mathbf{X}^{[2]^n}_{(1)}|\mathbf{S}^{[1]^n}, \mathbf{X}^{[3]^n}) \\
&\quad \leq: Nn\log\rho + nR + \hbar(\mathbf{X}^{[2]^n}_{(1)}|\mathcal{O}^n) && (148)\\
&\quad \leq: Nn\log\rho + nR + \hbar(\mathbf{X}^{[2]^n}_{(1)}, \mathcal{O}^n) - \hbar(\mathcal{O}^n) && (149)\\
&\quad \leq: Nn\log\rho + 2nR - \hbar(\mathcal{O}^n). && (150)
\end{aligned}$$

where $\mathcal{O}$ is the one dimensional observation of TX 2 after RX 1 removes its own signal $\mathbf{X}^{[1]}$, the provided genie signal $\bar{\mathbf{X}}^{[3]}$ and zero forces the interference from TX 4 and TX 5.

- Step 2: A genie provides $\bar{\mathbf{G}}_2 = \{\mathcal{O} + \mathbf{Z}, \bar{\mathbf{X}}^{[4]}\}$ to RX 1. Similar to Step 1, it can be easily seen that RX 1 can decode all the messages as well. Thus, we obtain the second inequality as follows:

$$\begin{aligned}
5nR - n\epsilon_n &\leq: \hbar(\mathbf{Y}^{[1]^n}) + \hbar(\mathbf{G}_2^n|\mathbf{S}^{[1]^n}) && (151)\\
&\leq: Nn\log\rho + \hbar(\mathcal{O}^n, \mathbf{X}^{[4]^n}) && (152)\\
&\leq: Nn\log\rho + \hbar(\mathcal{O}^n) + \hbar(\mathbf{X}^{[4]^n}) && (153)\\
&\leq: Nn\log\rho + nR + \hbar(\mathcal{O}^n). && (154)
\end{aligned}$$

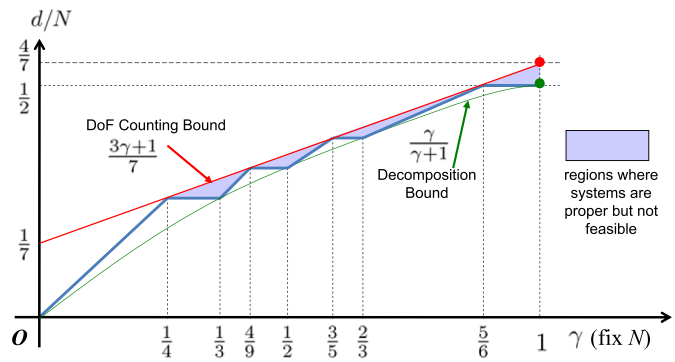

Fig. 3. d/N as a function of $\gamma = M/N$ for the Four-to-One MIMO interference channel.

Adding up the two sum rate inequalities we have so far, we obtain the following inequality:

$$10nR - 2n\epsilon_n \leq: 2Nn\log\rho + 3nR. \tag{155}$$

By letting first $n \to \infty$ and then $\rho \to \infty$ we obtain the desired the DoF outer bound:

$$d \leq \frac{2N}{7} = \frac{10}{7}. \tag{156}$$

■

2) Four-to-One MIMO Interference Channel: Besides the example of $K = 5$ setting shown above, we also present the DoF results of $K = 4$ setting in the following. In order to understand the interplay among spatial signal dimensions projected from interferers without zero forcing at the TX side, we only consider the $M \leq N$ setting. The DoF results are included in the following theorem.

Theorem 2: For a Four-to-One $M \times N$ MIMO Gaussian interference channel where each TX has M antennas, each RX has N antennas and $M \leq N$, the DoF value per user is given by:

$$d = \begin{cases} M, & M/N \leq 1/4, \\ N/4, & 1/4 \leq M/N \leq 1/3, \\ 3M/4, & 1/3 \leq M/N \leq 4/9, \\ N/3, & 4/9 \leq M/N \leq 1/2, \\ 2M/3, & 1/2 \leq M/N \leq 3/5, \\ 2N/5, & 3/5 \leq M/N \leq 2/3, \\ 3M/5, & 2/3 \leq M/N \leq 5/6, \\ N/2, & 5/6 \leq M/N \leq 1. \end{cases} \tag{157}$$

Proof: The DoF achievability relies on linear interference alignment schemes. Since the proof follows from the subspace alignment chains that we introduced in [2] and genie chains that we primarily investigate in this paper, we defer the proof to Appendix D. ■

The DoF results are shown in Figure 3 where the red line represents the DoF counting bound, derived in Appendix E. Notice that Theorem 2 implies that the DoF value is a piecewise linear function depending on M and N alternatively, which means that there are antenna redundancies at either the TX side or the RX side. While it is again similar to the DoF value of the three user MIMO interference channel recently shown by Wang *et al.* in [2], the DoF cruve only contains eight pieces, in contrast to infinitely many of pieces in the three user MIMO interference channel. In addition, because

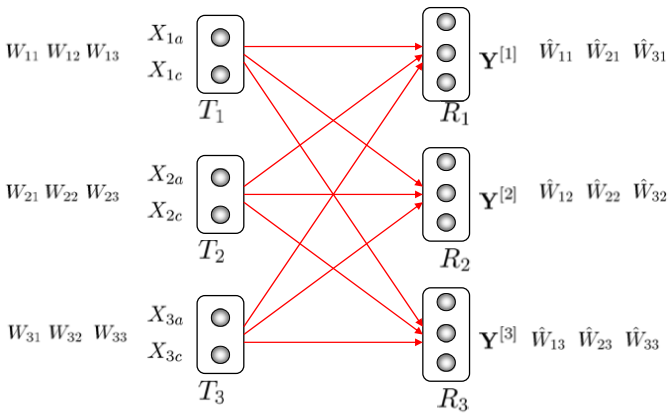

Fig. 4. $(M, N) = (2, 3)$ MIMO X Channel.

there is only one RX, we only need to deal with the signal dimensions projected from three interferers at that RX. Thus, understanding the spatial signal dimensions of this network is helpful for us to learn the spatial signal dimensions of more general networks.

In essence, Many-to-One MIMO interference channels are similar to cellular networks with two cells, as alignment is demanded to take place at only one interferer. Based on similar insights, the DoF value of MIMO two-cell cellular networks with 2 and 3 users per cell is found in [9] and [10], respectively.

D. DoF of the MIMO X Channel

Besides the multiuser interference channel, the tool of genie chains can also be applied in the MIMO X channel. We show one simple example in this section. Consider a $K = 3$ user MIMO X channel where each TX has $M = 2$ antennas and each RX has $N = 3$ antennas, as shown in Figure 4. Each TX T_i sends one independent message W_{ij} to RX R_j, $i, j \in \{1, 2, 3\}$. Again, the constant complex channel coefficients are assumed to be independently drawn from continuous distributions. Also, global channel knowledge is assumed to be available at all nodes. We refer to R_{ij} and d_{ij} as the rate and DoF, respectively, of the message W_{ij}. Again we are interested in the DoF of this network. Note that the value of DoF implied by the linear counting bound is less than that achieved by the decomposition bound. Thus, we expect the decomposition DoF bound, i.e., 10/7 DoF per user, is also the information theoretic DoF outer bound. Next, we show that this is true.

Proof: We need $M_T = 2$ sum rate bounds, produced as follows.

- Step 1: A genie provides $\bar{\mathbf{G}}_1 = \{\bar{X}_{2a}, W_{32}, W_{33}\}$ to RX 1. By the assumption of reliable communications, RX 1 is able to decode W_{11}, W_{21}, W_{31} from $\bar{\mathbf{Y}}^{[1]}$. Also, providing W_{32}, W_{33} to RX 1 allows it to reconstruct the signal $\mathbf{X}^{[3]} = [X_{3a}\ X_{3c}]^T$ and subtract it from $\bar{\mathbf{Y}}^{[1]}$. Then the remaining interference comes from TX 1 and TX 2. Since RX 1 has three antennas, providing X_{2a} to RX 1 allows it to invert the square matrix and reconstruct the transmit signal vectors from TX 1 and TX 2 subject to noise distortion, and thus RX 1 is able to decode all the messages subject to the distortion. Therefore, we have the following inequality:

$$
\begin{aligned}
&9nR - n\epsilon_n \\
&\leq: \hbar(\mathbf{Y}^{[1]^n}) + \hbar(\mathbf{G}_1^n|\mathbf{S}^{[1]^n}) && (158)\\
&=: Nn\log\rho + \hbar(X_{2a}^n, W_{32}, W_{33}|\mathbf{S}^{[1]^n}, W_{11}, W_{21}, W_{31}) \\
&\leq: Nn\log\rho + \hbar(W_{32}, W_{33}) \\
&\quad + \hbar(X_{2a}|\mathbf{Y}^{[1]^n}, W_{21}, W_{31}, W_{32}, W_{33}) && (159)\\
&\leq: Nn\log\rho + 2nR + \hbar(X_{2a}^n|\mathcal{O}^n, W_{21}) && (160)\\
&=: Nn\log\rho + 2nR + \hbar(X_{2a}^n, \mathcal{O}^n, W_{21}) \\
&\quad - \hbar(W_{21}) - \hbar(\mathcal{O}^n|W_{21}) && (161)\\
&=: Nn\log\rho + 2nR + 2nR - \hbar(\mathcal{O}^n|W_{21}). && (162)
\end{aligned}
$$

where $\mathcal{O}$ is the one dimensional linear combination of X_{2a} and X_{2c}.

- Step 2: A genie provides $\bar{\mathbf{G}}_2 = \{\mathcal{O} + \mathbf{Z}, W_{12}, W_{13}\}$ to RX 1. Similar to Step 1, it can be easily seen that RX 1 can decode all the messages. Thus, we obtain the second inequality as follows:

$$
\begin{aligned}
9nR - n\epsilon_n &\leq: \hbar(\mathbf{Y}^{[1]^n}) + \hbar(\mathbf{G}_2|\mathbf{S}^{[1]^n}) \\
&\leq: Nn\log\rho + \hbar(\mathcal{O}^n, W_{12}, W_{13}|\mathbf{S}^{[1]^n}, W_{21}) \\
&\leq: Nn\log\rho + \hbar(W_{12}, W_{13}) + \hbar(\mathcal{O}^n|W_{21}) \\
&\leq: Nn\log\rho + 2nR + \hbar(\mathcal{O}^n|W_{21}).
\end{aligned}
$$

Adding up the two sum rate inequalities we have so far, we obtain the following inequality:

$$18nR - 2n\epsilon_n \leq: 2Nn\log\rho + 6nR. \qquad (163)$$

By letting $n \to \infty$ and $\rho \to \infty$ we obtain the desired the DoF outer bound:

$$d \leq \frac{2N}{12} = \frac{1}{2} = \frac{MN}{3M + 2N}. \qquad (164)$$

■

VII. Discussion on the DoF Characterization of the K-User MIMO Gaussian Interference Channel

In this paper, since our primary goal is to introduce the genie chains approach, and highlight the principles that could be applied to not only MIMO interference channels, but also many other wireless networks such as X channel, etc., we focus mostly on the $K = 4$ user case with $M_T \leq M_R$ to present the main ideas in Section V. In this section, we continue to discuss the DoF results of the K-user $M_T \times M_R$ MIMO interference channel, not through rigorous proof for each case but based on the available observations that we obtained so far, to show a broad and fundamental DoF picture of the MIMO interference channel. Since the DoF results of the $K = 2$ and $K = 3$ User $M_T \times M_R$ MIMO Interference Channel have been reported in [2] and [11], respectively, we begin with $K \geq 4$ cases.

A. Unstructured Linear Schemes Achieving the Information Theoretic DoF Outer Bound

In Section V, we mention that the DoF result and corresponding proofs for $M/N \leq 3/8$ where $M = \min(M_T, M_R)$

and $N = \max(M_T, M_R)$ directly follows from the $K = 3$ user case [2]. In fact, we can also extend the results to the general K user case for $M/N \leq \frac{K-1}{K(K-2)}$. Similar to the $K = 3$ user case, we show that linear interference alignment schemes are sufficient to achieve the information theoretic DoF outer bound in the sense of spatial normalization. Spatial normalization is introduced in [2] and refers to assumption that DoF are normalized with respect to the spatial dimensions, wherein we allow symbol extensions in the spatial dimension through a scaling of antennas. Such a scaling of antennas would create generic structureless channel matrices (instead of structured (block-diagonal) channel matrices created by symbol extensions in time and frequency dimensions), which facilitates the achievability proof. The spatially normalized DoF formulation is motivated by the conjecture that the spatially normalized DoF are spatial scale invariant, e.g., if the number of antennas at every node is scaled up by a factor L, then the overall DoF will also scale by the same factor L. This assumption is relevant mainly for achievability arguments. The information theoretic outer bound part does not need the spatial normalization assumption. Our results are presented in the following lemmas and theorem.

