A Tractable and Accurate Cross-Layer Model for Multi-Hop MIMO Ad Hoc Networks

Jia Liu Yi Shi Cunhao Gao Y. Thomas Hou
Bradley Department of Electrical and Computer Engineering
Virginia Polytechnic Institute and State University, Blacksburg, VA 24061

***Abstract*—MIMO-based communications have great potential to improve network capacity for wireless ad hoc networks. Although there has been significant progress on MIMO at the physical layer or single-hop communication over the years, advances in the theory of MIMO for multi-hop ad hoc networks remain limited. This stagnation is mainly due to the lack of an accurate and more important, analytically tractable model that can be used by networking researchers. In this paper, we propose such a model to enable the networking community to carry out cross-layer research for multi-hop MIMO ad hoc networks. In particular, at the physical layer, we develop a simple model for MIMO link capacity computation that captures the essence of spatial multiplexing and transmit power limit without involving complex matrix operations and the water-filling algorithm. We show that the approximation gap in this model is negligible. At the link layer, we devise a space-time scheduling scheme called SUCCINCT that significantly advances the existing zero-forcing beamforming (ZFBF) to handle interference in a multi-hop network setting. The proposed SUCCINCT scheme employs simple numeric computation on matrix dimensions to simplify ZFBF in a multi-hop ad hoc network. As a result, we can characterize link layer scheduling behavior without entangling with beamforming details. Finally, we apply both our physical and link layer models in cross-layer performance optimization for a multi-hop MIMO ad hoc network.**

I. Introduction

Since its inception [5], [21], MIMO has been widely accepted as a key technology to increase wireless capacity. Researchers have shown that by employing multiple antennas on the transmitting and receiving nodes, wireless channel capacity can scale almost linearly with the number of antennas. Such capability is the driving force for the wide deployment of MIMO in wireless LAN (802.11n), WiMAX access networks (802.16), 4G cellular networks (LTE), etc.

Although there has been extensive studies on MIMO at the physical layer for point-to-point and cellular communications over the past decade (see, e.g., [2] for an overview), fundamental understanding and results on MIMO in multi-hop ad hoc networks remain limited, particularly from a cross-layer perspective. This stagnation is mainly due to the lack of an accurate and more importantly, tractable model that is amenable for analysis by networking researchers. Traditional signal processing and channel models for MIMO in communications research are clogged with complex matrix representations and operations, rendering enormous challenges for multi-hop ad hoc network optimizations. Due to these challenges, most efforts on multi-hop MIMO ad hoc networks to date [1], [4], [8], [10], [12], [15]–[17], [20] fall into the following two approaches.

The first approach is to formulate the problems by faithfully incorporating the MIMO channel and signal models without any loss of accuracy. However, the problem formulation under this approach is likely to become *intractable* due to the heavy burden from the underlying models. For example, Kim *et al.* studied a maxmin optimization problem in [12] for multi-hop MIMO backhaul networks where they formulated a nonlinear optimization problem to maximize the fair throughput of the access points in the network under the routing, MAC, and physical layer constraints. The physical layer in [12] is based on minimum mean square error (MMSE) beamforming. In [4], Chu and Wang also studied cross-layer algorithms for MIMO ad hoc networks where MMSE sequential interference cancelation technique (MMSE-SIC) is employed at the physical layer to maximize signal to interference and noise ratio (SINR). Due to the complex MMSE mechanics, the cross-layer optimization problems in [12] and [4] are intractable and the authors had to resort to heuristic algorithms.

The second approach is to simplify MIMO physical layer behavior so that tractable analysis can be developed for networking research. Although such approach is attractive, the problem with existing models under this approach suffer from "over simplification." That is, existing simple models ignore many important characteristics of MIMO for cross-layer design opportunities and thus lead to results far from MIMO's achievable performance. In [1], [8], a simplified MIMO cross-layer model was employed to study different throughput optimization problems. By using this model, the network throughput performance can be characterized simply by counting the number of degrees of freedom (DoF) in the network. However, this model does not consider transmit power constraint and power allocation at each node in the network. Also, although some ideas of zero-forcing beamforming (ZFBF) were employed to handle interference, the proposed interference cancelation scheme at the link layer was not designed appropriately, resulting in an unnecessary small DoF region and inferior throughput performance. Also, in [10], [15]–[17], [20], various studies on MAC designs and routing schemes are given based on very simple MIMO models that do not fully exploit MIMO physical capabilities.

The goal of this paper is to achieve the best of both approaches while avoiding their pitfalls. We want to construct a model for MIMO that is both *tractable* and *accurate* for cross-

layer optimization. Our main contributions are as follows.

- At the physical layer, we devise a simple model for computing MIMO channel capacity. This model captures the essence of both spatial multiplexing and transmit power constraint. More importantly, this model does not require complex matrices computation and the complicated water-filling process (which does not admit a close-form solution). We show that the gap between our proposed model and the exact capacity model is negligible.
- At the link layer, we construct a model that takes into account the interference nulling/supression by exploiting ZFBF. More specifically, we propose a space-time scheduling scheme called SUCCINCT (abbreviation of successive interference cancelation). The proposed SUCCINCT employs simple numeric computation on matrix dimensions to simplify ZFBF in a multi-hop ad hoc network. Moreover, by carefully arranging the cancelation order among the nodes, SUCCINCT does not waste unnecessary DoF resources on interference mitigation, thus offering superior throughput performance than those in [1], [8].
- As an application, we use the proposed new models to study cross-layer utility maximization problems for MIMO ad hoc networks. We show that the resulting optimization problems no longer involve complex matrix variables and operations. Further, the formulated problems share a lot of similarities with those cross-layer optimization problems under single-antenna ad hoc networks, which have been actively studied in recent years. This suggests that new solutions to MIMO ad hoc networks may be developed by drawing upon the experiences gained for single-antenna ad hoc networks.

The remainder of this paper is organized as follows. Section II presents a new channel capacity model for MIMO at the physical layer. Section III presents a new link layer model called SUCCINCT. In Section IV, as an application of our new models, we study a cross-layer optimization problem in a multi-hop MIMO ad hoc network. Section V concludes this paper.

II. A Model for Physical Layer Capacity Computation

From networking research perspective, the most important aspect of physical layer modeling for MIMO is its channel capacity computation. In Section II-A, we first give background on MIMO channel capacity computation and analyze why it is difficult to work with for networking research. Then, in Section II-B, we propose a new model for MIMO channel capacity that is both simple and accurate.

Before proceeding to the details of our proposed MIMO ad hoc network modeling, we first summarize all the key notation that appear in this and the subsequent sections in Table I.

TABLE I
Notation.

$t(l)$	The transmitting node of link l.
$r(l)$	The receiving node of link l.
$\mathbf{H}$	A MIMO matrix channel.
$\mathbf{H}_i$	A MIMO matrix channel for link i.
$\mathbf{H}_{i,j}$	A MIMO matrix channel from $t(i)$ to $r(j)$.
$\mathbf{T}_n$	Transmitting beamforming matrix for node n.
$\mathbf{R}_n$	Receiving beamforming matrix for node n.
$g_n(t)$	Node n's transmission status in time slot t.
$h_n(t)$	Node n's reception status in time slot t.
A_n	The number of antennas at node n.
$\mathcal{N}$	The set of nodes in the network.
$\mathcal{L}$	The set of links in the network.
N	The number of nodes in the network.
L	The number of links in the network.
$\mathcal{L}_n^{\text{Out}}$	The set of links originating from node n.
$\mathcal{L}_n^{\text{In}}$	The set of links coming into node n.
$\mathcal{I}_n$	The set of nodes within the interference range of node n.
$\pi_{mn}(t)$	The ordering relationship between nodes m and n in time t.
$z_l(t)$	The number of data streams on link l during time slot t.
$\mathcal{C}$	The DoF region of a ZFBF-based scheduling scheme.
$r_l^{(f)}$	The amount of flow of session f on link l.
r_f	The end-to-end session rate of session f.
$\text{src}(f)$	The source node of session f.
$\text{dst}(f)$	The destination node of session f.
$C_l(t)$	The capacity of link l during time slot t.
$\alpha_l(t)$	The power allocated on link l during time slot t.

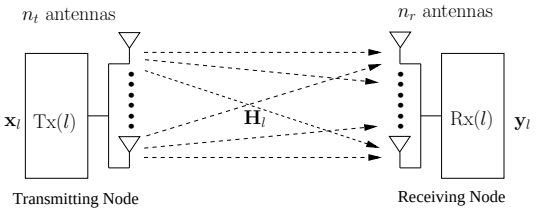

Fig. 1. A MIMO channel illustration.

A. Why Existing Physical Model for MIMO is Difficult to Use?

The channel of a MIMO link l is characterized by a matrix $\mathbf{H}_l$, as shown in Fig. 1. Communication over such a MIMO channel with n_t transmit antennas and n_r receive antennas can be described by

$$\mathbf{y}_l = \sqrt{\rho_l \alpha_l}\mathbf{H}_l\mathbf{x}_l + \mathbf{n}_l, \tag{1}$$

where $\mathbf{x}_l$, $\mathbf{y}_l$ and $\mathbf{n}_l$ denote the vectors of transmitted signal, received signal, and white Gaussian noise with unit variance, respectively.

In (1), ρ_l represents the received SNR of the channel, $\alpha_l \in [0, 1]$ represents the fraction of the transmit power that is assigned to link l (in the case when the source of link l also transmits on other links). As we shall see later in Section IV, α_l is useful to model the power allocation at each node if multi-path routing is employed in the network. For the single link case in Fig. 1, we have $\alpha_l = 1$.

The channel gain matrix $\mathbf{H}_l$ is typically assumed to be a complex random matrix with each of its entries being i.i.d. Gaussian distributed [23] with zero mean and unit variance.

To compute MIMO channel capacity, one needs to diagonalize $\mathbf{H}_l$ so that the channel is transformed into a set of parallel spatial channels. More specifically, by singular value decomposition (SVD), the channel model in (1) can be transformed as $\mathbf{y}_l = \sqrt{\rho_l \alpha_l}\mathbf{U}_l\mathbf{\Lambda}_l\mathbf{V}_l^\dagger\mathbf{x}_l + \mathbf{n}_l$, where $\mathbf{U}_l$ and $\mathbf{V}_l$

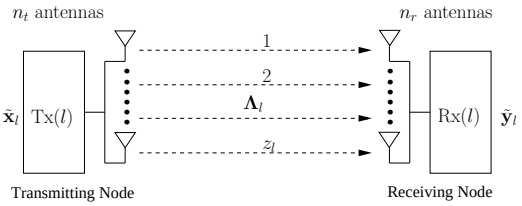

Fig. 2. The equivalent parallel scalar channels after transformation.

are unitary matrices, $\mathbf{\Lambda}_l$ is a diagonal matrix with the singular values of $\mathbf{H}_l$ on its main diagonal. By letting $\tilde{\mathbf{x}}_l = \mathbf{V}_l^\dagger \mathbf{x}_l$, $\tilde{\mathbf{y}}_l = \mathbf{U}_l^\dagger \mathbf{y}_l$, and $\tilde{\mathbf{n}}_l = \mathbf{U}_l^\dagger \mathbf{n}_l$, the channel model can be re-written as

$$\tilde{\mathbf{y}}_l = \sqrt{\rho_l \alpha_l}\mathbf{\Lambda}_l \tilde{\mathbf{x}}_l + \tilde{\mathbf{n}}_l, \tag{2}$$

which is equivalent to the channel in Fig. 2. The number of non-zero singular values (i.e., non-zero diagonal entries in $\mathbf{\Lambda}_l$) is $z_l \leq \min\{n_t, n_r\}$, i.e., the rank of $\mathbf{H}_l$. The rank of $\mathbf{H}_l$ is also called the *degrees of freedom* (DoF), which measures the number of independent signaling dimensions that are available in the channel.

By simple matrix manipulations, (2) can be re-written in the following form:

$$\tilde{\mathbf{y}}_l = \sqrt{\rho_l \alpha_l}\mathbf{R}_{\mathrm{Rx}(l)}^\dagger \mathbf{H}_l \mathbf{T}_{\mathrm{Tx}(l)} \tilde{\mathbf{x}}_l + \tilde{\mathbf{n}}_l, \tag{3}$$

where $\mathrm{Tx}(l)$ and $\mathrm{Rx}(l)$ denote the transmitting and receiving nodes of link l, respectively, and matrices $\mathbf{T}_{\mathrm{Tx}(l)} = \mathbf{V}_l$ and $\mathbf{R}_{\mathrm{Rx}(l)} = \mathbf{U}_l$. In MIMO communications literature (see, e.g., [22]), the matrices $\mathbf{T}_{\mathrm{Tx}(l)}$ and $\mathbf{R}_{\mathrm{Rx}(l)}$ are usually referred to as *transmit beamforming matrix* and *receive beamforming matrix*. As later shown in Section III, transmit and receive beamforming matrices will play an important role in MIMO networks interference mitigation at the link layer.

The capacity for the set of parallel channels in (2) can be found by the water-filling power allocation algorithm [21]:

$$\begin{aligned} C_l^{(\mathrm{wf})} &= \max_{\mathbf{Q}_l} W \log_2 \det(\mathbf{I} + \rho_l \alpha_l \mathbf{H}_l \mathbf{Q}_l \mathbf{H}_l^\dagger) \\ &= \sum_{i=1}^{z_l} W \left(\log_2(\rho_l \alpha_l \mu \lambda_i)\right)_+, \end{aligned}$$

where W represents the bandwidth of the channel; $\mathbf{Q}_l = \mathbb{E}\{\mathbf{x}_l \mathbf{x}_l^\dagger\}$ is the input covariance matrix representing the power allocation of signal $\mathbf{x}_l$; $\det(\cdot)$ represents matrix determinant; $\mathbf{I}$ represents an identity matrix; $(\cdot)_+$ represents $\max(0, \cdot)$; λ_i denotes an eigenvalue of matrix $\mathbf{H}_l \mathbf{H}_l^\dagger$ (having the same number of non-zero singular values in $\mathbf{\Lambda}_l$ and equal to the square of the singular values of $\mathbf{H}_l$); and μ is the optimal water-level satisfying $\sum_{i=1}^{z_l} (\mu - (\rho_l \alpha_l \lambda_i)^{-1})_+ = 1$.

Further, since $\mathbf{H}_l$ is a random matrix, the ergodic capacity of such a fading MIMO channel can be computed as [6]:

$$\begin{aligned} C_{l,\mathrm{ergodic}}^{(\mathrm{wf})} &= \mathbb{E}_{\mathbf{H}_l}[C_l^{(\mathrm{wf})}] = W \sum_{i=1}^{z_l} \mathbb{E}_{\lambda_i}\left[\left(\log_2(\rho_l \alpha_l \mu \lambda_i)\right)_+\right] \\ &= W \textstyle\sum_{i=1}^{z_l} \int \left(\log_2(\rho_l \alpha_l \mu \lambda_i)\right)_+ f_{\lambda_i}(\lambda) d\lambda \ , \end{aligned} \tag{4}$$

where $\mathbb{E}_{\lambda_i}[\cdot]$ represents the expectation taken over the distribution of λ_i and $f_{\lambda_i}(\cdot)$ denotes the distribution of λ_i. Although (4) is the exact formula for computing MIMO channel capacity, there are some issues that prevent (4) from being easily adopted in cross-layer optimization.

1) To determine the eigenvalues λ_i of $\mathbf{H}_l \mathbf{H}_l^\dagger$, one needs to solve the characteristic polynomial equation. However, even for a cubic or quartic polynomial equation (corresponding to 3×3 and 4×4 MIMO channels), the formula for roots computation is cumbersome to use and the polynomial equation is often solved approximately by numerical methods instead. Further, due to the complexity of computing λ_i from $\mathbf{H}_l \mathbf{H}_l^\dagger$, it is even harder to determine the distribution $f_{\lambda_i}(\cdot)$ from $\mathbf{H}_l$.

2) Even if we have solved λ_i's for a given $\mathbf{H}_l$, it remains to solve the optimal water-level μ for the optimal power allocation. However, due to the property of the water-filling algorithm, there is no closed-form solution to determine μ. Instead, μ can only be evaluated numerically.

3) Since it is difficult to determine λ_i, $f_{\lambda_i}(\cdot)$, and μ, computing the integration in (4) becomes a challenging task. Instead of integrating $\left(\log_2(\rho_l \alpha_l \mu \lambda_i)\right)_+ f_{\lambda_i}(\cdot)$, we can calculate a sample mean of $\left(\log_2(\rho_l \alpha_l \mu \lambda_i)\right)_+$ as an approximation for the expectation. However, this calculation requires a large number of random samples of $\mathbf{H}_l$ (so as to obtain a good approximation).

Due to the above difficulties, Eq. (4) cannot be readily used to offer tractable analysis in cross-layer optimization.

B. A Simple and Accurate Model for MIMO Channel Capacity

To avoid the difficulties incurred by using (4), we propose a simple model to approximate the MIMO link capacity computation as:

$$C_l^{(\mathrm{sim})} = W \cdot z_l \cdot \log_2\left(1 + \frac{\rho_l \alpha_l}{z_l}\right), \tag{5}$$

The construction of (5) to approximate the MIMO channel capacity is based on the following intuition. First, note that in (4), the capacity is determined by the averaging behavior of the eigenvalues of $\mathbf{H}_l \mathbf{H}_l^\dagger$. Although these eigenvalues are random, in practice they tend to be i.i.d. faded. As a result, when averaged over a large number of channel realizations, the mean channel gain for each parallel spatial channel (see Fig. 2) is roughly the same. Therefore, we approximate the random matrix $\mathbf{H}_l$ by a *deterministic* identity matrix (i.e., we replace $\mathbf{H}_l$ by $\mathbf{I}_{z_l}$), thus eliminating the expectation computation. With such a simplification, it is easy to verify that the optimal water-filling scheme degenerates into a trivial equal power allocation since all spatial channels have equal gain. It then follows that the channel capacity can be roughly approximated by (5). The main benefit of (5) is that we no longer need to explicitly compute the eigenvalues of $\mathbf{H}_l \mathbf{H}_l^\dagger$, the p.d.f. of λ_i, the optimal water level μ, and the expectation function. Also, note that when $z_l = 1$, (5) is reduced to Shannon formula.

We now formally examine the accuracy of (5). First, we quantify the gap between (4) and (5) for a single channel realization. We have the following lemma and its proof is given in Appendix A.

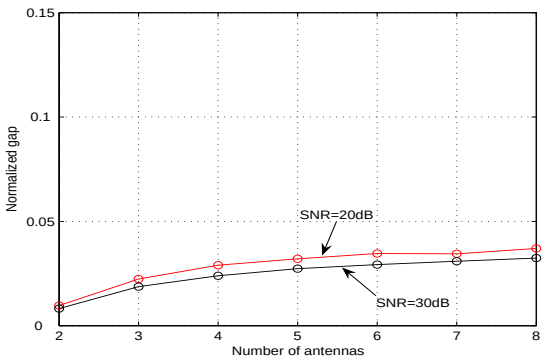

Fig. 3. Normalized gap vs. number of antennas.

Lemma 1. *For a MIMO link with instantaneous channel gain* $\mathbf{H}_l$ *of rank* z_l, $\Delta C_l \triangleq C_l^{(\mathrm{wf})} - C_l^{(\mathrm{sim})} \approx W \sum_{i=1}^{z_l} \log_2 \lambda_i$ *under high SNR regime.*

Based on Lemma 1, we show the gap between (4) and (5) is small by showing $\mathbb{E}_{\mathbf{H}_l}[\sum_{i=1}^{z_l} \log \lambda_i]$ is negligibly small. We state the result in the following theorem and give a detailed proof in Appendix B.