Definition 1: We define the following quantity:

$$d^* = \begin{cases} M, & 0 < \frac{M}{N} \leq \frac{1}{K}, \\ \frac{N}{K}, & \frac{1}{K} \leq \frac{M}{N} \leq \frac{1}{K-1}, \\ \frac{(K-1)M}{K}, & \frac{1}{K-1} \leq \frac{M}{N} \leq \frac{K}{K^2-K-1}, \\ \frac{(K-1)N}{K^2-K-1}, & \frac{K}{K^2-K-1} \leq \frac{M}{N} \leq \frac{K-1}{K(K-2)}. \end{cases} \tag{165}$$

Lemma 4: For the $K \geq 4$ user $M_T \times M_R$ MIMO interference channel where each TX has M_T and each RX has M_R antennas, if $M/N \leq \frac{K-1}{K(K-2)}$, then the DoF per user are outer bounded by $d \leq d^*$.

Proof: Since the idea behind the proof for this lemma directly follows from the $K = 3$ user case [2] yet requires much more cumbersome analysis, we defer the proof to Appendix B.1 and B.2.

Lemma 5: For the $K \geq 4$ user $M_T \times M_R$ MIMO interference channel where each TX has M_T and each RX has M_R antennas, if $M/N \leq \frac{K-1}{K(K-2)}$, then d^* DoF per user are achievable subject to spatial normalization.

Proof: The idea behind the proof for this lemma directly follows from the $K = 3$ user case [2]. Thus, we defer the proof to Appendix B.3.

Theorem 3: For the $K \geq 4$ user $M_T \times M_R$ MIMO interference channel where each TX has M_T and each RX has M_R antennas, if $M/N \leq \frac{K-1}{K(K-2)}$, then the spatially normalized DoF value per user is given by $d = d^*$.

Proof: The proof of this theorem directly follows from Lemma 4 and Lemma 5.

B. The Decomposition DoF Bound Achieving the Information Theoretic DoF Outer Bound

Next, let us consider the case where the value of M/N falls into the interval $\left(\frac{K-1}{K(K-2)}, 1\right)$. In Section V, Theorem 1 includes the cases of $M_T/M_R \in \mathcal{P}_1 \cup \mathcal{P}_2 \cup \mathcal{P}_3 = \mathcal{P}$ for the $K = 4$ setting, where $\mathcal{P}_1 = \{\frac{M_T}{M_R} | \frac{1}{2} \leq \frac{M_T}{M_R} < 1,\ M_T, M_R \in \mathbb{Z}^+,\ M_R \leq 20\}$, $\mathcal{P}_2 = [2/5, 1/2)$, and $\mathcal{P}_3 = \{\frac{8}{21}\} \cup \{\frac{2c-1}{5c-2} | c \in \mathbb{Z}^+, c \geq 2\}$. Next, we will discuss the remaining cases of the $K = 4$ setting outside $\mathcal{P}$ and the $K > 4$ settings, to shed light on the insights behind DoF results of the general K-user $M_T \times M_R$ MIMO Gaussian interference channel. We begin with the $M_T < M_R$ setting.

First, it can be easily shown that any information theoretical DoF outer bounds for the $K = K_0$ setting are valid information theoretical DoF outer bounds for the $K > K_0$ settings. This is because increasing the number of users in a network cannot increase the symmetric capacity per user. Thus, for $M_T/M_R \in \mathcal{P}$ cases, the DoF value per user $d = \frac{MN}{M+N}$ is also optimal for $K > 4$ settings.

Second, notice that for $K = 4$ setting, the left boundary value of $\mathcal{P}_2$, i.e., $M_T/M_R = 2/5$ is obtained in Section V by showing that we never need *successive two intermediate bounds* by applying the genie chains approach for the $K = 4$, $M_T/M_R \geq 2/5$ setting. This argument implies that at a given RX looking at one interferer, we do not need to exhaust the other *two unintended receivers* to produce two successive intermediate bounds. Therefore, if we apply the genie chains approach, then for the general $K \geq 4$ user setting, the decomposition DoF bound $d = \frac{MN}{M+N}$ per user is expected to be the information theoretic DoF outer bound as well, as long as we never need $K - 2$ *successive intermediate bounds*, i.e., we only need up to $K-3$ successive intermediate bounds. Equivalently, this implies that through the union of $N-(K-2)M$ dimensional observations of a given interferer at each unintended RX, we have a total of $(K-2)(N-(K-2)M)$ dimensional observations of that interferer, which contribute to recovery of the transmit signal vector of that TX up to $K-3$ times. Thus, we have the following inequality:

$$(K-2)(N-(K-2)M) \leq (K-3)M \tag{166}$$

$$\Rightarrow \frac{M}{N} \geq \frac{K-2}{K^2-3K+1}. \tag{167}$$

Therefore, for any cases of $\frac{M_T}{M_R} \in \left[\frac{K-2}{K^2-3K+1}, 1\right)$, we expect that the DoF value per user $d = \frac{MN}{M+N}$ is also the information theoretic DoF outer bound. Notice that when the value of K grows, $\frac{K-2}{K^2-3K+1}$ approaches zero, such that $\left[\frac{K-2}{K^2-3K+1}, 1\right)$ becomes the dominant interval.

What remains to be shown is the regime $\frac{M_T}{M_R} \in \left(\frac{K-1}{K(K-2)}, \frac{K-2}{K^2-3K+1}\right)$, which comprises of both regime 1 and regime 2 partially. In Theorem 1, we only show the DoF results of cases $\frac{M_T}{M_R} \in \mathcal{P}_3$ for the $K = 4$ setting, and the point sequence in $\mathcal{P}_3$ converges to the boundary $2/5$. Although there are infinitely many number of points in the set $\mathcal{P}_3$, they have only zero measure, i.e., $\mathcal{P}_3$ is not a dense set. Thus, the DoF characterization in this sub-regime is still open in general. Although we conjectured $\frac{MN}{M+N}$ is still the DoF value per user in [1], feasibility analysis and numerical evidence in [12] indicate that this is not the case. Moreover, [12] provides evidence that part of this regime behaves like the piece-wise linear regime discussed in the previous section.

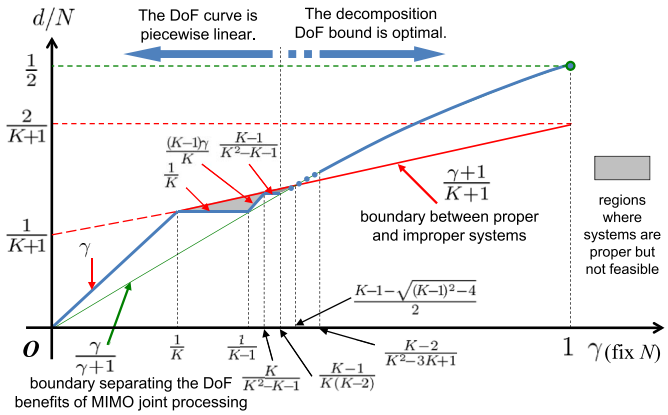

Fig. 5. d/N as a function of $\gamma = M/N$.

Finally, for all known DoF results, the information theoretic DoF satisfy the principle of duality. That is, the original channel and its reciprocal channel both have the same number of DoF.

C. Observations

Now let us collect all the DoF results in Figure 5. There are three curves in Figure 5. As we introduce in Section I, the red line and the green curve are the DoF counting outer bound and the DoF decomposition inner bound, respectively. It can be seen that if $M/N \leq \frac{K-1}{K(K-2)} = \gamma_0$, then the DoF curve is a piecewise linear function depending on M and N alternately, which is similar to that of the $K = 3$ setting [2]. Intuitively, it means that there are antenna redundancies at either TX or RX sides except when $\frac{M}{N} = \frac{1}{K}$ and $\frac{K}{K^2-K-1}$. The achievability relies on linear interference alignment schemes without symbol extensions or with finite number of symbol extensions (through numerical tests), i.e., asymptotical alignment is unnecessary. On the other hand, if $\frac{M}{N} \geq \gamma_0$, then the decomposition is expected to be optimal in many cases, and the achievability relies on the asymptotic alignment. Note that when the value of K grows, this cross point moves towards to the left, so that the interval $[\gamma_0, 1)$ becomes the dominant, and the DoF decomposition bound is optimal. Figure 5 also implies that whenever the decomposition bound is larger than the counting bound, the DoF decomposition bound is information theoretic optimal.

VIII. Conclusion

In this paper, we propose a novel tool, called genie chains, to study the information theoretic DoF outer bound of wireless interference networks, which essentially translates an information theoretic DoF outer bound problem into a much simpler linear algebraic problem. While this new tool has wide applications in various wireless interference networks, in this paper, we mainly study the MIMO interference channel as a typical example, followed by several other special examples including the many-to-one MIMO interference channel and the MIMO X channel.

Appendix A
Proof of Lemma 1

Since $B_i(\mathbf{L})$, $B_j(\mathbf{L})$ both represent the basis of the same subspace $\mathbf{L}$, there exists an invertible $|\mathbf{L}| \times |\mathbf{L}|$ square matrix A where $|\mathbf{L}|$ is the number of dimensions of the subspace $\mathbf{L}$, so that $B_i(\mathbf{L}) = A \cdot B_j(\mathbf{L})$. Then we have

$$\begin{aligned}
\hbar(B_i(\mathbf{L})\mathbf{X}) &= \hbar(A \cdot B_j(\mathbf{L})\mathbf{X}) && (168)\\
&= h(A \cdot B_j(\mathbf{L})\mathbf{X} + \mathbf{Z}) && (169)\\
&= h(A(B_j(\mathbf{L})\mathbf{X} + A^{-1}\mathbf{Z})) && (170)\\
&= h(B_j(\mathbf{L})\mathbf{X} + A^{-1}\mathbf{Z}) + \log|\det(A)| && (171)\\
&= h(B_j(\mathbf{L})\mathbf{X} + A^{-1}\mathbf{Z}) + o(\log\rho) && (172)
\end{aligned}$$

where $\log|\det(A)|$ is a constant that does not depend on the SNR, ρ. Notice that $\mathbf{Z} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$, thus $A^{-1}\mathbf{Z} \sim \mathcal{CN}(\mathbf{0}, \mathbf{K})$ where $\mathbf{K} = A^{-1}(A^{-1})^H$. Since $\tilde{\mathbf{Z}}$ is an independent noise vector, we have

$$\begin{aligned}
&h(B_j(\mathbf{L})\mathbf{X} + A^{-1}\mathbf{Z})\\
&\quad = h(B_j(\mathbf{L})\mathbf{X} + A^{-1}\mathbf{Z} + \tilde{\mathbf{Z}}|\tilde{\mathbf{Z}}) && (173)\\
&\quad \leq h(B_j(\mathbf{L})\mathbf{X} + A^{-1}\mathbf{Z} + \tilde{\mathbf{Z}}). && (174)
\end{aligned}$$

On the other hand, since $B_j(\mathbf{L})\mathbf{X} + A^{-1}\mathbf{Z} + \tilde{\mathbf{Z}}$ is a degraded version of $B_j(\mathbf{L})\mathbf{X} + A^{-1}\mathbf{Z}$, we have

$$\begin{aligned}
0 &\leq I(B_j(\mathbf{L})\mathbf{X}; B_j(\mathbf{L})\mathbf{X} + A^{-1}\mathbf{Z})\\
&\quad - I(B_j(\mathbf{L})\mathbf{X}; B_j(\mathbf{L})\mathbf{X} + A^{-1}_{|\mathbf{L}|\times|\mathbf{L}|}\mathbf{Z} + \tilde{\mathbf{Z}}) && (175)\\
&= h(B_j(\mathbf{L})\mathbf{X} + A^{-1}\mathbf{Z})\\
&\quad - h(B_j(\mathbf{L})\mathbf{X} + A^{-1}\mathbf{Z}|B_j(\mathbf{L})\mathbf{X})\\
&\quad - h(B_j(\mathbf{L})\mathbf{X} + A^{-1}\mathbf{Z} + \tilde{\mathbf{Z}})\\
&\quad + h(B_j(\mathbf{L})\mathbf{X} + A^{-1}\mathbf{Z} + \tilde{\mathbf{Z}}|B_j(\mathbf{L})\mathbf{X}) && (176)\\
&= h(B_j(\mathbf{L})\mathbf{X} + A^{-1}\mathbf{Z})\\
&\quad - h(B_j(\mathbf{L})\mathbf{X} + A^{-1}\mathbf{Z} + \tilde{\mathbf{Z}})\\
&\quad - h(A^{-1}\mathbf{Z}) + h(A^{-1}\mathbf{Z} + \tilde{\mathbf{Z}}) && (177)\\
&= h(B_j(\mathbf{L})\mathbf{X} + A^{-1}\mathbf{Z})\\
&\quad - h(B_j(\mathbf{L})\mathbf{X} + A^{-1}\mathbf{Z} + \tilde{\mathbf{Z}})\\
&\quad + \log(\det(\tilde{\mathbf{K}}\mathbf{K}^{-1} + \mathbf{I})). && (178)
\end{aligned}$$