Theorem 1. *Under high SNR regime, for a MIMO link with Gaussian random channel matrix* $\mathbf{H}_l$ *of rank* z_l, *the approximation gap incurred by the simple model in (5) is close to zero.*

To offer some quantitative insights on this gap, we plot in Fig. 3 the normalized gap between (4) and (5) for a MIMO channel under 20dB and 30dB of SNRs, respectively. We vary the number of antennas from 2 to 8 (range for practical MIMO systems). We can see that the gap between the approximation and the exact capacity is indeed negligibly small. For example, with 4 antennas under 20dB, the gap is only 2.3%.

III. Link Layer MIMO Modeling for Ad Hoc Networks

At the link layer, MIMO opens up new opportunities in space domain to mitigate interference. In Section III-A, we first describe zero forcing beamforming (ZFBF), which is a powerful MIMO interference mitigation technique. We also discuss its benefits and challenges in ad hoc networks applications. In Section III-B, we propose a space-time scheduling scheme called SUCCINCT and in Section III-C, we construct its mathematical model.

A. Zero-Forcing Beamforming: Benefits and Challenges

In MIMO point-to-point or cellular MIMO systems, one of the most powerful interference mitigation transceiver designs is called zero-forcing beamforming (ZFBF) [3], [19]. ZFBF uses multi-antenna arrays to steer beams toward the intended receiver to increase SNR, while forming nulls to unintended receivers to avoid interference. To see how ZFBF can be used in MIMO ad hoc networks, consider an ad hoc network having L links among which L_0 links are active. We denote $\bar{\mathcal{I}}_l$ the set of links that interfere with the reception of link l's intended receiver, $l = 1, 2, \ldots, L_0$. We denote $\mathbf{H}_{\mathrm{Tx}(m),\mathrm{Rx}(l)}$ the interference channel gain matrix from transmitting node of interference link m ($\mathrm{Tx}(m)$) to receiving node of link l ($\mathrm{Rx}(l)$).

Recall from Eq. (3) that, when extracting the transmitted signal through a MIMO channel, a transmit beamforming matrix and a receive beamforming matrix are employed on the channel. Thus, the received signal at link l can be written as (for simplicity, we drop the tildes for $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{n}$):

$$\mathbf{y}_l = \underbrace{\sqrt{\rho_l \alpha_l}\, \mathbf{R}_{\mathrm{Rx}(l)}^{\dagger} \mathbf{H}_l \mathbf{T}_{\mathrm{Tx}(l)} \mathbf{x}_l}_{\text{Desired signal part}} + \underbrace{\sum_{m \in \bar{\mathcal{I}}_l} \sqrt{\rho_{m,l} \alpha_m} \mathbf{R}_{\mathrm{Rx}(l)}^{\dagger} \mathbf{H}_{\mathrm{Tx}(m),\mathrm{Rx}(l)} \mathbf{T}_{\mathrm{Tx}(m)} \mathbf{x}_m}_{\text{Interference part}} + \mathbf{n}_l, \ \forall l, \quad (6)$$

where $\rho_{m,l}$ denotes the interference-to-noise ratio (INR) from node $\mathrm{Tx}(m)$ to node $\mathrm{Rx}(l)$.

By exploiting the multi-antenna array at each node, it is possible to cancel out all interferences by judiciously configuring $\mathbf{T}$'s and $\mathbf{R}$'s. Specifically, we can configure $\mathbf{T}$'s and $\mathbf{R}$'s in such a way that

$$\mathbf{R}_{\mathrm{Rx}(l)}^{\dagger} \mathbf{H}_{\mathrm{Tx}(m),\mathrm{Rx}(l)} \mathbf{T}_{\mathrm{Tx}(m)} = \mathbf{0}, \quad \forall l, m \in \bar{\mathcal{I}}_l. \quad (7)$$

Note that if there exist non-trivial solutions to (7) (i.e., $\mathbf{R}_{\mathrm{Rx}(l)} \neq \mathbf{0}$, $\mathbf{T}_{\mathrm{Tx}(m)} \neq \mathbf{0}$, $\forall l, m \in \bar{\mathcal{I}}_l$), then it means that all L_0 links can be active simultaneously in an *interference-free* environment. Moreover, the ranks of $\mathbf{T}_{\mathrm{Tx}(l)}$ and $\mathbf{R}_{\mathrm{Rx}(l)}$ determine the number of data streams z_l that can be transmitted over link l, i.e., $z_l = \min\{\mathrm{rank}(\mathbf{T}_{\mathrm{Tx}(l)}), \mathrm{rank}(\mathbf{R}_{\mathrm{Rx}(l)})\}$.

Although ZFBF's benefits are appealing, there remain significant challenge to employ it in large ad hoc networks. This is because finding an optimal set of $\mathbf{T}$'s and $\mathbf{R}$'s satisfying (7) requires solving a large number of *bilinear* equations. Unlike linear equation systems, a general solution to bilinear equation systems remains unknown [18]. Thus, it becomes an intractable problem to design scheduling schemes based on solving (7).

B. SUCCINCT: Basic Idea

We find that the specific element configurations in $\mathbf{T}$'s and $\mathbf{R}$'s are more closely tied to beamforming design than to link layer scheduling. Therefore, instead of focusing on solving (7), we propose to reposition ourselves to exploit matrix dimension constraints that are sufficient for (7) to hold. By doing so, we can characterize the link layer scheduling performance without entangling with the details of beamforming designs.

To understand how we can extract the matrix dimension constraints for ZFBF-based scheduling, we first use a simple two-link one-interfering-node network shown in Fig. 4 as an example. In this network, link l has 3 antennas on each side and link m has 5 antennas on each side. For this simple network, we can schedule the transmissions of links l and m as follows. First, we arbitrarily choose a transmit beamforming matrix $\mathbf{T}_{\mathrm{Tx}(l)}$ without considering link m's existence. Suppose that $\mathbf{T}_{\mathrm{Tx}(l)}$ is full-rank, meaning that 3 data streams are

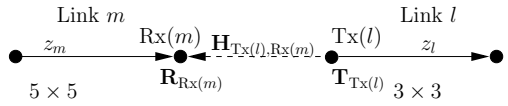

Fig. 4. A two-link one-interfering-node example.

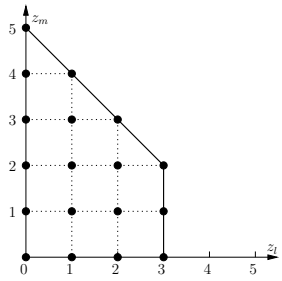

Fig. 5. DoF region of two-link one-interfering-node example.

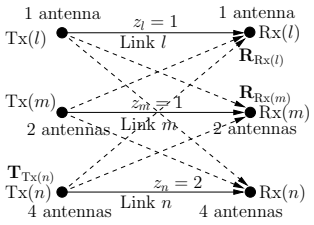

Fig. 6. A three-link multiple-interfering-node example.

being transmitted. Next, we choose a $\mathbf{R}_{\mathrm{Rx}(m)}$ to cancel the interference from link l, while receiving data streams from its desired transmitter. To do this, we need to solve for $\mathbf{R}_{\mathrm{Rx}(m)}$ such that $(\mathbf{H}_{\mathrm{Tx}(l),\mathrm{Rx}(m)}\mathbf{T}_{\mathrm{Tx}(l)})^\dagger\mathbf{R}_{\mathrm{Rx}(m)} = \mathbf{0}$, where $\mathbf{T}_{\mathrm{Tx}(l)}$ is already determined. From basic linear algebra, we know that all the receive beamforming vectors in $\mathbf{R}_{\mathrm{Rx}(m)}$ have to lie in the null space of $(\mathbf{H}_{\mathrm{Tx}(l),\mathrm{Rx}(m)}\mathbf{T}_{\mathrm{Tx}(l)})^\dagger$, denoted by $\mathrm{null}((\mathbf{H}_{\mathrm{Tx}(l),\mathrm{Rx}(m)}\mathbf{T}_{\mathrm{Tx}(l)})^\dagger)$. Note that the dimension of the null space in this case is $\dim(\mathrm{null}((\mathbf{H}_{\mathrm{Tx}(l),\mathrm{Rx}(m)}\mathbf{T}_{\mathrm{Tx}(l)})^\dagger)) = 5-3 = 2$. This implies that $\mathrm{Rx}(m)$ can receive up to 2 streams from $\mathrm{Tx}(m)$ (or link m can have 2 DoFs for receiving). Note that in this case links l and m are both active in an *interference-free* environment. Further, by varying the ranks of $\mathbf{T}_{\mathrm{Tx}(l)}$ and $\mathbf{R}_{\mathrm{Rx}(m)}$, it is easy to verify that the achievable DoF region under our ZFBF-based scheduling scheme is the trapezoid shown in Fig. 5.[1]

Observe that the scheduling scheme in the above two-link one-interfering-node example is performed in a *sequential* fashion: we arbitrarily choose a $\mathbf{T}_{\mathrm{Tx}(l)}$ first, and then choose a $\mathbf{R}_{\mathrm{Rx}(m)}$ such that the interference can be eliminated.

We now extend this successive interference cancelation idea to a three-link multiple-interfering-node example shown in Fig. 6, which is much more complicated than the earlier example. Here, each receiving node of a link is being interfered by the transmitting nodes of other links. To schedule transmission/reception, we can start with a scheduling order for the six nodes. Such scheduling order will be subject to optimization in Section III-C. For example, suppose the scheduling order for the six nodes is $\mathrm{Tx}(l) \to \mathrm{Rx}(m) \to \mathrm{Rx}(l) \to \mathrm{Tx}(m) \to \mathrm{Tx}(n) \to \mathrm{Rx}(n)$. Then the following scheduling decisions will take place.

1) $\mathrm{Tx}(l)$: Since nodes $\mathrm{Tx}(l)$ is the first to be scheduled, it does not have any interference to concern about. Also, since $\mathrm{Tx}(l)$ has only 1 antenna, we can let $\mathrm{Tx}(l)$ transmit 1 data stream;

[1]Note that the achievable DoF region in Fig. 5 coincides with the maximum DoF region described in [11, Theorem 2]. Thus, for this two-link example, the proposed scheduling scheme is an optimal scheduling scheme.

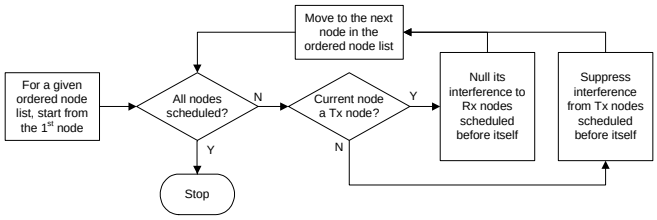

Fig. 7. The flow chart of SUCCINCT scheduling algorithm.

2) $\mathrm{Rx}(m)$: Since $\mathrm{Rx}(m)$ is scheduled after $\mathrm{Tx}(l)$, it needs to suppress the interference from $\mathrm{Tx}(l)$, i.e., solving $(\mathbf{H}_{\mathrm{Tx}(l),\mathrm{Rx}(m)}\mathbf{T}_{\mathrm{Tx}(l)})^\dagger\mathbf{R}_{\mathrm{Rx}(m)} = \mathbf{0}$. We have $\dim(\mathrm{null}((\mathbf{H}_{\mathrm{Tx}(l),\mathrm{Rx}(m)}\mathbf{T}_{\mathrm{Tx}(l)})^\dagger)) = 2-1 = 1$, i.e., we can let $\mathrm{Rx}(m)$ receive 1 stream in this case;

3) $\mathrm{Rx}(l)$: Since no interfering transmitting node is scheduled before $\mathrm{Rx}(l)$, $\mathrm{Rx}(l)$ does not need to concern about any interference. Given $\mathrm{Rx}(l)$ has only 1 antenna, we can let it receive 1 stream;

4) $\mathrm{Tx}(m)$: Following the similar argument as for $\mathrm{Rx}(m)$, we can let $\mathrm{Tx}(m)$ transmit 1 stream;

5) $\mathrm{Tx}(n)$: Since $\mathrm{Tx}(n)$'s transmission should not interfere with $\mathrm{Rx}(l)$ and $\mathrm{Rx}(m)$, it follows that $\mathbf{T}_{\mathrm{Tx}(n)}$ should satisfy $\begin{bmatrix} \mathbf{R}^\dagger_{\mathrm{Rx}(l)}\mathbf{H}_{\mathrm{Tx}(n),\mathrm{Rx}(l)} \\ \mathbf{R}^\dagger_{\mathrm{Rx}(m)}\mathbf{H}_{\mathrm{Tx}(n),\mathrm{Rx}(m)} \end{bmatrix}\mathbf{T}_{\mathrm{Tx}(n)} = \mathbf{0}$. Since $\dim\left(\mathrm{null}\begin{bmatrix} \mathbf{R}^\dagger_{\mathrm{Rx}(l)}\mathbf{H}_{\mathrm{Tx}(n),\mathrm{Rx}(l)} \\ \mathbf{R}^\dagger_{\mathrm{Rx}(m)}\mathbf{H}_{\mathrm{Tx}(n),\mathrm{Rx}(m)} \end{bmatrix}\right) = 4-(1+1) = 2$, we can schedule 2 data streams at $\mathrm{Tx}(n)$;

6) $\mathrm{Rx}(n)$: Similar analysis as in 5) can be done for $\mathrm{Rx}(n)$ to show that 2 streams can be scheduled at $\mathrm{Rx}(n)$.

The idea in the three-link multiple-interfering-node example can be synthesized for a general multiple-link setting. The essence of this scheduling scheme is to perform interference cancelation successively on each node in an ordered node list:

- If a node is transmitting, then it is only necessary to ensure that its transmissions do not interfere with previously scheduled receiving nodes in the ordered node list. It does not need to expend precious DoF resource to null its interference to those receiving nodes to be scheduled after itself in the node list.
- If a node is receiving, it only needs to suppress interference from transmitting nodes scheduled before itself in the node list. It does not need to concern interfering transmitting nodes to be scheduled after itself.

The interference cancelation behavior described above offers the basic idea for node-based scheduling algorithm. For easy reference, we call this scheduling scheme SUCCINCT (successive interference cancelation). Additional quantitative constraints on DoF on each transmitting and receiving node (as shown in previous two examples) will be discussed in the following section. Fig. 7 shows the flow chart of SUCCINCT algorithm.

Remark 1. In [8], Hamdaoui and Shin proposed several interference avoidance schemes based on ZFBF. For the so-called CiM scheme (the best among the proposed schemes

in [8]), the authors also recognized that interference can be canceled by either the transmitting or the receiving node of an interference link. However, without employing node-based sequential scheduling, it is impossible to know which node should perform interference mitigation. As a result, the CiM scheme requires both the transmitting and receiving nodes of an interference link to expend precious DoFs for interference cancelation (c.f. [8, Eq. (10)]). This approach adversely leads to a much smaller DoF region. As an example, we compare the performance of SUCCINCT and the CiM model on the simple two-link one-interfering-node example in Fig. 4. Under the CiM model, it is not difficult to show that the achievable DoF region is the shaded triangle in Fig. 8, which is much smaller than the achievable DoF region by SUCCINCT. The detailed analysis on why this is the case is provided in Appendix C. In general, it can be shown that the DoF region achieved under the CiM model is always a subset of that under the SUCCINCT scheme [13]. The details of the proof can also be found in Appendix C.

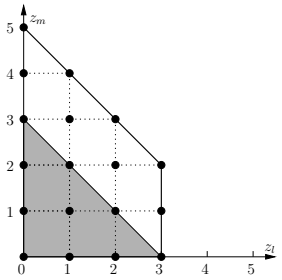

Fig. 8. Achievable DoFs comparison between SUCCINCT and CiM for the example in Fig. 4.

C. SUCCINCT: A Mathematical Model

Having introduced the basic idea of SUCCINCT, we now develop its mathematical model. We represent the topology of a MIMO ad hoc network by a directed graph, denoted by $\mathcal{G} = \{\mathcal{N}, \mathcal{L}\}$, where $\mathcal{N}$ and $\mathcal{L}$ are the sets of nodes and all possible MIMO links, respectively. Suppose that the cardinalities of the sets $\mathcal{N}$ and $\mathcal{L}$ are $|\mathcal{N}| = N$ and $|\mathcal{L}| = L$, respectively. In this paper, we assume that scheduling operates in on period frame-by-frame system with T time slots in each frame.

Ordered Node List: An Optimization Model. Referring to Fig. 7 and the discussion in Section III-B, before we start scheduling on a node, we must have an ordered node list, which itself is subject to optimization.

To model an ordered node list that can be optimized, we define the following binary variable. For $i, j \in \mathcal{N}$, $i \neq j$, we let

$$\pi_{ij}(t) = \begin{cases} 1 & \text{if node } j \text{ is after node } i \text{ in time slot } t, \\ 0 & \text{if node } j \text{ is before node } i \text{ in time slot } t. \end{cases}$$

To model an ordered node list, $\pi_{ij}(t)$ must satisfy the following two properties.

1) *Mutual exclusiveness*: If node j is scheduled after node i (i.e., $\pi_{ij}(t) = 1$), then it also implies that node i is before node j (i.e., $\pi_{ji}(t) = 0$). This relationship can be modeled as

$$\pi_{ij}(t) + \pi_{ji}(t) = 1, \quad \forall i, j \in \mathcal{N} : i \neq j, \ \forall t. \tag{8}$$

Transitivity: If node j is scheduled after i (i.e., $\pi_{ij} = 1$) and node k is scheduled after node j (i.e., $\pi_{jk} = 1$), then it implies that node k is scheduled after node i (i.e., $\pi_{ik} = 1$). We enumerate all possible cases for π_{ij} and π_{jk} in Table II and show what π_{ik} will be. In Table II, "indefinite" means that $\pi_{ik}(t)$ cannot be determined by the current settings of $\pi_{ij}(t)$ and $\pi_{jk}(t)$. Mathematically, the relationship in Table II can be modeled as

TABLE II
ENUMERATING THE TRANSITIVITY RELATIONSHIP.

$\pi_{ij}(t)$	$\pi_{jk}(t)$	$\pi_{ik}(t)$
1	1	1
1	0	indefinite
0	1	indefinite
0	0	0

$$\pi_{ik}(t) \leq \pi_{ij}(t) + \pi_{jk}(t), \quad \forall i, j, k \in \mathcal{N}, \forall t, \tag{9}$$

$$\pi_{ik}(t) \geq \pi_{ij}(t) + \pi_{jk}(t) - 1, \quad \forall i, j, k \in \mathcal{N}, \forall t. \tag{10}$$

For example, for the second row in Table II, when $\pi_{ij}(t) = 1$ and $\pi_{jk}(t) = 0$, (9) gives $\pi_{ik}(t) \leq 1$ and (10) gives $\pi_{ik}(t) \geq 0$. Given that $\pi_{ik}(t)$ is a binary variable, this is equivalent to saying that $\pi_{ik}(t)$ remains to be determined (or indefinite).