Combining (174) and (178) produces

$$\begin{aligned}
&h(B_j(\mathbf{L})\mathbf{X} + A^{-1}\mathbf{Z})\\
&\quad \leq h(B_j(\mathbf{L})\mathbf{X} + A^{-1}\mathbf{Z} + \tilde{\mathbf{Z}}) && (179)\\
&\quad \leq h(B_j(\mathbf{L})\mathbf{X} + A^{-1}\mathbf{Z})) + \log(\det(\tilde{\mathbf{K}}\mathbf{K}^{-1} + \mathbf{I})) && (180)
\end{aligned}$$

where $\log(\det(\tilde{\mathbf{K}}\mathbf{K}^{-1} + \mathbf{I}))$ is a constant that does not depend on ρ. Thus we have

$$\begin{aligned}
&h(B_j(\mathbf{L})\mathbf{X} + A^{-1}\mathbf{Z})\\
&\quad = h(B_j(\mathbf{L})\mathbf{X} + A^{-1}\mathbf{Z} + \tilde{\mathbf{Z}}) + o(\log\rho). && (181)
\end{aligned}$$

Following the similar procedure, we also obtain

$$\begin{aligned}
&h(B_j(\mathbf{L})\mathbf{X} + \tilde{\mathbf{Z}})\\
&\quad = h(B_j(\mathbf{L})\mathbf{X} + A^{-1}\mathbf{Z} + \tilde{\mathbf{Z}}) + o(\log\rho). && (182)
\end{aligned}$$

Thus, we have

$$h(B_j(\mathbf{L})\mathbf{X} + A^{-1}\mathbf{Z}) = h(B_j(\mathbf{L})\mathbf{X} + \tilde{\mathbf{Z}}) + o(\log\rho). \quad (183)$$

Finally, substituting (183) into (172), we obtain

$$\hbar(B_i(\mathbf{L})\mathbf{X}) = h(B_j(\mathbf{L})\mathbf{X} + A^{-1}\mathbf{Z}) + o(\log\rho) \quad (184)$$
$$= h(B_j(\mathbf{L})\mathbf{X} + \tilde{\mathbf{Z}}) + o(\log\rho). \quad (185)$$

Thus, we complete the proof of Lemma 1.

Appendix B
DoF of the $\frac{M}{N} \leq \frac{K-1}{K(K-2)}$ Setting for the K User $M_T \times M_R$ MIMO Interference Channel

For the K user $M_T \times M_R$ MIMO interference channel, the DoF value is a piecewise linear function of M and N alternately if $\frac{M}{N} \leq \frac{K-1}{K(K-2)}$ where $M = \min(M_T, M_R)$ and $N = \max(M_T, M_R)$. As shown in Theorem 3, the DoF value per user is given by:

$$d = \begin{cases} M, & 0 < \frac{M}{N} \leq \frac{1}{K}, \\ \frac{N}{K}, & \frac{1}{K} \leq \frac{M}{N} \leq \frac{1}{K-1}, \\ \frac{(K-1)M}{K}, & \frac{1}{K-1} \leq \frac{M}{N} \leq \frac{K}{K^2-K-1}, \\ \frac{(K-1)N}{K^2-K-1}, & \frac{K}{K^2-K-1} \leq \frac{M}{N} \leq \frac{K-1}{K(K-2)}. \end{cases} \quad (186)$$

In this section, we investigate the DoF converse and the achievability. Note that all the techniques applied in proofs presented in this section follow similarly from the $K = 3$ user interference channel setting [2]. In the following, we first show the outer bound for the $M_T < M_R$ and $M_T > M_R$ settings, respectively, and then provide the achievability proof.

A. The Information Theoretic DoF Outer Bound for $M_T < M_R$

We consider the $M_T < M_R$ setting in this section. $M_T < M_R$ implies that $M = M_T$, $N = M_R$.

Among the four regions shown in (186), the DoF outer bound of the first three regions can be established by the single user DoF bound and the cooperation DoF outer bound. Specifically, let us consider $\frac{M}{N} \in (0, \frac{1}{K}]$ first. The DoF outer bound $d \leq M$ follows trivially from the single user bound. Next, consider $\frac{M}{N} \in [\frac{1}{K}, \frac{1}{K-1}]$ and $\frac{M}{N} \in [\frac{1}{K-1}, \frac{K}{K^2-K-1}]$. Since collaboration among the users does not decrease the capacity region, we allow the $K-1$ users from User 2 to User K to cooperate as one user, such that the network becomes a two user MIMO interference channel where the two transmitters have M and $(K-1)M$ antennas respectively, and corresponding receivers have N and $(K-1)N$ antennas, respectively. The sum DoF of this network, as reported in [11], are outer bounded by $\min(\max((K-1)M, N), \max(M, (K-1)N)$, which produces the desired DoF bound per user $d \leq \max((K-1)M, N)/K = \frac{N}{K}$ if $\frac{M}{N} \in [\frac{1}{K}, \frac{1}{K-1}]$, and $d \leq \max((K-1)M, N)/K = \frac{(K-1)M}{K}$ if $\frac{M}{N} \in [\frac{1}{K-1}, \frac{K}{K^2-K-1}]$.

Now let us focus on the remaining case $\frac{M}{N} \in [\frac{K}{K^2-K-1}, \frac{K-1}{K(K-2)}]$. We apply similar linear transformations as introduced in [2]. Consider RX 2, which is able to decode its own message W_2 due to the reliable communications

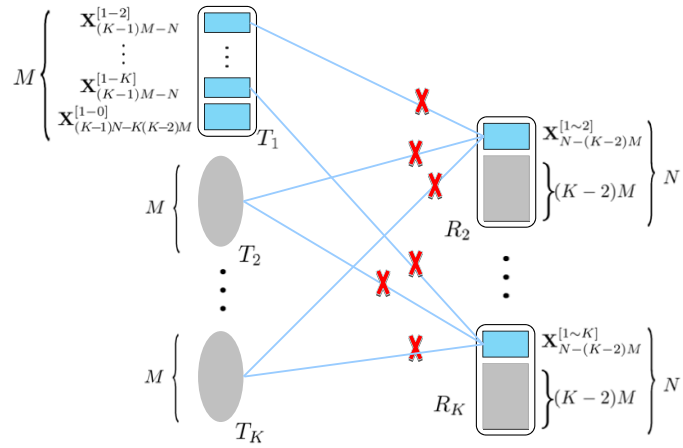

Fig. 6. The Linear Transformations for the $M \times N$ Case, $\frac{M}{N} \in [\frac{K}{K^2-K-1}, \frac{K-1}{K(K-2)}]$ where the red cross symbols stand for nulling. Note that we only show the transformed channels for a clear presentation.

assumption. Thus, after removing the desired signal carrying message W_2, RX 2 obtains an N-dimensional interference vector space $\mathbf{S}^{[2]}$. By zero forcing the interference from TX 3 to TX K, RX 2 extracts the exposed subspace $\mathbf{X}^{[1\sim2]}_{N-(K-2)M}$ from $\mathbf{S}^{[2]}$. This can be done by left-multiplying the received signal with an invertible $N \times N$ matrix whose first $N-(K-2)M$ rows are orthogonal to the channel vectors from each antenna of TX $3, \ldots, K$ to RX 2, and last $(K-2)M$ rows are the last $(K-2)M$ rows of the $N \times N$ identity matrix, as shown in Figure 6. After this operation, the first $N-(K-2)M$ antennas at RX 2 only hear TX 1. Similarly, we proceed to RX $3, \ldots, K$ where we apply linear transformations such that the first $N-(K-2)M$ antennas of each RX only hear TX 1. Therefore, we obtain the exposed subspaces $\mathbf{X}^{[1\sim3]}_{N-(K-2)M}, \ldots, \mathbf{X}^{[1\sim K]}_{N-(K-2)M}$ at RX $3, \ldots, K$, respectively. Now we complete the linear transformations of the RX basis and then we switch to TX 1. We multiply an $M \times M$ matrix to the right-hand side of its channel matrix such that the first $(K-1)M-N$ antennas of TX 1 are *not* heard by the first $N-(K-2)M$ antennas at RX 2, the next $(K-1)M-N$ antennas are *not* heard by the first $N-(K-2)M$ antennas of RX 3 and so forth. This can be done by choosing the first $(K-1)M-N$ columns of the transformation matrix as the basis of the null space of the channel matrix from the M antennas of TX 1 to the first $N-(K-2)M$ antennas at RX 2, the next $(K-1)M-N$ columns associated with RX 3 and so forth. Note that the dimension matches as $(K-1)M-N = M-[N-(K-2)M]$, and thus corresponding transmit signals of TX 1, not heard by RX k, will be heard by RX k' where $k' \in \mathcal{K} \setminus \{1, k\}$. Continuing to RX K, we fix the directions of the first $(K-1) \times [(K-1)M-N]$ antennas of TX 1, and leave the last $M-[(K-1)^2M-(K-1)N] = (K-1)N-K(K-2)M$ antennas which do not need change of basis and these columns can be chosen as the corresponding columns of the $M \times M$ identity matrix. Notice that $(K-1)N-K(K-2)M \geq 0$ as $\frac{M}{N} \leq \frac{K-1}{K(K-2)}$. For brevity, we label the corresponding transmit signals from TX 1 to RX k after the invertible linear transformations as $\mathbf{X}^{[1-k]}_{(K-1)M-N}$ where $k \in \mathcal{K} \setminus \{1\}$ and the

signals from the last $(K-1)N-K(K-2)M$ antennas of TX 1 are represented as $\mathbf{X}^{[1-0]}_{(K-1)N-K(K-2)M}$ (See Figure 6). Now we also complete the linear transformations at the TX side.

Note that $\mathbf{X}^{[1-2]}_{(K-1)M-N}$ has $(K-1)M-N$ dimensions. Also note that $\mathbf{X}^{[1-2]}_{(K-1)M-N}$ is linearly independent of the exposed subspace $\mathbf{X}^{[1\sim 2]}_{N-(K-2)M}$. This is because $\mathbf{X}^{[1-2]}_{(K-1)M-N}$ is determined by the channel matrices associated with RX 2. Thus the genie signal $\mathbf{X}^{[1-2]}_{(K-1)M-N}$ allows RX 2 to decode all the messages subject to noise distortion. From Fano's inequality, we have

$$\begin{aligned}
&KnR - n\epsilon_n && (187)\\
&\leq: nN\log\rho + \hbar(\mathbf{X}^{[1-2]^n}_{(K-1)M-N}|\mathbf{S}^{[2]^n}) && (188)\\
&\leq: nN\log\rho + \hbar(\mathbf{X}^{[1-2]^n}_{(K-1)M-N}|\mathbf{X}^{[1\sim 2]^n}_{N-(K-2)M}) && (189)\\
&\leq: nN\log\rho + \hbar(\mathbf{X}^{[1-2]^n}_{(K-1)M-N}|\mathbf{X}^{[1-3]^n}_{(K-1)M-N},\cdots\\
&\qquad \ldots,\mathbf{X}^{[1-K]^n}_{(K-1)M-N},\mathbf{X}^{[1-0]^n}_{(K-1)N-K(K-2)M}) && (190)
\end{aligned}$$

where (190) follows from Property (P2) in Lemma 2. For compactness, we omit the subscript which denotes dimension and is clear from the context. Thus, we can rewrite the equation above as

$$\begin{aligned}
&KnR - n\epsilon_n\\
&\leq: nN\log\rho + \hbar(\mathbf{X}^{[1-2]^n}|\mathbf{X}^{[1-3]^n},\ldots\\
&\qquad \ldots,\mathbf{X}^{[1-K]^n},\mathbf{X}^{[1-0]^n}). && (191)
\end{aligned}$$

Similarly at RX k where $k\in\{3,\cdots,K\}$, a genie provides $\mathbf{X}^{[1-i]^n}$ to RX k such that it can decode all the messages subject to noise distortion. Thus we have

$$\begin{aligned}
&KnR - n\epsilon_n\\
&\leq: nN\log\rho + \hbar(\mathbf{X}^{[1-k]^n}|\mathbf{X}^{[1-2]^n},\ldots\\
&\qquad \ldots,\mathbf{X}^{[1-(k-1)]^n},\mathbf{X}^{[1-(k+1)]^n},\ldots,\mathbf{X}^{[1-K]^n},\mathbf{X}^{[1-0]^n}) && (192)
\end{aligned}$$