Note that according to (9) and (10), we can write 12 different constraints in 6 groups for three nodes i, j, $k \in \mathcal{N}$ as follows.

$$\begin{cases} \pi_{ij}(t) \leq \pi_{ik}(t) + \pi_{kj}(t) \\ \pi_{ij}(t) \geq \pi_{ik}(t) + \pi_{kj}(t) - 1, \end{cases} \tag{11}$$

$$\begin{cases} \pi_{ji}(t) \leq \pi_{jk}(t) + \pi_{ki}(t) \\ \pi_{ji}(t) \geq \pi_{jk}(t) + \pi_{ki}(t) - 1, \end{cases} \tag{12}$$

$$\begin{cases} \pi_{ik}(t) \leq \pi_{ij}(t) + \pi_{jk}(t) \\ \pi_{ik}(t) \geq \pi_{ij}(t) + \pi_{jk}(t) - 1, \end{cases} \tag{13}$$

$$\begin{cases} \pi_{ki}(t) \leq \pi_{kj}(t) + \pi_{ji}(t) \\ \pi_{ki}(t) \geq \pi_{kj}(t) + \pi_{ji}(t) - 1, \end{cases} \tag{14}$$

$$\begin{cases} \pi_{jk}(t) \leq \pi_{ji}(t) + \pi_{ik}(t) \\ \pi_{jk}(t) \geq \pi_{ji}(t) + \pi_{ik}(t) - 1, \end{cases} \tag{15}$$

$$\begin{cases} \pi_{kj}(t) \leq \pi_{ki}(t) + \pi_{ij}(t) \\ \pi_{kj}(t) \geq \pi_{ki}(t) + \pi_{ij}(t) - 1. \end{cases} \tag{16}$$

A closer look at these 6 groups of constraints show that any one group can be used to derive the other 5 groups of constraints (see Appendix D). In other words, any one group from the six groups in (11)–(16) is sufficient to describe the transitivity property for a node triplet $\{i, j, k\}$.

To select only one group out of (11)–(16) to describe the transitivity property, we need to maintain certain consistency in such selection. Our approach to achieve such consistency is to create a mapping $\Omega(\cdot) : \mathcal{N} \to \mathbb{N}$, where each node in $\mathcal{N}$ is mapped to an integer number in $\mathbb{N} = \{1, 2, \ldots, N\}$. Now,

without loss of generality, suppose that nodes i, j, and k satisfy $\Omega(i) < \Omega(j) < \Omega(k)$. We will select group constraint (11), which is the same as selecting two constraints from (9) and (10) such that $\Omega(i) < \Omega(j) < \Omega(k)$. Using such a mapping and the constraint $\Omega(i) < \Omega(j) < \Omega(k)$, we can consistently and uniquely identify a group out of 6 groups for any node triplet. Now, combining the ordered mapping and (9) and (10), we have:

$$\pi_{ik}(t) \leq \pi_{ij}(t) + \pi_{jk}(t), \quad \forall i,j,k \in \mathcal{N} : \Omega(i) < \Omega(j) < \Omega(k), \forall t, \tag{17}$$

$$\pi_{ik}(t) \geq \pi_{ij}(t) + \pi_{jk}(t) - 1, \quad \forall i,j,k \in \mathcal{N} : \Omega(i) < \Omega(j) < \Omega(k), \forall t. \tag{18}$$

We further show in Appendix D that (17) and (18) can be written in a more compact form: $1 \leq \pi_{ij}(t)+\pi_{jk}(t)+\pi_{ki}(t) \leq 2$. We summarize this result in the following lemma.

Lemma 2. *For nodes $i,j,k \in \mathcal{N}$ such that $\Omega(i) < \Omega(j) < \Omega(k)$, the following two constraints are sufficient to describe the transitivity relationship among nodes node triplet i, j, and k:*

$$1 \leq \pi_{ij}(t) + \pi_{jk}(t) + \pi_{ki}(t) \leq 2. \tag{19}$$

It is not hard to see that Lemma 2 will significantly reduce computational complexity in finding the optimal spatial scheduling order. Also note that loop is not allowed in the transitivity relationship in (19). This is because that (19) implies that at least 1 and at most 2 π-variables can be equal to one. The cases when loop occurs, i.e., either $\pi_{ij}(t) = \pi_{jk}(t) = \pi_{ki}(t) = 0$ or $\pi_{ij}(t) = \pi_{jk}(t) = \pi_{ki}(t) = 1$, are not allowed in (19).

A Model for Transmitting Node Behavior. Next, we model the block in Fig. 7 where a node i is scheduled to be a transmitting node. Note that in each time slot t, $1 \leq t \leq T$, due to half-duplex, each node either transmit, receive, or be idle. To model half-duplex, we introduce two groups of binary variables $g_i(t)$'s and $h_i(t)$'s as follows. $g_i(t) = 1$ if node i is transmitting in time slot t and 0 otherwise; $h_i(t) = 1$ if node i is receiving in time slot t and 0 otherwise. Then, the half-duplex constraint can be characterized by

$$g_i(t) + h_i(t) \leq 1, \quad \forall i,t. \tag{20}$$

We assume that scattering is rich enough in the environment such that all channel matrices are of full-rank. As a result, the number of data streams that a node can transmit or receive is limited by its number of antennas and we have the following two constraints:

$$g_i(t) \leq \sum_{l \in \mathcal{L}_i^{\mathrm{Out}}} z_l(t) \leq g_i(t) A_i, \tag{21}$$

$$h_i(t) \leq \sum_{l \in \mathcal{L}_i^{\mathrm{In}}} z_l(t) \leq h_i(t) A_i, \tag{22}$$

where $\mathcal{L}_i^{\mathrm{Out}}$ and $\mathcal{L}_i^{\mathrm{In}}$ represent the sets of outgoing and incoming links at node i, respectively; $z_l(t)$ denotes the number of data streams over link l in time slot t, and A_i represents the number of antennas at node i.

From Fig. 7, we see that the data streams transmitted by node i should not interfere with those receiving nodes scheduled previously. This is equivalent to saying that the transmission beamforming vectors in $\mathbf{T}_i$ should lie in the null space of the stacked matrix formed by stacking all $\mathbf{R}_j^\dagger \mathbf{H}_{i,j}$ matrices, where j denotes a previously scheduled receiving node that could be interfered by node i. That is,

$$\mathbf{T}_i \in \mathrm{null} \begin{bmatrix} \vdots \\ \mathbf{R}_j^\dagger \mathbf{H}_{i,j} \\ \vdots \end{bmatrix}, \begin{array}{l} j \in \mathcal{I}_i \text{ and } j \text{ is} \\ \text{scheduled before } i \\ \text{(i.e., } \pi_{ji} = 1\text{)}, \end{array} \tag{23}$$

where $\mathcal{I}_i$ represents the set of nodes within the interference range of node i. For convenience, we let $\mathbf{S}$ denote the stacked matrix in (23). Note that $\sum_{l \in \mathcal{L}_i^{\mathrm{Out}}} z_l(t)$ is the number of data streams that node i transmits in time slot t. Thus, from (23), we have that $\sum_{l \in \mathcal{L}_i^{\mathrm{Out}}} z_l(t)$ should be less than or equal to the nullity of of $\mathbf{S}$, i.e., $\sum_{l \in \mathcal{L}_i^{\mathrm{Out}}} z_l(t) \leq \mathrm{null}(\mathbf{S})$. Also, note that the rank of $\mathbf{S}$ is $\sum_{j \in \mathcal{I}_i} \pi_{ji}(t) \sum_{l:\mathrm{Rx}(l)=j}^{\mathrm{Tx}(l) \neq i} z_l(t)$. Therefore, according to rank-nullity theorem [9] (i.e., $\mathrm{rank}(\mathbf{S}) + \mathrm{null}(\mathbf{S})$ is equal to the number of columns in $\mathbf{S}$), we can model the dimensional constraint as follows: for all $i,j \in \mathcal{N}$ and for all t,

$$\sum_{l \in \mathcal{L}_i^{\mathrm{Out}}} z_l(t) + \sum_{j \in \mathcal{I}_i} \pi_{ji}(t) \sum_{\substack{l:\mathrm{Rx}(l)=j \\ \mathrm{Tx}(l) \neq i}} z_l(t) \leq A_i + (1 - g_i(t))M. \tag{24}$$

In (24), M is a sufficiently large number (e.g., we can set $M = \sum_{j \in \mathcal{I}_i} A_i$). When node i is a transmission node (i.e., $g_i(t) = 1$), then (24) is reduced to the rank-nullity condition with respect to $\mathbf{S}$. Otherwise, if node i is scheduled to be a receiving node or in idle status (i.e., $g_i(t) = 0$), then (23) becomes irrelevant due to the large value of M.

We note that the nonlinear terms $\pi_{ji}(t) \sum_{l:\mathrm{Rx}(l)=j}^{\mathrm{Tx}(l) \neq i} z_l(t)$ in (24) could complicate optimization algorithms design. To remove these nonlinear terms, we can introduce a new integer variable $\phi_{ji}(t)$ and reformulate (24) as follows:

$$\sum_{l \in \mathcal{L}_i^{\mathrm{Out}}} z_l(t) + \sum_{j \in \mathcal{I}_i} \phi_{ji}(t) \leq A_i + (1 - g_i(t))M, \tag{25}$$

where $\phi_{ji}(t)$, $\forall i \in \mathcal{N}$, $j \in \mathcal{I}_i$, $\forall t$, satisfies the following constraints:

$$\phi_{ji}(t) \leq \textstyle\sum_{l:\mathrm{Rx}(l)=j}^{\mathrm{Tx}(l) \neq i} z_l(t), \tag{26}$$

$$\phi_{ji}(t) \leq A_i \pi_{ji}(t), \tag{27}$$

$$\phi_{ji}(t) \geq A_i \pi_{ji}(t) - A_i + \textstyle\sum_{l:\mathrm{Rx}(l)=j}^{\mathrm{Tx}(l) \neq i} z_l(t). \tag{28}$$

It is easy to verify that this set of new constraints (25)–(28) is equivalent to (24).

A Model for Receiving Node Behavior. Similarly, for the block in Fig. 7 where a node i is scheduled to be a receiving node, we can derive the following constraints: for all $i,j \in \mathcal{N}$

and for all t,

$$\sum_{l\in\mathcal{L}_i^{\rm In}} z_l(t) + \sum_{j\in\mathcal{I}_i} \varphi_{ji}(t) \le A_i + (1-h_i(t))M, \quad (29)$$

$$\varphi_{ji}(t) \le \sum_{l:{\rm Tx}(l)=j}^{{\rm Rx}(l)\neq i} z_l(t), \quad (30)$$

$$\varphi_{ji}(t) \le A_i \pi_{ji}(t), \quad (31)$$

$$\varphi_{ji}(t) \ge A_i \pi_{ji}(t) - A_i + \sum_{l:{\rm Tx}(l)=j}^{{\rm Rx}(l)\neq i} z_l(t). \quad (32)$$

IV. Application in Multi-Hop Ad Hoc Networks

In Sections II and III, we have developed two models for the physical layer and the link layer in multi-hop MIMO ad hoc networks, respectively. In this section, we will show how to apply them for cross-layer optimization in multi-hop MIMO ad hoc networks. We consider a generic utility maximization problem involving a set of sessions, $\mathcal{F}$, in an ad hoc network. Denote ${\rm src}(f)$ and ${\rm dst}(f)$ the source and destination nodes of session $f \in \mathcal{F}$, respectively. Denote $r(f)$ the rate of session f and $r_l(f)$ the amount of rate on link l that is attributed to session $f \in \mathcal{F}$, respectively. Denote $C_l(t)$ the capacity of link l in time-slot t. For stability, we have the following constraints on the flow rates:

$$\sum_{f\in\mathcal{F}} r_l(f) \le \tfrac{1}{T}\sum_{t=1}^{T} C_l(t), \quad \forall l. \quad (33)$$

At the network layer, different routing schemes can be adopted. But regardless of specific routing schemes, the flow balance constraints must hold at each node $i \in \mathcal{N}$ in the network.

$$\sum_{l\in\mathcal{L}_i^{\rm Out}} r_l(f) = r(f), \quad \text{if } i = {\rm src}(f), \quad (34)$$

$$\sum_{l\in\mathcal{L}_i^{\rm Out}} r_l(f) = \sum_{l\in\mathcal{L}_i^{\rm In}} r_l(f), \text{ if } i \neq {\rm src}(f), {\rm dst}(f), \quad (35)$$

$$\sum_{l\in\mathcal{L}_i^{\rm In}} r_l(f) = r(f), \quad \text{if } i = {\rm dst}(f). \quad (36)$$

It can be easily verified that if (34) and (35) are satisfied, then (36) is automatically satisfied. As a result, there is not necessary to list (36) in a formulation once we have both (34) and (35).

When a node is transmitting simultaneously on more than one outgoing links, it is necessary to consider power allocation among $\mathcal{L}_i^{\rm Out}$ at node i. Recall that $\alpha_l(t) \in [0,1]$ represent a fraction of transmit power allocated onto link l in time-slot t. Then, for each node i in time-slot t, we have

$$\sum_{l\in\mathcal{L}_i^{\rm Out}} \alpha_l(t) \le g_n(t), \quad \forall n, t. \quad (37)$$

The constraint in (37) ensures that the sum of transmit power of all outgoing links at node i does not exceed the transmit power limit. In the case when node i is not in transmission mode, then $g_i(t) = 0$ and $\alpha_l(t) = 0$ for all $l \in \mathcal{L}_i^{\rm Out}$.

Consider an objective function for each session, $u\big(r(f)\big)$, which we assume is concave. Then a general MIMO network utility maximization (MIMO-NUM) problem can be formulated as follows.

MIMO-NUM

max $\sum_{f=1}^{F} u\big(r(f)\big)$

s.t. Network layer flow-balance routing constraints in (35) and (34);
Coupling constraints between network layer and link layer in (33);
SUCCINCT based link layer modeling constraints in (20), (21), (22), (8), (19) and (25)–(32);
Simplified MIMO physical layer model in (5) and (37).

Two remarks for the MIMO-NUM problem are in order. 1) *Tractability.* Recall that a MIMO cross-layer optimization involves many matrix variables in the capacity calculation and ZFBF scheduling, making network level research quite challenging. With our simple models, matrix variables no longer appear in MIMO-NUM problem, which significantly simplifies formulation and reduce computational complexity. 2) *Solvability.* By using our simple models, MIMO-NUM problem is reduced to a similar mathematical form as in NUM problems for single-antenna ad hoc networks. Note that although SUCCINCT part is unique, it is of linear form and can be easily handled mathematically. This suggests that new solutions to MIMO-NUM problems may be developed by drawing upon the rich experiences gained for single-antenna ad hoc networks.

As an example to illustrate the solvability of our MIMO-NUM formulation, we consider a MIMO ad hoc network consisting of 50 nodes that are uniformly distributed in a square region of 1500m $\times$ 1500m (see Fig. 9). Each node in the network is equipped with 4 antennas and the maximum power for each node is 100 mW. The channel bandwidth is 20 MHz. The path-loss index is 3.5. There are 5 sessions in the network: N26 to N19, N44 to N18, N24 to N15, N48 to N2, and N9 to N32, respectively. Suppose that minimum-hop routing is employed at the network layer. The objective is to maximize the sum of the end-to-end session rates, i.e., $u\big(r(f)\big) = r(f)$. Suppose that there are 4 time slots in each time frame, i.e., $T = 4$. Given these parameters and network settings, the MIMO-NUM problem is now completely specified. We can use CPLEX solver to obtain optimal solution.

The optimal scheduling ordering for each node in each time slot is listed in Table III. In this table, each column gives the node ordering for scheduling in a given time slot of the frame. For example, in the first time slot, the optimal ordering of the nodes is $N19 \to N18 \to \ldots \to N2$.

Fig. 10 shows the routing paths for each session and optimal scheduling solution (shown in shaded boxes). As an example, the shaded box next to the link from N6 to N18 contains $1:1; 2:1; 3:2$, which means that in time slots 1, 2, 3, there are 1, 1, and 2 streams on this link, respectively. In time slot 4, the link is not transmitting. Based on the number of streams, the simple physical layer model (5), and the constraint in (33), the optimal session rates are found to be (in Mb/s): 60.4 for

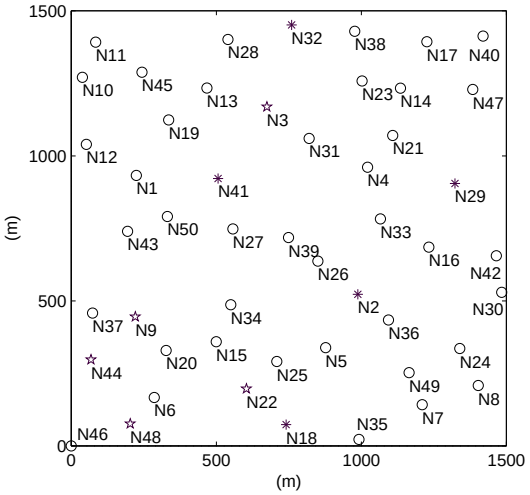

Fig. 9. A 50-node 5-session MIMO-based ad hoc network.

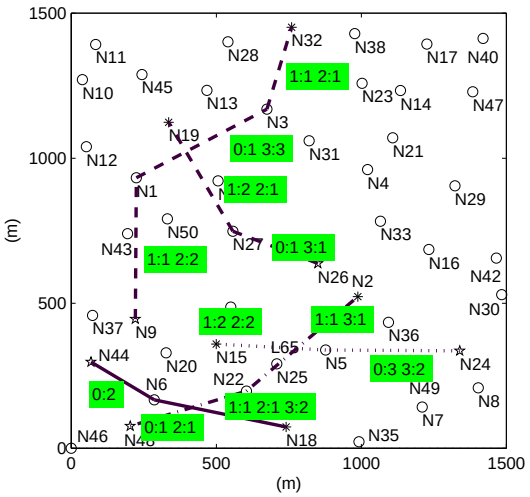

Fig. 10. Scheduling result on each link.

N26 → N19, 151 for N44 → N18, 102 for N24 → N15, 36.6 for N48 → N2, and 57.2 for N9 → N32.

TABLE III
OPTIMAL NODE ORDERING IN EACH TIME SLOT OF A FRAME.

	Time Slot 1	Time Slot 2	Time Slot 3	Time Slot 4
1st	N19	N48	N24	N24
2nd	N18	N27	N22	N32
3rd	N44	N32	N44	N26
4th	N15	N15	N9	N48
5th	N26	N9	N15	N18
6th	N22	N19	N27	N24
7th	N5	N44	N32	N9
8th	N48	N26	N26	N22
9th	N6	N18	N19	N5
10th	N27	N2	N18	N6
11th	N24	N24	N6	N15
12th	N32	N22	N5	N19
13th	N3	N5	N48	N2
14th	N1	N3	N1	N27
15th	N9	N1	N3	N1
16th	N2	N6	N2	N3

V. CONCLUSION

Existing models for MIMO suffer from either intractability or inaccuracy when they are employed for multi-hop ad hoc network. In this paper, we proposed a tractable and accurate model for MIMO that is amenable for cross-layer analysis in multi-hop ad hoc networks. Our contributions included a model at the physical layer and a model at the link layer. At the physical layer, we proposed a simple model to compute MIMO channel capacity that captures the essence of spatial multiplexing and transmit power limit without involving complex matrix operations and the water-filling algorithm. We proved that the approximation gap in this physical layer model is negligible. At the link layer, we proposed a scheduling scheme called SUCCINCT that is based on ZFBF interference mitigation. The proposed SUCCINCT scheme cuts through the complexity associated with beamforming designs in a multi-hop ad hoc network by using simple numeric computation on matrix dimensions. As a result, we can characterize the link layer scheduling performance without entangling with beamforming details. By applying the proposed cross-layer model to a general network utility maximization problem, we validate its efficacy in practice. The results in this paper offer an important analytical tool to fully exploit the potential of MIMO in multi-hop ad hoc networks.