Adding (191) and (192), we have:

$$\begin{aligned}
&(K-1)KnR - n\epsilon_n && (193)\\
&\leq: \sum_{i=2}^{K}\hbar(\mathbf{X}^{[1-i]^n}|\mathbf{X}^{[1-2]^n},\ldots\\
&\qquad \ldots,\mathbf{X}^{[1-(i-1)]^n},\mathbf{X}^{[1-(i+1)]^n},\ldots,\mathbf{X}^{[1-K]^n},\mathbf{X}^{[1-0]^n})\\
&\qquad + (K-1)nN\log\rho && (194)\\
&\leq: (K-1)nN\log\rho + \hbar(\mathbf{X}^{[1-2]^n},\ldots,\mathbf{X}^{[1-K]^n}|\mathbf{X}^{[1-0]^n})\\
&\leq: (K-1)nN\log\rho + \hbar(\mathbf{X}^{[1]^n}) && (195)\\
&\leq: (K-1)nN\log\rho + nR. && (196)
\end{aligned}$$

Rearranging the terms in (196) we obtain:

$$(K^2-K-1)nR - n\epsilon_n \leq: (K-1)nN\log\rho. \quad (197)$$

By letting $n\to\infty$ first and then $\rho\to\infty$, we have the desired DoF outer bound

$$d \leq \frac{(K-1)N}{K^2-K-1}. \quad (198)$$

Thus, we complete the outer bound proof for the $M_T < M_R$ setting. ■

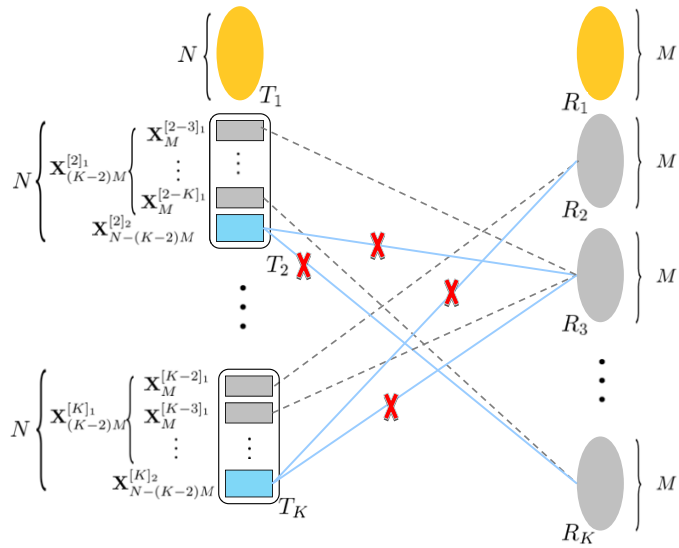

Fig. 7. The First Stage of the Linear Transformations for the $N\times M$ Case, $\frac{M}{N}\in[\frac{K}{K^2-K-1},\frac{K-1}{K(K-2)}]$ where the red cross stands for nulling and the dashed lines stand for an identity matrix. Note that we only show the transformed channels for a clear presentation.

B. The Information Theoretic DoF Outer Bound for $M_T > M_R$

In this section, we consider the reciprocal $M_T > M_R$ setting, where $N=M_T$, $M=M_R$. For the K user $N\times M$ MIMO interference channel, again if $\frac{M}{N}\leq\frac{K}{K^2-K-1}$, the DoF outer bound can be directly obtained by the single-user DoF bound and the cooperation DoF bound, as stated in Appendix B-A. What remains to be shown is the case of $\frac{M}{N}\in[\frac{K}{K^2-K-1},\frac{K-1}{K(K-2)}]$. We will use a two-stage approach as follows.

Stage 1: In the first stage, we consider Users $2,\ldots,K$. Let us first consider TX 2. Denote its first $(K-2)M$ antennas as layer 1 symbol $\mathbf{X}^{[2]_1}_{(K-2)M}$ and last $N-(K-2)M$ antennas as layer 2 symbol $\mathbf{X}^{[2]_2}_{N-(K-2)M}$. Now we take invertible linear transforms on the symbols of these two layers. As shown in Figure 7, for layer 1 symbol $\mathbf{X}^{[2]_1}_{(K-2)M}=[\mathbf{X}^{[2-3]_1}_{M};\ldots;\mathbf{X}^{[2-K]_1}_{M}]$, invert its channel matrices to RX k where $k\in\{3,\ldots,K\}$ such that $\mathbf{X}^{[2-k]_1}_{M}$ is only heard by RX k. For layer 2 symbol $\mathbf{X}^{[2]_2}_{N-(K-2)M}$, let it *not* be heard by all RX $3,\ldots,K$ by nulling the channel matrices to those receivers from TX 2. Once we complete the transformation at TX 2, we can apply similar transformations at TX $k\in\{3,\ldots,K\}$, i.e., dividing $\mathbf{X}^{[k]}$ to layer 1 symbols $\mathbf{X}^{[k]_1}_{(K-2)M}$ associated with the first $(K-2)M$ antennas which sequentially have projections to RX $k'\in\mathcal{K}\setminus\{1,k\}$, and the remaining layer 2 symbols $\mathbf{X}^{[k]_2}_{N-(K-2)M}$ which are not heard by RX $k'\in\mathcal{K}\setminus\{1,k\}$. Now we have that if a genie provides $K-2$ messages $W^{[1]},W^{[4]},\ldots,W^{[K]}$ to RX 3, it can remove the interference signals carrying those messages and only hear $\mathbf{X}^{[2-3]_1}_{M}$. Hence, if a genie further provides $\mathbf{X}^{[2]}\setminus\mathbf{X}^{[2-3]_1}_{M}$ to RX 3, it is then able to decode $W^{[2]}$ subject to noise distortion. That is, providing $\mathbf{G}_3=\{W^{[1]},W^{[4]},\ldots,W^{[K]},\mathbf{X}^{[2]}\setminus\mathbf{X}^{[2-3]_1}_{M}\}$ to RX 3 allows it to decode all the messages subject to

noise distortion. From Fano's inequality, we have

$$\begin{aligned}
&KnR - n\epsilon_n \\
&\quad \leq: nM\log\rho \\
&\qquad + \hbar(W^{[1]}, W^{[4]}, \ldots, W^{[K]}, \mathbf{X}^{[2]^n} \setminus \mathbf{X}_M^{[2-3]_1^n} | \mathbf{S}^{[3]^n}) \\
&\quad \leq: nM\log\rho + n(K-2)R + \hbar(\mathbf{X}^{[2]^n} \setminus \mathbf{X}_M^{[2-3]_1^n} | \mathbf{X}_M^{[2-3]_1^n}) \\
&\quad \leq: nM\log\rho + n(K-2)R + \hbar(\mathbf{X}^{[2]^n}) - \hbar(\mathbf{X}_M^{[2-3]_1^n}) \\
&\quad \leq: nM\log\rho + n(K-1)R - \hbar(\mathbf{X}_M^{[2-3]_1^n}) \\
&\quad \Rightarrow nR - n\epsilon_n \leq: nM\log\rho - \hbar(\mathbf{X}_M^{[2-3]_1^n}). \qquad (199)
\end{aligned}$$

Following the same line, if a genie provides

$$\begin{aligned}
\mathbf{G}_k = \{&W^{[1]}, W^{[3]}, \ldots, W^{[k-1]}, W^{[k+1]}, \ldots \\
&\ldots, W^{[K]}, \mathbf{X}^{[2]} \setminus \mathbf{X}_M^{[2-k]_1}\}
\end{aligned}$$

to RX $k \in \{4, \ldots, K\}$, RX k can also decode all messages subject to noise distortion. Therefore, we have the sum rate inequality as follows:

$$nR - n\epsilon_n \leq: nM\log\rho - \hbar(\mathbf{X}_M^{[2-k]_1^n}). \qquad (200)$$

Adding all $K-2$ sum rate inequalities associated with RX $k \in \{3, \cdots, K\}$ above, we have:

$$\begin{aligned}
&(K-2)nR - n\epsilon_n \qquad (201) \\
&\quad \leq: (K-2)nM\log\rho - \sum_{k=3}^{K} \hbar(\mathbf{X}_M^{[2-k]_1^n}) \qquad (202) \\
&\quad \leq: (K-2)nM\log\rho - \hbar(\mathbf{X}_M^{[2-3]_1^n}, \ldots, \mathbf{X}_M^{[2-K]_1^n}) \qquad (203) \\
&\quad =: (K-2)nM\log\rho - \hbar(\mathbf{X}_{(K-2)M}^{[2]_1^n}). \qquad (204)
\end{aligned}$$

After obtaining the inequality above by considering layer 1 symbols $\mathbf{X}_{(K-2)M}^{[2]_1}$ at TX 2, we proceed to TX $k \in \{3, \ldots, K\}$, and obtain the following sum rate inequality similarly:

$$\begin{aligned}
(K-2)nR - n\epsilon_n \leq: (K-2)nM\log\rho - \hbar(\mathbf{X}_{(K-2)M}^{[k]_1^n}), \\
k \in \{3, \ldots, K\}. \qquad (205)
\end{aligned}$$

Adding (204) and (205), we have

$$\begin{aligned}
&(K-1)(K-2)nR - n\epsilon_n \\
&\quad \leq: (K-1)(K-2)nM\log\rho - \sum_{k=2}^{K} \hbar(\mathbf{X}^{[k]_1^n}) \qquad (206)
\end{aligned}$$

where we omit the subscript that denotes dimension, for convenience.

Stage 2: Next we consider the second stage where we focus on RX 1. Note that all the linear transformations in the first stage are *not* associated with the channel matrix to RX 1 which guarantees that all the transmit symbols from TX $2, \ldots, K$ are still generic for RX 1. Again, we will apply invertible linear transformations at both the TX side and the RX side. We first describe the linear transformation at RX 1. For $k \in \{3, \ldots, K\}$, as each layer 2 symbol $\mathbf{X}_{N-(K-2)M}^{[k]_2}$ has $N-(K-2)M$ dimensions, colored in blue in Figure 8, RX 1 can find $M-(N-(K-2)M) = (K-1)M-N$ dimensions that do not hear the layer 2 symbol from one TX through zero forcing. Therefore, at RX 1, let the first $(K-1)M-N$ antennas not hear $\mathbf{X}_{N-(K-2)M}^{[3]_2}$, the next $(K-1)M-N$ antennas not hear $\mathbf{X}_{N-(K-2)M}^{[4]_2}$ and so on. That is, the kth $(K-1)M-N$ antennas at RX 1 do not hear $\mathbf{X}_{N-(K-2)M}^{[k]_2}$ where $k \in \{3, \ldots, K\}$. Now we complete the linear transformations at RX 1. Note that we have only considered the first $(K-2)[(K-1)M-N] < M$ antennas as $\frac{M}{N} \leq \frac{K-1}{K(K-2)}$. Next we consider the layer 2 symbols $\mathbf{X}_{N-(K-2)M}^{[2]_2}$ at TX 2. We invert the channel from its first $(K-2)[(K-1)M-N]$ antennas (the first $K-2$ blue boxes of TX 2 in Figure 8) to the first $(K-2)[(K-1)M-N]$ antennas at RX 1 such that the channel between them becomes an identity matrix. Moreover, at TX 2, the remaining $N-(K-2)M-(K-2)[(K-1)M-N] = (K-1)N-K(K-2)M$ dimensions of $\mathbf{X}_{N-(K-2)M}^{[2]_2}$ (the last blue box of TX 2 in Figure 8) are chosen to be zero forced at the first $(K-2)[(K-1)M-N]$ antennas at RX 1. Denote the symbols of $\mathbf{X}_{N-(K-2)M}^{[2]_2}$ in sequence after linear transforms as $\mathbf{X}_{N-(K-2)M}^{[2]_2} = [\mathbf{X}_{(K-1)M-N}^{[2:1]_2}; \ldots; \mathbf{X}_{(K-1)M-N}^{[2:(K-2)]_2}; \mathbf{X}_{(K-1)N-K(K-2)M}^{[2:0]_2}]$.

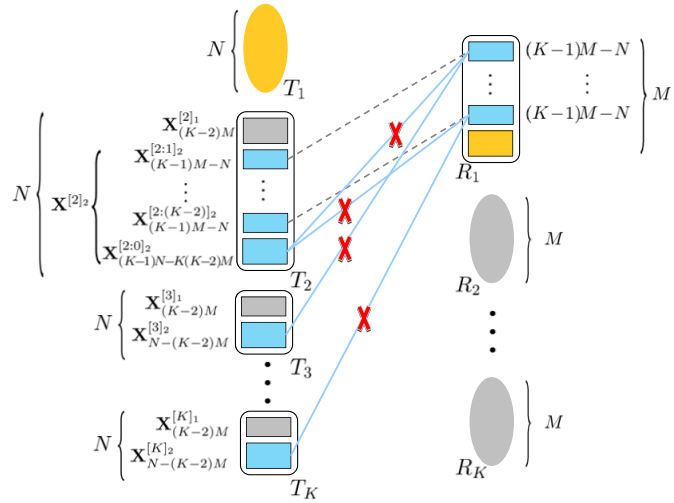

Fig. 8. The Second Stage of the Linear Transformations for the $N \times M$ Case, $\frac{M}{N} \in [\frac{K}{K^2-K-1}, \frac{K-1}{K(K-2)}]$, where the red cross stands for nulling and the dashed lines stand for an identity matrix. Note that we only show the transformed channels for a clear presentation.

At the RX side, owing to our linear transformations, the received signals from $\mathbf{X}_{N-(K-2)M}^{[2]_2}$ seen by the first $(K-2)[(K-1)M-N]$ antennas of RX 1 are given by $[\mathbf{X}_{(K-1)M-N}^{[2:1]_2}; \ldots; \mathbf{X}_{(K-1)M-N}^{[2:(K-2)]_2}]$. Now we finish the linear transformations at the second stage.

When a genie provides to RX 1 the $K-3$ messages $W^{[4]}, \ldots, W^{[K]}$ and layer 1 symbols $\mathbf{X}^{[2]_1}, \mathbf{X}^{[3]_1}$, RX 1 only hears interference caused by layer two symbols $\mathbf{X}_{N-(K-2)M}^{[2]_2}$, $\mathbf{X}_{N-(K-2)M}^{[3]_2}$ from TX 2 and 3. Further, with these genie signals, RX 1 hears clean $\mathbf{X}_{(K-1)M-N}^{[2:1]_2}$ at the first $(K-1)M-N$ antennas of RX 1, where $\mathbf{X}_{N-(K-2)M}^{[3]_2}$ are zero forced. So further giving $(\mathbf{X}_{N-(K-2)M}^{[2]_2} \setminus \mathbf{X}_{(K-1)M-N}^{[2:1]_2})$ allows RX 1 to decode messages $W^{[2]}$ and $W^{[3]}$ subject to noise distortion. Therefore, providing $\mathbf{G}_3 = \{W^{[4]}, \ldots, W^{[K]}, \mathbf{X}^{[2]_1}, \mathbf{X}^{[3]_1}, \mathbf{X}_{N-(K-2)M}^{[2]_2} \setminus \mathbf{X}_{(K-1)M-N}^{[2:1]_2}\}$ to RX 3 allows it to decode all the messages subject to

noise distortion. From Fano's inequality, we have

$$\begin{aligned}
&KnR - n\epsilon_n \\
&\quad \leq: nM\log\rho + \hbar(W^{[4]},\ldots,W^{[K]},\mathbf{X}^{[2]_1},\mathbf{X}^{[3]_1},\ldots \\
&\qquad \ldots,\mathbf{X}^{[2]_2}_{N-(K-2)M} \setminus \mathbf{X}^{[2:1]_2}_{(K-1)M-N}|\mathbf{S}^{[3]^n}) && (207)\\
&\quad \leq: nM\log\rho + (K-3)nR + \hbar(\mathbf{X}^{[3]^n_1}) \\
&\qquad + \hbar(\mathbf{X}^{[2]^n_1},\mathbf{X}^{[2]^n_2}\setminus\mathbf{X}^{[2:1]^n_2}|\mathbf{X}^{[2:1]^n_2}) && (208)\\
&\quad =: nM\log\rho + (K-3)nR + \hbar(\mathbf{X}^{[3]^n_1}) \\
&\qquad + \hbar(\mathbf{X}^{[2]^n}) - \hbar(\mathbf{X}^{[2:1]^n_2}) && (209)\\
&\Rightarrow 2nR - n\epsilon_n \leq: nM\log\rho + \hbar(\mathbf{X}^{[3]^n_1}) - \hbar(\mathbf{X}^{[2:1]^n_2}). && (210)
\end{aligned}$$

Similarly, for $k \in \{4,\ldots,K\}$, if a genie provides to RX 1 the signals set

$$\begin{aligned}
\mathbf{G}_k = \{&W^{[3]},\ldots,W^{[k-1]},W^{[k+1]},\ldots \\
&\ldots,W^{[K]},\mathbf{X}^{[k]^n_1},\mathbf{X}^{[2]}\setminus\mathbf{X}^{[2:(k-2)]_2}_{(K-1)M-N}\},
\end{aligned}$$

RX 1 can also decode all the messages subject to noise distortion. Therefore, we have the sum rate inequality:

$$2nR - n\epsilon_n \leq: nM\log\rho + \hbar(\mathbf{X}^{[k]^n_1}) - \hbar(\mathbf{X}^{[2:k]^n_2}), \quad k \in \{4,\ldots,K\}. \tag{211}$$

In order to make the addition operation shown later simple, we rewrite the inequality for $k = K$, similar to (208), as follows:

$$\begin{aligned}
&3nR - n\epsilon_n \\
&\quad \leq: nM\log\rho + \hbar(\mathbf{X}^{[K]^n_1}) \\
&\qquad + \hbar(\mathbf{X}^{[2]^n_1},\mathbf{X}^{[2]^n_2}\setminus\mathbf{X}^{[2:(K-2)]^n_2}|\mathbf{X}^{[2:(K-2)]^n_2}) && (212)\\
&\quad \leq: nM\log\rho + \hbar(\mathbf{X}^{[K]^n_1}) \\
&\qquad + \hbar(\mathbf{X}^{[2]^n_1},\mathbf{X}^{[2:1]^n_2},\ldots,\mathbf{X}^{[2:(K-3)]^n_2},\mathbf{X}^{[2:0]^n_2}) && (213)\\
&\quad \leq: nM\log\rho + \hbar(\mathbf{X}^{[K]^n_1}) + \hbar(\mathbf{X}^{[2]^n_1}) \\
&\qquad + \sum_{k=1}^{K-3}\hbar(\mathbf{X}^{[2:k]^n_2}) + \hbar(\mathbf{X}^{[2:0]^n_2}). && (214)
\end{aligned}$$

Now adding (210), (211) for $k \in \{4,\cdots,K-1\}$ and (214), we have

$$\begin{aligned}
&2(K-3)nR + 3nR - n\epsilon_n \\
&\quad \leq: (K-2)nM\log\rho + \sum_{k=2}^{K}\hbar(\mathbf{X}^{[k]^n_1}) + \hbar(\mathbf{X}^{[2:0]^n_2}) && (215)
\end{aligned}$$

Finally, adding (206) and (215), we have:

$$\begin{aligned}
&(K^2-K-1)nR - n\epsilon_n \\
&\quad \leq: K(K-2)nM\log\rho + \hbar(\mathbf{X}^{[2:0]^n_2}) && (216)\\
&\quad \leq: K(K-2)nM\log\rho \\
&\qquad + [(K-1)N - K(K-2)M]n\log\rho && (217)\\
&\quad =: (K-1)Nn\log\rho && (218)
\end{aligned}$$

where (217) follows from the fact that $\mathbf{X}^{[2:0]_2}$ has a total of $(K-1)N - K(K-2)M$ dimensions and Property 1 in Lemma 2.

By letting $n \to \infty$ first and then $\rho \to \infty$, we have the desired DoF outer bound

$$d \leq \frac{(K-1)N}{K^2-K-1}. \tag{219}$$

Thus, we complete the outer bound proof for the $M_T > M_R$ setting. ■

C. The DoF Achievability

We will show that linear beamforming at the transmitters and zero forcing at the receivers are sufficient to achieve the optimal DoF values. Due to the duality of linear schemes, we only need to consider the $M_T < M_R$ setting, i.e., $M = M_T,\ N = M_R$.

We begin with the first two cases. First, if $\frac{M}{N} \in (0, \frac{1}{K}]$, i.e., $KM \leq N$, each RX has enough antennas to distinguish all the transmit signals from all transmitters. Thus, each user can achieve its interference-free DoF whose value is given by $d = \min(M,N) = M$. Second, if $\frac{M}{N} \in [\frac{1}{K}, \frac{1}{K-1}]$, N sum DoF are achievable because each RX, after decoding its own message and subtracting the signal carrying that message, still has enough antennas to distinguish all the interference signals owing to $(K-1)M \leq N$. Thus, the DoF value $d = N/K$ per user is achievable.

Next, we consider the remaining two cases $\frac{M}{N} \in [\frac{1}{K-1}, \frac{K}{K^2-K-1}]$ and $\frac{M}{N} \in [\frac{K}{K^2-K-1}, \frac{K-1}{K(K-2)}]$. We want to show that $d = \frac{(K-1)M}{K}$ and $d = \frac{(K-1)N}{K^2-K-1}$ are achievable for these two cases, respectively. To do so, we first prove $d = \frac{(K-1)M}{K} = \frac{(K-1)N}{K^2-K-1}$ are achievable at $\frac{M}{N} = \frac{K}{K^2-K-1}$. Then by increasing the RX antenna redundancies N such that M/N falls into the region $[\frac{1}{K-1}, \frac{K}{K^2-K-1}]$, the achievability of $d = \frac{(K-1)M}{K}$ DoF should remain. Similarly, by increasing the TX antenna redundancies M such that M/N falls into the region $[\frac{K}{K^2-K-1}, \frac{K-1}{K(K-2)}]$, the achievability of $d = \frac{(K-1)N}{K^2-K-1}$ should not be affected as well.

We first investigate the case $\frac{M}{N} = \frac{K}{K^2-K-1}$, i.e., $(M,N) = (\beta K, \beta(K^2-K-1))$ where $\beta \in \mathbb{Z}^+$. Our goal is to show $d = \frac{(K-1)M}{K} = \frac{(K-1)N}{K^2-K-1} = \beta(K-1)$ DoF are achievable. At each time slot, TX $k \in \mathcal{K}$ sends $\beta(K-1)$ independent symbols using a $\beta K \times \beta(K-1)$ beamforming matrix $\mathbf{V}^{[k]} = [\mathbf{V}^{[k]}_1,\ldots,\mathbf{V}^{[k]}_{K-1}]$ where each block $\mathbf{V}^{[k]}_i,\ i \in \{1,\cdots,K-1\}$ is a $\beta K \times \beta$ matrix. In the $\beta(K^2-K-1)$ dimensional vector space at each RX, the desired signal will occupy $\beta(K-1)$ dimensions, thus leaving only a subspace with $\beta(K^2-K-1)-\beta(K-1) = \beta((K-1)^2-1)$ dimensions to accommodate a total of $\beta(K-1)^2$ interference symbols. Therefore, we need to align $\beta(K-1)^2-\beta((K-1)^2-1) = \beta$ dimensional interference at each RX. Specifically, we design the beamforming vectors in the following way to satisfy these alignment constraints. At RX 1, β interference vectors from TX 2 are aligned into the subspace spanned by $(K-2)\beta$ interference vectors, β from each of TX $3,\ldots,K$, respectively. This operation gives us one alignment equation:

$$\mathbf{H}^{[12]}\mathbf{V}^{[2]}_1 = -(\mathbf{H}^{[13]}\mathbf{V}^{[3]}_1 + \cdots + \mathbf{H}^{[1K]}\mathbf{V}^{[K]}_1) \tag{220}$$

$$\Rightarrow \underbrace{[\mathbf{H}^{[12]}\ \mathbf{H}^{[13]}\ \ldots\ \mathbf{H}^{[1K]}]}_{\bar{\mathbf{H}}_1}\underbrace{\begin{bmatrix}\mathbf{V}^{[2]}_1\\ \mathbf{V}^{[3]}_1\\ \vdots\\ \mathbf{V}^{[K]}_1\end{bmatrix}_{\beta K(K-1)\times\beta}}_{\bar{\mathbf{V}}_1} \tag{221}$$

$$= \mathbf{O}. \tag{222}$$

Since $\bar{\mathbf{H}}_1$ is a $\beta(K^2-K-1)\times\beta K(K-1)$ generic matrix which only consists of interference carrying channel matrices from TX $2,\ldots,K$ to RX 1, $\bar{\mathbf{V}}_1$ can be obtained as the basis vectors of the null space of $\bar{\mathbf{H}}_1$, and thus all $\mathbf{V}_1^{[2]},\ldots,\mathbf{V}_1^{[K]}$ are obtained. Note that from the equation above which is associated with RX 1, we have determined the directions of $\beta(K-1)$ symbols, β from each of TX $2,\ldots,K$, respectively. Similarly, at RX $k\in\{2,\ldots,K\}$, by aligning only β dimensional interference, we obtain the following alignment equations:

$$\begin{aligned}\mathbf{H}^{[k1]}\mathbf{V}_{k-1}^{[1]} &= -(\mathbf{H}^{[k2]}\mathbf{V}_{k-1}^{[2]}+\cdots+\mathbf{H}^{[k(k-1)]}\mathbf{V}_{k-1}^{[k-1]}\\ &\quad+\mathbf{H}^{[k(k+1)]}\mathbf{V}_{k}^{[k+1]}+\cdots+\mathbf{H}^{[kK]}\mathbf{V}_{k}^{[K]})\\ &\Rightarrow \underbrace{[\mathbf{H}^{[k1]}\ \cdots\ \mathbf{H}^{[k(k-1)]}\ \mathbf{H}^{[k(k+1)]}\ \cdots\ \mathbf{H}^{[iK]}]}_{\bar{\mathbf{H}}_k}\\ &\quad\times\underbrace{\begin{bmatrix}\mathbf{V}_{k-1}^{[1]}\\ \vdots\\ \mathbf{V}_{k-1}^{[k-1]}\\ \mathbf{V}_{k}^{[k+1]}\\ \vdots\\ \mathbf{V}_{k}^{[K]}\end{bmatrix}_{\beta K(K-1)\times\beta}}_{\bar{\mathbf{V}}_k}=\mathbf{O}.\end{aligned}$$

Again, $\bar{\mathbf{H}}_i$ is a $\beta(K^2-K-1)\times\beta K(K-1)$ generic matrix and $\bar{\mathbf{V}}_k$ can be determined as the basis vectors of its null space. For the kth alignment equation we show above, we determine the beamforming directions of β symbols per user. Therefore, we have established the beamforming directions of all $\beta(K-1)$ symbols per user. After aligning interference at each RX, we still need to ensure that the desired signals do not overlap with the interference space. In fact, this is guaranteed since the direct channels $\mathbf{H}^{[kk]}$ do not appear in the alignment equations (222) and (223), $k\in\mathcal{K}$. In addition, note that the kth alignment equation only involves the interference carrying links associated with RX k and channel matrices are generic. Thus, the beamforming directions of all $\beta(K-1)$ symbols at each user are linearly independent, almost surely, and can be separated from the interference at each RX. Therefore, each user is able to achieve $d=\beta(K-1)$ DoF, almost surely.