REFERENCES

[1] R. Bhatia and L. Li, "Throughput optimization of wireless mesh networks with MIMO links," in *Proc. IEEE INFOCOM*, Anchorage, AK, May 6-12, 2007, pp. 2326–2330.

[2] E. Biglieri, R. Calderbank, A. Constantinides, A. Goldsmith, A. Paulraj, and H. V. Poor, *MIMO Wireless Communications*. Cambridge University Press, Jan. 2007.

[3] L.-U. Choi and R. D. Murch, "A transmit preprocessing technique for multiuser MIMO systems using a decomposition approach," *IEEE Trans. Wireless Commun.*, vol. 3, no. 1, pp. 20–24, Jan. 2004.

[4] S. Chu and X. Wang, "Opportunistic and cooperative spatial multiplexing in MIMO ad hoc networks," in *Proc. ACM Mobihoc*, Hong Kong SAR, China, May 26-30, 2008, pp. 63–72.

[5] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading envorinment when using multi-element antennas," *Bell Labs Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.

[6] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 1, pp. 684–702, Jun. 2003.

[7] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*. San Diego: Academic Press, 2000.

[8] B. Hamdaoui and K. G. Shin, "Characterization and analysis of multi-hop wireless MIMO network throughput," in *Proc. ACM Mobihoc*, Montréal, Québec, Canada, Sep. 2007, pp. 120–129.

[9] R. A. Horn and C. R. Johnson, *Matrix Analysis*. New York: Cambridge University Press, 1990.

[10] M. Hu and J. Zhang, "MIMO ad hoc networks: Medium access control, saturation throughput, and optimal hop distance," *Special Issue on Mobile Ad Hoc Networks, Journal of Communications and Networks*, pp. 317–330, Dec. 2004.

[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] S.-J. Kim, X. Wang, and M. Madihian, "Cross-layer design of wireless multihop backhaul networks with multiantenna beamforming," *IEEE Trans. Mobile Comput.*, vol. 6, no. 11, pp. 1259–1269, Nov. 2007.

[13] J. Liu, Y. Shi, and Y. T. Hou, "A tractable and accurate cross-layer model for multi-hop MIMO ad hoc networks," *Technical Report, Department of ECE, Virginia Tech*, Jul. 2009. [Online]. Available: http://filebox.vt.edu/users/kevinlau/publications

[14] M. L. Metha, *Random Matrices*, 3rd ed. London, UK: Academic Press, 2004.

[15] S. Y. Oh, M. Gerla, and J.-S. Park, "MIMO and TCP: A case for cross-layer design," in *Proc. IEEE MILCOM*, Orlando, FL, Oct. 29-31, 2007.

[16] S. Y. Oh, M. Gerla, P. Zhao, B. Daneshrad, G. Pei, and J. H. Kim, "MIMO-CAST: A cross-layer ad hoc multicast protocol using MIMO radios," in *Proc. IEEE MILCOM*, Orlando, FL, Oct. 29-31, 2007.

[17] J.-S. Park, A. Nandan, M. Gerla, and H. Lee, "SPACE-MAC: Enabling spatial reuse using MIMO channel-aware MAC," in *Proc. IEEE ICC*, Seoul, Korea, May 16-20, 2005, pp. 3642–3646.

[18] S. Roman, *Advanced Linear Algebra*. New York, NY: Springer, 2007.

[19] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser mimo channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461–471, Feb. 2004.

[20] K. Sundaresan and R. Sivakumar, "Routing in ad hoc networks with MIMO links," in *Proc. IEEE International Conf. on Network Protocols*, Boston, MA, U.S.A., Nov. 2005, pp. 85–98.

[21] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecomm.*, vol. 10, no. 6, pp. 585–596, Nov. 1999.

[22] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, UK: Cambridge University Press, 2005.

[23] A. M. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications*. Hanover, MA: now Publishers Inc., 2004.

APPENDIX A
PROOF OF LEMMA 1

Under high SNR regime, $\rho_l\alpha_l$ is large enough such that all spatial channels are active with high probability, i.e., $\sum_{i=1}^{z_l}(\mu-(\rho_l\alpha_l\lambda_i)^{-1})_+ = \sum_{i=1}^{z_l}(\mu-(\rho_l\alpha_l\lambda_i)^{-1}) = 1$. Thus, we have

$$z_l\mu = 1 + \textstyle\sum_{i=1}^{z_l}(\rho_l\alpha_l\lambda_i)^{-1}. \quad (38)$$

Also, we have that $C_l^{(\mathrm{wf})} = W\sum_{i=1}^{z_l}\log_2(\rho_l\alpha_l\mu\lambda_i)$. On the other hand, we can re-write the capacity formula for the simple model as $C_l^{(\mathrm{sim})} = W\sum_{i=1}^{z_l}\log_2\left(1+\frac{\rho_l\alpha_l}{z_l}\right)$. As a result, the capacity gap can be computed as

$$\begin{aligned}
\Delta C_l &= C_l^{(\mathrm{wf})} - C_l^{(\mathrm{sim})} = W\textstyle\sum_{i=1}^{z_l}\log_2\frac{\rho_l\alpha_l\mu\lambda_i}{1+\rho_l\alpha_l/z_l} \\
&\approx W\textstyle\sum_{i=1}^{z_l}\log_2 z_l\mu\lambda_i \quad (39)\\
&= W\textstyle\sum_{i=1}^{z_l}\log_2\left(1+(\rho_l\alpha_l)^{-1}\sum_{i=1}^{z_l}\lambda_i^{-1}\right)\lambda_i \quad (40)\\
&\approx W\textstyle\sum_{i=1}^{z_l}\log_2\lambda_i, \quad (41)
\end{aligned}$$

where (39) follows from the fact that in high SNR regime, $\rho_l\alpha_l/z_l \gg 1$ so that 1 can be ignored; (40) follows from substituting (38) into (39), and (41) follows from the fact that in high SNR regime, the term $(\rho_l\alpha_l)^{-1}\sum_{i=1}^{z_l}\lambda_i^{-1} \ll 1$ (none of λ_i can be too small since $\mathbf{H}_l$ is a well-conditioned matrix) and can be ignored. This completes the proof.

APPENDIX B
PROOF OF THEOREM 1

First of all, from Lemma 1, we have

$$\mathbb{E}_{\mathbf{H}_l}[\Delta C_l] \approx W\sum_{i=1}^{z_l}\mathbb{E}_{\lambda_i}[\log_2\lambda_i] \le W\sum_{i=1}^{z_l}\log_2\mathbb{E}_{\lambda_i}[\lambda_i].$$

where the last inequality follows from the concavity of the log function and Jensen's inequality. The Marčenko-Pastur theorem [14] says that for a matrix $\mathbf{H}_l$ with $\frac{n_r}{n_t} = \beta$, the limiting p.d.f. of the eigenvalues of the corresponding Wishart matrix $\mathbf{H}_l\mathbf{H}_l^\dagger$ as $n_t, n_r \to \infty$ is:

$$f_\lambda^\beta(x) = \left(1-\frac{1}{\beta}\right)_+\delta(x) + \frac{\sqrt{(x-l)_+(u-x)_+}}{2\pi\beta x}, \quad (42)$$

where $l = (1-\sqrt{\beta})^2$, $u = (1+\sqrt{\beta})^2$, and $(\cdot)_+ = \max(0,\cdot)$, and $\delta(x)$ is the Dirac delta function. Moreover, even for small values of n_t and n_r, the p.d.f function in (42) can be used to serve as an excellent approximation [23].

Now, let us first consider the case with $\beta \le 1$. In this case, the p.d.f. can be simplified to $f_\lambda^\beta(x) = \sqrt{(x-l)_+(u-x)_+}/2\pi\beta x$. Since all eigenvalues are i.i.d. distributed, we have

$$\begin{aligned}
&\sum_{i=1}^{z_l}\log_2\mathbb{E}_{\lambda_i}[\lambda_i] = z_l\log_2\mathbb{E}[\lambda]\\
&= z_l\log_2\left(\frac{1}{2\pi\beta}\int_l^u\sqrt{-x^2+2(1+\beta)x-(1-\beta)^2}dx\right).
\end{aligned}$$

For convenience, let $R(x) = -x^2+2(1+\beta)x-(1-\beta)^2$. By using [7, Eq. 2.262], we can derive that

$$\int_l^u\sqrt{R(x)}dx = \frac{2x-2(1+\beta)}{4}\sqrt{R(x)}\Big|_l^u + \frac{16\beta}{8}\int_l^u\frac{dx}{\sqrt{R(x)}}. \quad (43)$$

Note that the first term in the summation in (43) is zero. Then by using [7, Eqn. 2.261], we can further derive that

$$\begin{aligned}
\int_l^u\sqrt{R(x)}dx &= 2\beta\arcsin\left(\frac{-2x+2(1+\beta)}{\sqrt{16\beta}}\right)\Big|_l^u\\
&= -2\beta(\arcsin(-1)-\arcsin(1)) = 2\pi\beta.
\end{aligned}$$

It thus follows that

$$\mathbb{E}_{\mathbf{H}_l}[\Delta C_l] \le W\sum_{i=1}^{z_l}\log_2\mathbb{E}_{\lambda_i}[\lambda_i] = Wz_l\log_2\left(\frac{2\pi\beta}{2\pi\beta}\right) = 0.$$

For the case where $\beta > 1$, the first term in the p.d.f. function in (42) becomes relevant. Thus, we need to further evaluate the expectation of the first term. In this case, it is easy to see that

$$\int_l^u x\left(1-\frac{1}{\beta}\right)\delta(x)dx = \left(1-\frac{1}{\beta}\right)\int_l^u x\delta(x)dx = 0. \quad (44)$$

Combining two cases, we finally have $\mathbb{E}_{\mathbf{H}_l}[\Delta C_l] \approx 0$ for all β, and the proof is complete.