After we establish the DoF achievability at $M/N=\frac{K}{K^2-K-1}$, let us consider $\frac{M}{N}\in[\frac{1}{K-1},\frac{K}{K^2-K-1}]$, i.e., $\frac{K^2-K-1}{K}M\le N$. In this region, the DoF value only depends on M, so we can reduce the number of RX antennas N to $N'=\frac{K^2-K-1}{K}M$, without affecting the DoF, such that it becomes the case $\frac{M}{N'}=\frac{K}{K^2-K-1}$ that we have solved. Note that if the value of $N-N'$ is not an integer, then we can scale the number of both TX and RX antennas by the the same factor α such that $\alpha(N-N')$ is an integer, i.e., we resort to spatial normalization [2]. Similarly, for the case where $M/N\in[\frac{K}{K^2-K-1},\frac{K-1}{K(K-2)}]$, i.e., $M\ge\frac{K}{K^2-K-1}N$, the DoF value $d=\frac{(K-1)N}{K^2-K-1}$ only depends on N and we can reduce the number of transmit antennas from M to $M'=\frac{KN}{K^2-K-1}$, such that M'/N becomes $\frac{K}{K^2-K-1}$ again. Also, if $M-M'$ is not an integer, then we can again use channel extensions over space. ■

Appendix C
The Linear Independence Proofs for the $K=4$ User $M\times N$ MIMO Interference Channel

A. $M/N\in[2/5,1/2)$ Case (Algorithm 2)

Proof: As shown in Algorithm 2, $\mathbf{G}$ may contain two kinds of components, i.e., $\mathcal{O}$ and randomly generated linear combinations of one interferer's symbols. Providing $\mathcal{O}$ releases $|\mathcal{O}|$ dimensional observations of the corresponding TX. Therefore, we need to show that the $|\mathbf{G}|=(3M-N)$ dimensional observations of this TX are linearly independent of the $N-2M$ dimensional observations that the RX has originally. In order to do this, we need to show that the $M\times M$ square matrix whose entries are the linear combination coefficients of the M equations has full rank, i.e., the determinant of this matrix, a polynomial function of its entries, is non-zero almost surely. This polynomial is either a zero polynomial or not equal to zero almost surely for randomly generated channel coefficients. Next, we show that the polynomial is not a zero polynomial. To do that, we only need to find one specific set of channel coefficients such that the polynomial is not equal to zero. Next, we construct the channels for all interference carrying links, i.e., $\mathbf{H}^{[ji]},\ i,j\in\{1,2,3,4\},\ j\ne i$, as shown at the bottom of this page and $a=\gcd(M,N)$. While it is easy to check the specific matrices above have full rank,

$$\mathbf{H}^{[k\ k+1]}=\begin{bmatrix}\mathbf{I}_M\\ \mathbf{O}_M\\ \mathbf{O}_{N-2M}\end{bmatrix},\quad \mathbf{H}^{[k\ k+2]}=\begin{bmatrix}\mathbf{O}_M\\ \mathbf{I}_M\\ \mathbf{O}_{N-2M}\end{bmatrix},$$
$$\mathbf{H}^{[k\ k-1]}=\begin{bmatrix}\mathbf{O}_{(N-2M-a)\times(N-2M)} & \mathbf{O}_{(N-2M-a)\times(N-2M-a)} & \mathbf{O}_{(N-2M-a)\times(5M-2N+a)}\\ \mathbf{O}_{(5M-2N+a)\times(N-2M)} & \mathbf{O}_{(5M-2N+a)\times(N-2M-a)} & \mathbf{I}_{(5M-2N+a)\times(5M-2N+a)}\\ \mathbf{O}_{(N-2M-a)\times(N-2M)} & \mathbf{I}_{(N-2M-a)\times(N-2M-a)} & \mathbf{O}_{(N-2M-a)\times(5M-2N+a)}\\ \mathbf{O}_{a\times(N-2M)} & \mathbf{O}_{a\times(N-2M-a)} & \mathbf{O}_{a\times(5M-2N+a)}\\ \mathbf{O}_{(3M-N)\times(N-2M)} & \mathbf{I}_{(3M-N)\times(N-2M-a)} & \mathbf{O}_{(3M-N)\times(5M-2N+a)}\\ \mathbf{O}_{(3M-N)\times(N-2M)} & \mathbf{O}_{(3M-N)\times(N-2M-a)} & \mathbf{I}_{(3M-N)\times(5M-2N+a)}\\ \mathbf{O}_{(N-2M)\times(N-2M)} & \mathbf{O}_{(N-2M)\times(N-2M-a)} & \mathbf{O}_{(N-2M)\times(5M-2N+a)}\\ \mathbf{I}_{(N-2M)\times(N-2M)} & \mathbf{O}_{(N-2M)\times(N-2M-a)} & \mathbf{O}_{(N-2M)\times(5M-2N+a)}\end{bmatrix}\tag{223}$$

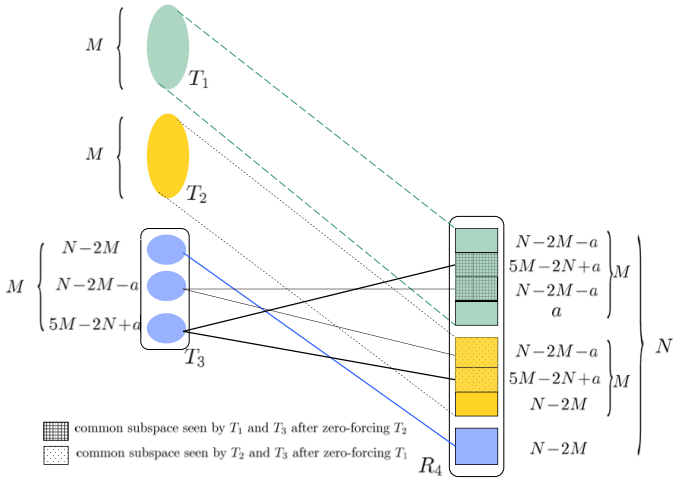

Fig. 9. Linear Dimension Counting of Subspaces Participating in the Interference Alignment for the $M \times N$ setting (the values denote the dimensions of each corresponding subspace.)

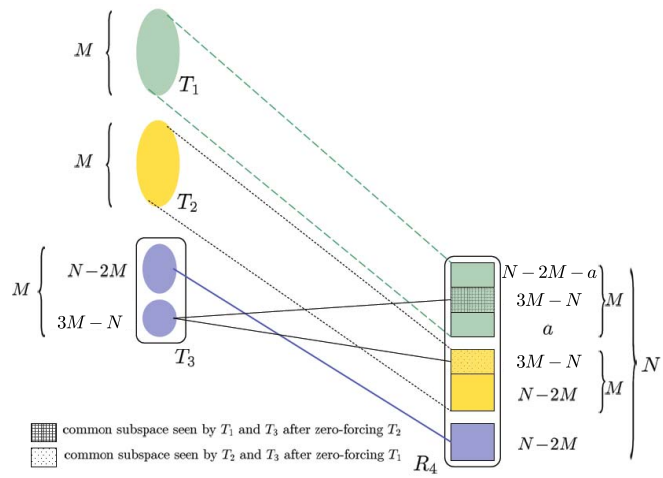

Fig. 10. Linear Dimension Counting of Subspaces Participating in the Interference Alignment for the $M \times N$ setting (the values denote the dimensions of each corresponding subspace.)

we will show through Figure 9 that these matrices keep the generic properties of linear subspaces. For brevity, we only show the channels associated with RX 4. The interference carrying links associated with other receivers can be obtained by advancing user indices. Note that the channel matrices of desired links are generic.

From Figure 9, it can be easily seen that in the N dimensional space at RX 4, after zero forcing the two M-dimensional subspaces from two interferers, RX 4 has $N - 2M$ dimensional clean observations of the remaining interferer. For example, after zero forcing T_1 and T_2, the remaining $N - 2M$ dimensional observations of T_3 are with blue color. Similarly, the remaining $N - 2M$ dimensional observations of T_1 and T_2 after zero forcing the rest of interferers are with green and yellow colors, respectively.

Substituting the specific channel matrices in (223) into Algorithm 2, we can easily check that in each step the $M \times M$ square matrix that we obtain has full rank. In our work, we use programming to check all cases for the values of M, N up to 100. ■

B. Special Case of $M/N \in [3/8, 2/5)$ (Algorithm 2)

Proof: Similar to the proof for $M/N \in [2/5, 1/2)$ case in Appendix C-A, again, we only need to find one specific set of channel coefficients such that the polynomial is not equal to zero. We construct the channels for all interference carrying links, i.e., $\mathbf{H}^{[ji]},\ i, j \in \{1, 2, 3, 4\},\ j \neq i$.

$$\mathbf{H}^{[k\ k+1]} = \begin{bmatrix} \mathbf{I}_M \\ \mathbf{O}_M \\ \mathbf{O}_{N-2M} \end{bmatrix}, \quad \mathbf{H}^{[k\ k+2]} = \begin{bmatrix} \mathbf{O}_M \\ \mathbf{I}_M \\ \mathbf{O}_{N-2M} \end{bmatrix},$$

$$\mathbf{H}^{[k\ k-1]} = \begin{bmatrix} \mathbf{O}_{(N-2M-a)\times(N-2M)} & \mathbf{O}_{(N-2M-a)\times(3M-N)} \\ \mathbf{O}_{(3M-N)\times(N-2M)} & \mathbf{I}_{(3M-N)\times(3M-N)} \\ \mathbf{O}_{a\times(N-2M)} & \mathbf{O}_{a\times(3M-N)} \\ \mathbf{O}_{(3M-N)\times(N-2M)} & \mathbf{I}_{(3M-N)\times(3M-N)} \\ \mathbf{O}_{(N-2M)\times(N-2M)} & \mathbf{O}_{(N-2M)\times(3M-N)} \\ \mathbf{I}_{(N-2M)\times(N-2M)} & \mathbf{O}_{(N-2M)\times(3M-N)} \end{bmatrix} \tag{224}$$

where $a = \gcd(M, N)$. We show through Figure 10 that these matrices keep the generic properties of linear subspaces. For brevity, we only show the channels associated with RX 4. The interference carrying links associated with other receivers can be obtained by advancing user indices. Note that the channel matrices of desired links are generic. Substituting the specific channel matrices in (224) into Algorithm 2, we can easily check that in each step the $M \times M$ square matrix that we obtain has full rank. In our work, we test all (M, N) cases where $M/N = (2c-1)/(5c-2),\ c \in \mathbb{Z}^+ \setminus \{1\},\ M/N \geq 3/8,\ M \leq 100,\ N \leq 100$. ■

Appendix D
DoF of the Four-to-One MIMO Interference Channel

We want to show that

$$d = \begin{cases} M, & M/N \leq 1/4, \\ N/4, & 1/4 \leq M/N \leq 1/3, \\ 3M/4, & 1/3 \leq M/N \leq 4/9, \\ N/3, & 4/9 \leq M/N \leq 1/2, \\ 2M/3, & 1/2 \leq M/N \leq 3/5, \\ 2N/5, & 3/5 \leq M/N \leq 2/3, \\ 3M/5, & 2/3 \leq M/N \leq 5/6, \\ N/2, & 5/6 \leq M/N \leq 1. \end{cases} \tag{225}$$

First, let us consider the case where $M/N \leq 1/3$, i.e., $3M \leq N$. In this case, after decoding and subtracting its own message W_1, RX 1 is able to invert the channel from three interferers, such that RX 1 can reconstruct the three interfering signal vectors subject to noise distortion. Therefore, the DoF value of each user is given by $d = \min(M, N/4)$, i.e., $d = M$ for $M/N \leq 1/4$, and $d = N/4$ for $1/4 \leq M/N \leq 1/3$.

We next consider the remaining cases, where $1/3 \leq M/N$. In the following, we first investigate the outer bound, and then the achievability.

C. The Information Theoretic DoF Outer Bound

For all cases, we apply the genie aided argument to produce the outer bound. For the $M \times N$ Four-to-One MIMO

interference channel, to allow RX 1 to decode all 4 messages, a genie needs to provide RX 1 genie signals with $|\bar{\mathbf{G}}| = 3M - N$ dimensions. We now prove the outer bound for each case.

Case: $1/3 \leq M/N \leq 4/9 \Rightarrow d \leq 3M/4$

In this case, notice that $3M - N \leq N/3 \leq M$. A genie provides $\bar{\mathbf{G}} = \bar{\mathbf{X}}^{[2]}_{(3M-N)}$ to RX 1. Since $\mathbf{G}$ is linearly independent of $\mathbf{S}^{[1]}$, RX 1 is able to decode all the messages subject to noise distortion. Thus, we obtain the following inequality:

$$\begin{aligned} &4nR - n\epsilon_n \\ &\quad \leq: \hbar(\mathbf{Y}^{[1]^n}) + \hbar(\mathbf{G}^n|\mathbf{S}^{[1]^n}) && (226)\\ &\quad \leq: Nn\log\rho + \hbar(\mathbf{X}^{[2]^n}_{3M-N}) && (227)\\ &\quad \leq: Nn\log\rho + (3M-N)n\log\rho = 3Mn\log\rho. && (228)\end{aligned}$$

By letting $n \to \infty$ first and then $\rho \to \infty$, we have the desired DoF outer bound

$$d \leq 3M/4. \tag{229}$$

Case: $4/9 \leq M/N \leq 1/2 \Rightarrow d \leq N/3$

A genie provides $\bar{\mathbf{G}} = \bar{\mathbf{X}}^{[2]}$ to RX 1, such that after removing its desired signal and the interference from TX 2, RX 1 is able to invert the channels from TX 3 and TX 4 as $N \geq 2M$. Thus, we obtain the following inequality:

$$\begin{aligned} 4nR - n\epsilon_n &\leq: \hbar(\mathbf{Y}^{[1]^n}) + \hbar(\mathbf{G}^n|\mathbf{S}^{[1]^n})\\ &\leq: Nn\log\rho + \hbar(\mathbf{X}^{[2]^n}) \leq: Nn\log\rho + nR.\end{aligned}$$

By letting $n \to \infty$ first and then $\rho \to \infty$, we have the desired DoF outer bound

$$d \leq N/3. \tag{230}$$

Case: $1/2 \leq M/N \leq 3/5 \Rightarrow d \leq 2M/3$

In this case, $3M - N \geq M$. A genie provides $\bar{\mathbf{G}} = \{\bar{\mathbf{X}}^{[2]}, \bar{\mathbf{X}}^{[3]}_{(2M-N)}\}$ to RX 1. After removing its desired signal and the interference from TX 2, RX 1 is able to recover the signal vectors from TX 3 and TX 4 from the N-dimensional observations of $\mathbf{X}^{[3]}, \mathbf{X}^{[4]}$ and $\mathbf{G}$ subject to noise distortion. Thus, we obtain the following inequality:

$$\begin{aligned} 4nR - n\epsilon_n &\leq: \hbar(\mathbf{Y}^{[1]^n}) + \hbar(\mathbf{G}^n|\mathbf{S}^{[1]^n}) && (231)\\ &\leq: Nn\log\rho + \hbar(\mathbf{X}^{[2]^n}, \mathbf{X}^{[3]^n}_{(2M-N)}) && (232)\\ &\leq: Nn\log\rho + nR + (2M-N)n\log\rho. && (233)\end{aligned}$$

By letting $n \to \infty$ first and then $\rho \to \infty$, we have the desired DoF outer bound

$$d \leq 2M/3. \tag{234}$$

Case: $3/5 \leq M/N \leq 2/3 \Rightarrow d \leq 2N/5$

In this case, $3M - N \geq M$. First, a genie provides $\bar{\mathbf{G}}_1 = \{\bar{\mathbf{X}}^{[2]}, \bar{\mathbf{X}}^{[3]}_{(2M-N)}\}$ to RX 1, such that it can decode all the messages subject to noise distortion. Therefore, we have the first inequality:

$$\begin{aligned} &4nR - n\epsilon_n \\ &\quad \leq: \hbar(\mathbf{Y}^{[1]^n}) + \hbar(\mathbf{G}^n_1|\mathbf{S}^{[1]^n}) && (235)\\ &\quad \leq: Nn\log\rho + \hbar(\mathbf{X}^{[2]^n}, \mathbf{X}^{[3]^n}_{(2M-N)}|\mathbf{S}^{[1]^n}) && (236)\\ &\quad \leq: Nn\log\rho + \hbar(\mathbf{X}^{[2]^n}) + \hbar(\mathbf{X}^{[3]^n}_{(2M-N)}|\mathbf{S}^{[1]^n}, \mathbf{X}^{[2]^n})\\ &\quad \leq: Nn\log\rho + nR + \hbar(\mathbf{X}^{[3]^n}_{(2M-N)}|\mathbf{X}^{[3\sim1]^n}_{N-M}) && (237)\\ &\quad \leq: Nn\log\rho + 2nR - \hbar(\mathbf{X}^{[3\sim1]^n}_{N-M}) && (238)\end{aligned}$$

where (237) is obtained by zero forcing the interference from TX 4.

Second, a genie provides $\bar{\mathbf{G}}_2 = \{\bar{\mathbf{X}}^{[4]}, \bar{\mathbf{X}}^{[3\sim1]^n}_{N-M}\}$ to RX 1. Notice that $|\bar{\mathbf{G}}_2| = M + (N-M) \geq 3M - N$ and $\mathbf{X}^{[3\sim1]^n}_{N-M}$ depends only on the channel from TX 4. Thus, providing $\bar{\mathbf{G}}_2$ to RX 1 allows it to decode all the messages subject to noise distortion and we have the second inequality:

$$\begin{aligned} 4nR - n\epsilon_n &\leq: \hbar(\mathbf{Y}^{[1]^n}) + \hbar(\mathbf{G}^n_1|\mathbf{S}^{[1]^n}) && (239)\\ &\leq: Nn\log\rho + \hbar(\mathbf{X}^{[4]^n}, \mathbf{X}^{[3\sim1]^n}_{N-M}) && (240)\\ &\leq: Nn\log\rho + nR + \hbar(\mathbf{X}^{[3\sim1]^n}_{N-M}). && (241)\end{aligned}$$

Adding (238) and (241), we have the desired DoF outer bound:

$$8nR - n\epsilon_n \leq: 2Nn\log\rho + 3nR \Rightarrow d \leq 2N/5. \tag{242}$$

Case: $2/3 \leq M/N \leq 5/6 \Rightarrow d \leq 3M/5$

The proof for this case is similar to that for the $3/5 \leq M/N \leq 2/3$ case. The first inequality is the same as (238). To obtain the second inequality, a genie provides $\bar{\mathbf{G}}_2 = \{\bar{\mathbf{X}}^{[4]}, \bar{\mathbf{X}}^{[3\sim1]}_{N-M}, \bar{\mathbf{X}}^{[3]}_{3M-2N}\}$ to RX 1 to allow it to decode all the messages subject to noise distortion. As a result, we have the second inequality:

$$\begin{aligned} &4nR - n\epsilon_n \\ &\quad \leq: \hbar(\mathbf{Y}^{[1]^n}) + \hbar(\mathbf{G}^n_2|\mathbf{S}^{[1]^n}) && (243)\\ &\quad \leq: Nn\log\rho + \hbar(\mathbf{X}^{[4]^n}, \mathbf{X}^{[3\sim1]^n}_{N-M}, \mathbf{X}^{[3]^n}_{3M-2N}) && (244)\\ &\quad \leq: Nn\log\rho + \hbar(\mathbf{X}^{[4]^n}) + \hbar(\mathbf{X}^{[3\sim1]^n}_{N-M}) + \hbar(\mathbf{X}^{[3]^n}_{3M-2N})\\ &\quad \leq: Nn\log\rho + nR + \hbar(\mathbf{X}^{[3\sim1]^n}_{N-M}) + (3M-2N)n\log\rho. && (245)\end{aligned}$$

Adding up (238) and (245) we have the desired DoF outer bound:

$$\begin{aligned} 8nR - n\epsilon_n &\leq: 2Nn\log\rho + 3nR + (3M-2N)n\log\rho\\ &\Rightarrow d \leq 3M/5.\end{aligned}$$

Case: $5/6 \leq M/N \leq 1 \Rightarrow d \leq N/2$

A genie provides $\bar{\mathbf{G}} = \{\bar{\mathbf{X}}^{[2]}, \bar{\mathbf{X}}^{[3]}\}$ to RX 1. After removing its desired signal and the interference from TX 2 and TX 3, RX 1 is able to invert the channels from TX 4 to recover $\mathbf{X}^{[4]}$ and thus decode W_4 subject to noise distortion. Thus, we obtain the following inequality:

$$\begin{aligned} 4nR - n\epsilon_n &\leq: \hbar(\mathbf{Y}^{[1]^n}) + \hbar(\mathbf{G}^n|\mathbf{S}^{[1]^n}) && (246)\\ &\leq: Nn\log\rho + \hbar(\mathbf{X}^{[2]^n}, \mathbf{X}^{[3]^n}) && (247)\\ &\leq: Nn\log\rho + 2nR. && (248)\end{aligned}$$

By letting $n \to \infty$ first and then $\rho \to \infty$, we have the desired DoF outer bound

$$d \leq N/2. \tag{249}$$

■

D. The DoF Achievability

It suffices to show the achievability at $M/N = 4/9$, $3/5$, $5/6$, and all the other cases directly follow from the property that increasing the number of antennas does not hurt the DoF. As only RX 1 suffers from interference, we merely need to design the precoding matrices $\mathbf{V}^{[i]}$, $i \in \mathcal{K}$ at each TX to minimize the dimension of the interference seen at RX 1.

Case: $M/N = 4/9 \Rightarrow d = 3M/4 = N/3$.

Suppose $(M, N) = (4\beta, 9\beta)$ where $\beta \in \mathbb{Z}^+$. In this case, we want to show that $d = 3\beta$ DoF per user are achievable. Consider the 9β-dimensional received signal subspace at RX 1. The desired signal occupies 3β dimensions, leaving the remaining 6β dimensions for interference. Since the interference from TX 2 to TX 4 has $3d = 9\beta$ variables in total, we need to align $9\beta - 6\beta = 3\beta$ interference symbols. Note that TX 2 and TX 3 together project at RX 1 an 8β-dimensional subspace, which intersects with the 4β-dimensional subspace seen from TX 4 at RX 1 at an $8\beta+4\beta-9\beta = 3\beta$ dimensional subspace. Therefore, we can align the 3β symbols from TX 4 into the subspace spanned by the interference from TX 2 and TX 3. Mathematically, we have

$$\begin{aligned}\mathbf{H}^{[14]}\mathbf{V}^{[4]} &= -(\mathbf{H}^{[12]}\mathbf{V}^{[2]} + \mathbf{H}^{[13]}\mathbf{V}^{[3]}) \\ &\Rightarrow \begin{bmatrix}\mathbf{H}^{[12]} & \mathbf{H}^{[13]} & \mathbf{H}^{[14]}\end{bmatrix}_{9\beta\times 12\beta}\begin{bmatrix}\mathbf{V}^{[2]}\\ \mathbf{V}^{[3]}\\ \mathbf{V}^{[4]}\end{bmatrix}_{12\beta\times 3\beta} \quad (250)\\ &= \mathbf{0}, \quad (251)\end{aligned}$$

and we can solve $\mathbf{V}^{[2]}, \mathbf{V}^{[3]}, \mathbf{V}^{[4]}$ by truncating the 3β basis vectors of the null space of $\begin{bmatrix}\mathbf{H}^{[12]} & \mathbf{H}^{[13]} & \mathbf{H}^{[14]}\end{bmatrix}$. Finally, we randomly generate $\mathbf{V}^{[1]}$ to ensure the linear independence between the desired signals and interference at RX 1, such that RX 1 is able to decode its 3β symbols. Since RX 2 to RX 4 hear no interference, they are able to decode their own messages as well.