APPENDIX C
PERFORMANCE COMPARISON BETWEEN SUCCINCT AND THE H-S MODEL

For the CiM scheme in [8], two sets of integer variables are defined: For every pair of mutually-interfered links i and j in time slot t, $\theta_{ij}(t)$ represents the number of DoFs assigned by $t(i)$ to null its interference at $r(j)$, and $\vartheta_{ij}(t)$ represents the number of DoFs assigned by $r(j)$ to suppress the interference coming from $t(i)$. The H-S link layer mode is given as follows (in this paper's notation):

$$\textstyle\sum_{l\in\mathcal{L}_i^{\mathrm{Out}}} z_l(t) + \sum_{l\in\mathcal{I}_i^+}\theta_{il}(t) \le A_{t(i)}, \quad (45)$$

$$\textstyle\sum_{l\in\mathcal{L}_i^{\mathrm{In}}} z_l(t) + \sum_{l\in\mathcal{I}_i^-}\vartheta_{il}(t) \le A_{r(i)}, \quad (46)$$

$$z_i(t) \le \vartheta_{ij}(t) + \alpha_{t(i)}(1-y_i(t)), \quad (47)$$

$$z_j(t) \le \theta_{ij}(t) + \beta_{r(j)}(1-y_j(t)). \quad (48)$$

Now, let us compare the performance using the 2-link example shown in Fig. 4. Clearly, when both links i and j are active, i.e., $y_i(t) = y_j(t) = 1$, we have from (47) and (48) that $z_i(t) \leq \vartheta_{ij}(t)$ and $z_j(t) \leq \theta_{ij}(t)$. In Section III-A, we can see that $z_i(t) = 3$ and $z_j(t) = 2$ can be scheduled without causing any mutual interference. However, $z_i(t) = 3$ and $z_j(t) = 2$ clearly violates (45)-(48). This is because $z_i(t) = 3 \Rightarrow \theta_{ij} = 0$ (by (45)) and $\theta_{ij} = 0 \Rightarrow z_j(t) = 0$ (by (48)), a contradiction to $z_j(t) = 2$. Thus, we can see that SIM can schedule more data streams than CiM.

In fact, for the simple 2-link example in Fig. 4, we can derive the DoF region under the H-S model as follows. First, note that there is only one interference going from link i to link j. Thus, when both links are active, we can simplify (45)-(48) as: $z_i + \theta_{ij} \leq 3$, $z_j + \vartheta_{ij} \leq 5$, and $z_j \leq \theta_{ij}$. Since $z_j \leq \theta_{ij}$ is tighter than $z_j + \vartheta_{ij} \leq 5$ (because $\theta_{ij} \leq 3$), the latter can be ignored. As a result, we have DoF region as $z_i + z_j \leq z_i + \theta_{ij} \leq 3$. We plot this DoF region (as squares) in Fig. 8. It can be seen that the H-S model can only achieve a portion of the entire DoF region.

To formally state the superior performance of SUCCINCT, we have the following fact.

Fact 1. *The achievable DoF region $\mathcal{C}_{\rm SUCCINCT}$ contains the achievable DoF region $\mathcal{C}_{\rm HS}$ achieved by the H-S model, i.e., $\mathcal{C}_{\rm HS} \subseteq \mathcal{C}_{\rm SUCCINCT}$.*

Proof: Recall that in the SUCCINCT scheme, the number of data streams and the number of DoFs used for interference mitigation satisfy the following constraints:

$$\sum_{l\in\mathcal{L}_i^{\rm Out}} z_l(t) + \sum_{j\in\mathcal{I}_i} \pi_{ji}(t) \sum_{\substack{l:{\rm Rx}(l)=j\\ {\rm Tx}(l)\neq i}} z_l(t) \leq A_i + (1-g_i(t))M, \quad (49)$$

$$\sum_{l\in\mathcal{L}_i^{\rm In}} z_l(t) + \sum_{j\in\mathcal{I}_i} \pi_{ji}(t) \sum_{\substack{l:{\rm Tx}(l)=j\\ {\rm Rx}(l)\neq i}} z_l(t) \leq A_i + (1-h_i(t))M. \quad (50)$$

On the other hand, the link layer constraints for transmissions and interference mitigation in the H-S model can be re-written as:

$$\sum_{l\in\mathcal{L}_i^{\rm Out}} z_l(t) + \sum_{j\in\mathcal{I}_i} \sum_{\substack{l:{\rm Rx}(l)=j\\ {\rm Tx}(l)\neq i}} z_l(t) \leq A_i + (1-g_i(t))M, \quad (51)$$

$$\sum_{l\in\mathcal{L}_i^{\rm Out}} z_l(t) + \sum_{j\in\mathcal{I}_i} \sum_{\substack{l:{\rm Rx}(l)=j\\ {\rm Tx}(l)\neq i}} z_l(t) \leq A_i + (1-g_i(t))M. \quad (52)$$

It can see that the RHS are identical. However, in the SUCCINCT scheme, the number of terms in the summation for interference mitigation for each node is strictly less than that in the H-S model except for the last node to be scheduled in the SUCCINCT scheme. This is because for each node in the network, it only has to perform interference mitigation for the nodes that are scheduled earlier. Mathematically, this can be reflected by the fact that only some of γ-variables in the summation are equal to one.

By contrast, in the H-S model, each node has to perform interference mitigation task for *all* nodes within its interference range. As a result, the number of terms in the summation for interference mitigation is strictly larger than that in the SUCCINCT scheme. Mathematically, this is equivalent to setting all π-variables equal to 1 in (49) and (50).

Since more DoF resources are used for interference mitigation in the H-S model, we have that the number of DoF used for data transmissions is strictly less than that in the SUCCINCT scheme, thus resulting in a strictly larger DoFs region. ■

APPENDIX D
PROOF OF LEMMA 2

For simplicity, we drop the time slot index t of $\pi-$ variables in this proof. We first show that given any one group in (11)–(16), the remaining 5 are redundant. Without loss generality, let us consider the two constraints in group (11), i.e.,

$$\begin{cases} \pi_{ik} \leq \pi_{ij} + \pi_{jk} \\ \pi_{ik} \geq \pi_{ij} + \pi_{jk} - 1. \end{cases}$$

Multiply -1 and adding 1 on both sides of the above two inequalities, we have

$$\begin{cases} 1-\pi_{ik} \geq 1-\pi_{ij}-\pi_{jk} \\ 1-\pi_{ik} \leq 1-\pi_{ij}-\pi_{jk}+1 \end{cases} \Rightarrow \begin{cases} \pi_{ki} \geq \pi_{ji} + \pi_{kj} - 1 \\ \pi_{ki} \leq \pi_{ji} + \pi_{kj}. \end{cases}$$

This shows that the set of constraints in group (12) can be derived from group (11) and the mutual exclusiveness constraints. Thus, group (12) is not independent and can be removed. Following the same token, we can also show that

$$\begin{cases} \pi_{ij} \leq \pi_{ik} + \pi_{kj} \\ \pi_{ij} \geq \pi_{ik} + \pi_{kj} - 1 \end{cases} \Rightarrow \begin{cases} \pi_{ji} \geq \pi_{jk} + \pi_{ki} - 1 \\ \pi_{ji} \leq \pi_{jk} + \pi_{ki}, \end{cases}$$

and

$$\begin{cases} \pi_{jk} \leq \pi_{ji} + \pi_{ik} \\ \pi_{jk} \geq \pi_{ji} + \pi_{ik} - 1 \end{cases} \Rightarrow \begin{cases} \pi_{kj} \geq \pi_{ki} + \pi_{ij} - 1 \\ \pi_{kj} \leq \pi_{ki} + \pi_{ij}. \end{cases}$$

That is, the groups (14) and (16) can derived from groups (13) and (15). So, they can also be removed. Now, let us consider the remaining 6 constraints associated with groups (11), (13), and (15). Note that

$$\begin{cases} \pi_{ik} \leq \pi_{ij} + \pi_{jk} & \Rightarrow & \pi_{ik} \leq \pi_{ij} + 1 - \pi_{kj} \\ & \Rightarrow & \pi_{ij} \geq \pi_{ik} + \pi_{kj} - 1 \\ \pi_{ik} \geq \pi_{ij} + \pi_{jk} - 1 & \Rightarrow & \pi_{ik} \geq \pi_{ij} - \pi_{kj} \\ & \Rightarrow & \pi_{ij} \leq \pi_{ik} + \pi_{kj}. \end{cases}$$

This shows that group (13) can be derived from group (11) and can also removed. Following exactly the same token, we can show that group (15) can be derived from group (11) and also can be removed. So finally, we have only 2 remaining constraints associated with group (11).

Next, we show that (11) can be written in a more compact form. Note that

$$\begin{cases} \pi_{ik} \leq \pi_{ij} + \pi_{jk} & \Rightarrow & 1-\pi_{ki} \leq \pi_{ij} + \pi_{jk} \\ & \Rightarrow & \pi_{ij} + \pi_{jk} + \pi_{ki} \geq 1 \\ \pi_{ik} \geq \pi_{ij} + \pi_{jk} - 1 & \Rightarrow & 1-\pi_{ki} \geq \pi_{ik} + \pi_{jk} - 1 \\ & \Rightarrow & \pi_{ij} + \pi_{jk} + \pi_{ki} \leq 2, \end{cases}$$

which give the constraints stated in Lemma 2. This completes the proof.