Case: $M/N = 3/5 \Rightarrow d = 2M/3 = 2N/5$

Suppose $(M, N) = (3\beta, 5\beta)$ where $\beta \in \mathbb{Z}^+$. In this case, we want to show that $d = 2\beta$ DoF per user are achievable. In the 5β-dimensional received signal space at RX 1, the desired signal occupies 2β dimensions, thus leaving the rest 3β-dimensional subspace for interference. As a result, a total of $3d = 6\beta$ dimensional interference from TX 2 to 4 should be aligned into a 3β-dimensional subspace. Note that each TX projects a 3β-dimensional subspace at RX 1, and any two of them have a $3\beta + 3\beta - 5\beta = \beta$ dimensional intersection. Therefore, for any two users among the three interferers, we align β symbols and thus save 3β interference dimensions. Suppose the beamforming matrix of TX i is given by $\mathbf{V}^{[i]} = [\mathbf{V}_1^{[i]}, \mathbf{V}_2^{[i]}]$. We have

$$\begin{aligned}&\mathbf{H}^{[12]}\mathbf{V}_1^{[2]} = -\mathbf{H}^{[13]}\mathbf{V}_1^{[3]}\\ &\Rightarrow \begin{bmatrix}\mathbf{H}^{[12]}\mathbf{H}^{[13]}\end{bmatrix}_{5\beta\times 6\beta}\begin{bmatrix}\mathbf{V}_1^{[2]}\\ \mathbf{V}_1^{[3]}\end{bmatrix}_{6\beta\times\beta} = \mathbf{0}, \quad (252)\\ &\mathbf{H}^{[12]}\mathbf{V}_2^{[2]} = -\mathbf{H}^{[14]}\mathbf{V}_1^{[4]}\\ &\Rightarrow \begin{bmatrix}\mathbf{H}^{[12]}\mathbf{H}^{[14]}\end{bmatrix}_{5\beta\times 6\beta}\begin{bmatrix}\mathbf{V}_2^{[2]}\\ \mathbf{V}_1^{[4]}\end{bmatrix}_{6\beta\times\beta} = \mathbf{0}, \quad (253)\\ &\mathbf{H}^{[13]}\mathbf{V}_2^{[3]} = -\mathbf{H}^{[14]}\mathbf{V}_2^{[4]}\\ &\Rightarrow \begin{bmatrix}\mathbf{H}^{[13]}\mathbf{H}^{[14]}\end{bmatrix}_{5\beta\times 6\beta}\begin{bmatrix}\mathbf{V}_2^{[3]}\\ \mathbf{V}_2^{[4]}\end{bmatrix}_{6\beta\times\beta} = \mathbf{0}, \quad (254)\end{aligned}$$

and therefore we can solve the equations by truncating the basis vectors of the null spaces of the corresponding matrices. The linear independence among 2β columns of each $\mathbf{V}^{[i]}$, $i = 2, 3, 4$ can be established by choosing a set of special matrices and showing that it has full rank. Finally, we choose $\mathbf{V}^{[1]}$ randomly to ensure the linear independence of the desired signal and interference at RX 1.

Case: $M/N = 5/6 \Rightarrow d = 3M/5 = N/2$

Suppose $(M, N) = (5\beta, 6\beta)$ where $\beta \in \mathbb{Z}^+$. In this case, we want to show that $d = 3\beta$ DoF per user are achievable. In the 6β-dimensional received signal space at RX 1, the desired signal occupies 3β dimensions, leaving the rest 3β-dimensional subspace for interference. Because we have a total of $3d = 9\beta$ symbols from three interferers, we need to align the interference from TX 3 and TX 4 into the same subspace projected from TX 2. Therefore, we have

$$\begin{aligned}\mathbf{H}^{[12]}\mathbf{V}^{[2]} &= -\mathbf{H}^{[13]}\mathbf{V}^{[3]} = -\mathbf{H}^{[14]}\mathbf{V}^{[4]} \quad (255)\\ &\Rightarrow \begin{bmatrix}\mathbf{H}^{[12]} & \mathbf{H}^{[13]} & \mathbf{O}\\ \mathbf{H}^{[12]} & \mathbf{O} & \mathbf{H}^{[14]}\end{bmatrix}_{12\beta\times 15\beta}\begin{bmatrix}\mathbf{V}^{[2]}\\ \mathbf{V}^{[3]}\\ \mathbf{V}^{[4]}\end{bmatrix}_{15\beta\times 3\beta} \quad (256)\\ &= \mathbf{0}, \quad (257)\end{aligned}$$

and $\mathbf{V}^{[i]}, i = 2, 3, 4$ can be solved similarly. Finally, we choose $\mathbf{V}^{[1]}$ randomly to ensure the linear independence of the desired signal and interference at RX 1. ■

APPENDIX E
DoF COUNTING BOUND OF THE MANY-TO-ONE MIMO INTERFERENCE CHANNEL

In the K-user many-to-one $M \times N$ MIMO Gaussian interference channel where each TX has M antennas and each RX has N antennas, only RX 1 hears interference from TX 2 to TX K while all the other receivers only hear their desired signals. We consider the feasibility of interference alignment over such a channel, i.e., the DoF achieved by linear schemes with no symbol extension. The DoF counting bound is based on counting the number of free variables and alignment bilinear equations. This idea of counting variables and equations is originally proposed by Yetis *et al.* [4] in the fully connected MIMO interference channel, and then applied in the X channel setting [13].

Using linear schemes, TX i intends to send d independent streams using an $M \times d$ precoding matrix $\mathbf{V}^{[i]}$ and RX i extracts its desired signal by using an $N \times d$ interference filtering matrix $\mathbf{U}^{[i]}$, $i \in \mathcal{K}$. Then the feasibility of linear interference alignment is equivalent to solving the following algebraic equations:

$$\mathbf{U}^{[1]^\dagger}\mathbf{H}^{[1i]}\mathbf{V}^{[i]} = 0, \qquad i \in \{2, 3, \ldots, K\} \quad (258)$$

which essentially means that RX 1 may zero force all the interference. Now we count the number of equations N_e and

variables N_v in (258):

$$N_e = (K-1)d^2, \tag{259}$$

$$N_v = (K-1)(M-d)d + (N-d)d. \tag{260}$$

The DoF counting bound is given by the inequality

$$N_e \leq N_v \tag{261}$$

$$\Rightarrow d \leq \frac{(K-1)M+N}{2K-1}. \tag{262}$$

When $K = 4$, the DoF counting bound is $\frac{3M+N}{7}$ and it is easily checked that the counting bound $\frac{3M+N}{7}$ is always larger than the decomposition bound $\frac{MN}{M+N}$.

REFERENCES

[1] C. Wang, H. Sun, and S. A. Jafar, "Genie chains and the degrees of freedom of the K-user MIMO interference channel," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Cambridge, MA, USA, Jul. 2012, pp. 2476–2480.

[2] C. Wang, T. Gou, and S. A. Jafar, "Subspace alignment chains and the degrees of freedom of the three-user MIMO interference channel," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2432–2479, May 2014.

[3] G. Bresler, D. Cartwright, and D. Tse, "Feasibility of interference alignment for the MIMO interference channel," *IEEE Trans. Inf. Theory*, vol. 60, no. 9, pp. 5573–5586, Sep. 2014.

[4] C. M. Yetis, T. Gou, S. A. Jafar, and A. H. Kayran, "On feasibility of interference alignment in MIMO interference networks," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4771–4782, Sep. 2010.

[5] M. Razaviyayn, G. Lyubeznik, and Z.-Q. Luo, "On the degrees of freedom achievable through interference alignment in a MIMO interference channel," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 812–821, Feb. 2012.

[6] S. Jafar, "Understanding spatial signal dimensions: Recent results and some conjectures on the DoF of MIMO interference networks," in *Proc. Inf. Theory Appl. Workshop*, Jan. 2012.

[7] A. Ghasemi, A. S. Motahari, and A. K. Khandani. (Sep. 2009). "Interference alignment for the K user MIMO interference channel." [Online]. Available: https://arxiv.org/pdf/0909.4604

[8] V. R. Cadambe and S. A. Jafar, "Interference alignment and the degrees of freedom of the K user interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3425–3441, Aug. 2008.

[9] G. Sridharan and W. Yu, "Degrees of freedom of MIMO cellular networks with two cells and two users per cell," in *Proc. ISIT*, Jul. 2013, pp. 1774–1778.

[10] G. Sridharan and W. Yu, "Degrees of freedom of MIMO cellular networks: Two-cell three-user-per-cell case," in *Proc. GLOBECOM*, Dec. 2013, pp. 1968–1973.

[11] S. A. Jafar and M. J. Fakhereddin, "Degrees of freedom for the MIMO interference channel," *IEEE Trans. Inf. Theory*, vol. 53, no. 7, pp. 2637–2642, Jul. 2007.

[12] Ó. González, C. Beltrán, and I. Santamaría, "A feasibility test for linear interference alignment in MIMO channels with constant coefficients," *IEEE Trans. Inf. Theory*, vol. 60, no. 3, pp. 1840–1856, Mar. 2014.

[13] H. Sun, C. Geng, T. Gou, and S. A. Jafar, "Degrees of freedom of MIMO X networks: Spatial scale invariance, one-sided decomposability and linear feasibility," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2012, pp. 2082–2086.

Chenwei Wang (S'05–M'13) received his Ph.D. in Electrical and Computer Engineering from the University of California, Irvine, USA, in December 2012. Prior to that, he received his B.Eng. in Information Engineering and M.S. in Communications and Information Systems from Beijing University of Posts and Telecommunications, China, in 2005 and 2008, respectively. From February 2013, Dr. Wang has been a Research Engineer in the Mobile Network Technology Group of DOCOMO Innovations Inc., Palo Alto, CA, USA. His current research interests include the 5G key technologies, network information theory, wireless communications and networks. Before joining DOCOMO Innovations Inc., his industry experience includes intern positions in the Research, Technology and Platforms Group at Nokia Siemens Networks, Beijing, China in 2008, and the Network Innovation Lab of DoCoMo USA Labs, Palo Alto, CA, USA in 2010. He was a recipient of 2014 IEEE Signal Processing Society Young Author Best Paper Award, and also a recipient of the honor of IEEE COMMUNICATIONS LETTERS Exemplary Reviewer in 2011 and 2014.

Hua Sun (S'12) received the B.E. degree in Communications Engineering from Beijing University of Posts and Telecommunications, Beijing, China, in 2011, the M.S. degree in Electrical and Computer Engineering from University of California Irvine, USA, in 2013. He is currently pursuing the Ph.D. degree at the University of California Irvine. His research interests include information theory and its applications to communications, networking, and security. He received the IEEE Jack Keil Wolf ISIT Student Paper Award in 2016.

Syed Ali Jafar (S'99–M'04–SM'09–F'14) received his B. Tech. from IIT Delhi, India, in 1997, M.S. from Caltech, USA, in 1999, and Ph.D. from Stanford, USA, in 2003, all in Electrical Engineering. His industry experience includes positions at Lucent Bell Labs, Qualcomm Inc. and Hughes Software Systems. He is a Professor in the Department of Electrical Engineering and Computer Science at the University of California Irvine, Irvine, CA USA. His research interests include multiuser information theory, wireless communications and network coding.

Dr. Jafar received the Blavatnik National Laureate in Physical Sciences and Engineering in 2015, the UCI Academic Senate Distinguished Mid-Career Faculty Award for Research in 2015, the School of Engineering Mid-Career Excellence in Research Award in 2015, the School of Engineering Maseeh Outstanding Research Award in 2010, the ONR Young Investigator Award in 2008, and the NSF CAREER award in 2006. His co-authored papers received the IEEE Information Theory Society Best Paper Award in 2009, IEEE Communications Society Best Tutorial Paper Award in 2013, IEEE Communications Society Heinrich Hertz Award in 2015, IEEE Signal Processing Society Young Author Best Paper Award (to student co-authors) in 2015, an IEEE GLOBECOM Best Paper Award in 2012 and an IEEE GLOBECOM Best Paper Award in 2014. Dr. Jafar received the UC Irvine EECS Professor of the Year award five times, in 2006, 2009, 2011, 2012, and 2014, from the Engineering Students Council and the Teaching Excellence Award in 2012 from the School of Engineering. He was a University of Canterbury Erskine Fellow in 2010 and an IEEE Communications Society Distinguished Lecturer for 2013-2014. Dr. Jafar was recognized as a Thomson Reuters Highly Cited Researcher and included by Sciencewatch among The World's Most Influential Scientific Minds in 2014 and again in 2015. He served as Associate Editor for IEEE TRANSACTIONS ON COMMUNICATIONS 2004- 2009, for IEEE COMMUNICATIONS LETTERS 2008-2009 and for IEEE TRANSACTIONS ON INFORMATION THEORY 2009-2012.