Optimal Power Allocation in Multi-Relay MIMO Cooperative Networks: Theory and Algorithms

Jia Liu, *Member, IEEE,* Ness B. Shroff, *Fellow, IEEE,* and Hanif D. Sherali

***Abstract*—Cooperative networking is known to have significant potential in increasing network capacity and transmission reliability. Although there have been extensive studies on applying cooperative networking in multi-hop ad hoc networks, most works are limited to the basic three-node relay scheme and single-antenna systems. These two limitations are interconnected and both are due to a limited theoretical understanding of the optimal power allocation structure in MIMO cooperative networks (MIMO-CN). In this paper, we study the structural properties of the optimal power allocation in MIMO-CN with per-node power constraints. More specifically, we show that the optimal power allocations at the source and each relay follow a matching structure in MIMO-CN. This result *generalizes* the power allocation result under the basic three-node setting to the multi-relay setting, for which the optimal power allocation structure has been heretofore unknown. We further quantify the performance gain due to cooperative relay and establish a connection between cooperative relay and pure relay. Finally, based on these structural insights, we reduce the MIMO-CN rate maximization problem to an equivalent scalar formulation. We then propose a global optimization method to solve this simplified and equivalent problem.**

***Index Terms*—Cooperative networks, multiple-input multiple-output (MIMO), relay, amplify and forward, power allocation, nonconvex optimization, algorithms.**

I. Introduction

THE CONCEPT of cooperative networking [1], [2] traces its roots back to the 1970s, when information-theoretic studies were first conducted in [3], [4] under the theme of "relay channels." In recent years, cooperative networking has received substantial interest from the wireless networking and communications research communities. Many interesting problems for cooperative networks have been actively researched, such as throughput-optimal scheduling [5], network lifetime maximization [6], distributed routing [7], and MAC layer protocol design [8], just to name a few.

Although there have been extensive studies concerning cooperative networks, most works on optimizing the performance of cooperative networks are limited by: **i) the basic three-node relay scheme.** The basic three-node relay scheme is shown in Fig. 1, where the message transmitted from source S to destination D is relayed by a node R, which can overhear the message. In an ad hoc network environment, however, the message from the source is likely to be overheard by multiple neighboring nodes. A common cooperative approach in this situation is *relay assignment*, i.e., we choose only one of the neighboring nodes as a relay for which the three-node relay scheme can be applied (see, e.g., [9], [10] and references therein). Despite its simplicity, this relay assignment scheme is clearly suboptimal since all such neighboring nodes can potentially serve as relays to further improve the system performance, as shown in Fig. 2; **ii) single-antenna systems.** In the current literature, research on cooperative networks with MIMO-enabled nodes remains limited. In cooperative networks, it is interesting to explore the idea of deploying multiple antennas at each node. With multiple antennas, the source and the relays can multiplex independent data streams by exploiting the inherent independent spatial channels.

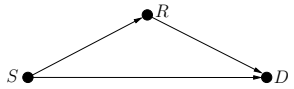

Fig. 1. The basic three-node relay scheme.

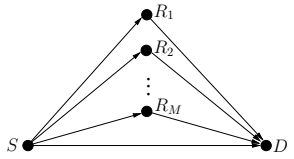

Fig. 2. A cooperative network with multiple relays.

While the above two limitations are seemingly unrelated, they are in fact both associated with the limited theoretical understanding of MIMO cooperative networks (MIMO-CN). To see this, let us consider the single-antenna multi-relay network in Fig. 2. Here, we can treat all single-antenna relays $R_1, \ldots, R_M$ as a single virtual relay node with M antennas. In this sense, analyzing this multi-relay network is closely related to analyzing a three-node cooperative network with a MIMO-enabled relay. Thus, besides the attractive capacity benefits, research results of MIMO-CN can also generalize previous studies on single-antenna-based cooperative communication, which can be viewed as special cases of MIMO-CN.

In this paper, we consider the optimal power allocation at the source and each relay to maximize the end-to-end achievable rate of multi-relay MIMO-CN. Our focus is on the amplify-and-forward (AF) relay strategy. An obvious reason is that AF has low complexity since no decoding/encoding is needed. This benefit is even more attractive in MIMO-

Manuscript received 15 February 2011; revised 20 July 2011. This work has been partially funded via NSF Grant CNS-1012700, a grant from the Army Research Office MURI W911NF08-1-0238, and NSF Grant CMMI-0969196.

J. Liu and N. B. Shroff are with the Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH, 43210 USA (e-mail: {liu, shroff}@ece.osu.edu).

H. D. Sherali is with the Grado Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg, VA, 24061 USA (e-mail: hanifs@vt.edu).

Digital Object Identifier 10.1109/JSAC.2012.1202xx.

CN, where decoding multiple data streams could be computationally intensive. In addition to simplicity, a more important reason is that AF outperforms decode-and-forward (DF) in terms of network capacity scaling: in general, as the number of relays increases in MIMO-CN, the effective signal-to-noise ratio (SNR) under AF scales linearly, as opposed to being a constant under DF (see [11] for more detailed discussions). The main results and contribution of this work are as follows:

- We show that the optimal power allocation at each relay follows a *matching* structure: the diagonalization of each relay's amplification matrix matches with certain eigendirections of the joint source-relay and relay-destination channels. Our result *generalizes* the matching structure under the basic three-node setting to the multi-relay setting, for which the optimal power allocation structure has been heretofore unknown. We remark that our finding is not a straightforward extension of the "matching structure" in the three-node relay setting [12]–[15] due to the more complex per-node power constraint. As a result, showing the matching structure in our case requires a new proof.
- By analyzing the channel structures in MIMO-CN, we establish the relationship between MIMO-CN and multi-hop MIMO networks with pure relay links. We specifically quantify the performance gain due to cooperative relay. This result is novel because: 1) it simplifies the proof of the optimal power allocation structure under CR; and 2) it is the first result that connects cooperative relay and pure relay in a multi-relay setting, which further advances our understanding of the benefits of cooperative communications.
- Based on the above structural insights, we reduce the MIMO-CN rate maximization problem to an equivalent scalar formulation. We analyze the structure and convexity properties of the equivalent problem, and then propose a *global* optimization algorithm based on a branch-and-bound framework coupled with a custom-designed convex programming relaxation (BB/CPR), which *guarantees* finding a global optimal solution for this *nonconvex* problem. To our knowledge, this is the first work that employs a global optimization technique for MIMO-CN.

The remainder of this paper is organized as follows. Section II discusses the related work. Section III presents the network model and problem formulation. In Sections IV and V, we study the optimal power allocation structures in pure relay and cooperative relay paradigms, and point out their connections. Based on these structural results, we simplify and reformulate the optimization problems. In Section VI, we propose a novel global optimization method called BB/CPR to solve the reformulated problem. Numerical results are provided to show the efficacy of the proposed algorithms. Section VII concludes this paper.

II. Related Work

Since the benefits of cooperative networking were recognized [1], [2], several initial attempts on extending cooperative networking to MIMO have been reported [12]–[17]. In [12], Tang and Hua first considered the optimal relay amplification matrix for the basic three-node MIMO-CN under the assumption that the source-relay channel state information (CSI) is unknown. Their main conclusion is that when the direct link between the source and the destination is not present (i.e., pure relay), the optimal amplification matrix adopts a "matching" structure. Coincidentally, Muñoz-Medina *et al*. [13] independently arrived at the same conclusion via a different proof technique. Later in [18], Fang *et al*. generalized the matching result to the three-node MIMO-CN network where the source has full CSI.[1] Recent works on MIMO-CN started to consider more complex relay settings. In [19], Fu *et al*. studied MIMO-CN with multiple AF relays, which is similar to our setting. However, their work differs from ours in that they assumed a sum power constraint across all relay nodes, which is usually not realistic since each relay has its own power budget. Thus, a per-node power constraint on each relay is more appropriate. As we shall see later, imposing a per-node power constraint results in a more challenging power allocation problem. It is worth pointing out that the three-node multi-carrier MIMO-CN considered in [16] can also be viewed as a MIMO-CN with multiple relays. Compared to our network setting, the major difference is that each source-relay-destination path in [20] operates under orthogonal channels (subcarriers) that do not cooperate with each other. This yields a more tractable problem, which can be thought of as a special case of the model we consider in this paper.

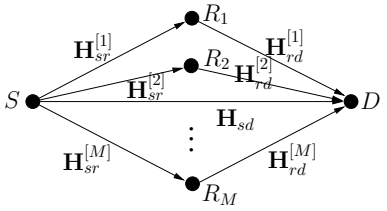

Fig. 3. An AF-based MIMO-CN with M relay nodes.

III. Network Model and Problem Formulation

In this paper, we use boldface to denote matrices and vectors. For a complex-valued matrix $\mathbf{M}$, we let $\mathbf{M}^*$, $\mathbf{M}^\dagger$, and $|\mathbf{M}|$ denote the conjugate, conjugate transpose, and determinant of $\mathbf{M}$, respectively. $\mathrm{Tr}\{\mathbf{M}\}$ denotes the trace of $\mathbf{M}$. We let $\mathbf{I}_N$ denote the $N \times N$ identity matrix. $\mathbf{M} \succeq 0$ represents that $\mathbf{M}$ is Hermitian and positive semidefinite (PSD). $\mathrm{Diag}\{\mathbf{M}_1 \dots \mathbf{M}_n\}$ represents the block diagonal matrix with matrices $\mathbf{M}_1, \dots, \mathbf{M}_n$ on its main diagonal. $\mathbf{M} \circ \mathbf{N}$ represents the Hadamard product of matrices $\mathbf{M}$ and $\mathbf{N}$. We let $(\mathbf{M})_{ij}$ denote the entry in the i-th row and j-th column in matrix $\mathbf{M}$. We let $(\mathbf{v})_j$ denote the j-th entry of the vector $\mathbf{v}$.

We consider an AF-based MIMO-CN as shown in Fig. 3. The source node S transmits messages to the destination node D, assisted by M relays R_1, R_2, ..., R_M. S and D have N_s and N_d antennas, respectively. For ease of exposition, we assume that each R_i has N_r antennas. We note that the generalization to the case where each R_i has a distinct number of antennas can also be similarly incorporated in our subsequent analysis but at the expense of more complicated notation. We let $\mathbf{H}_{sr}^{[i]} \in \mathbb{C}^{N_r \times N_s}$ and $\mathbf{H}_{rd}^{[i]} \in \mathbb{C}^{N_d \times N_r}$ denote

[1]Interestingly, the actual publication dates of [12], [13] turned out to be even later than that of [18] due to journal appearance delay.

the channel gain matrices between S and R_i and between R_i and D, respectively. The channel gain matrix for the direct link between S and D is denoted by $\mathbf{H}_{sd} \in \mathbb{C}^{N_d \times N_s}$.

Due to self-interference, each relay node cannot transmit and receive in the same channel at the same time. Thus, a transmission from S to D takes two time slots. In the first time slot, a message is transmitted from S to D, which is overheard by all R_i. In the second time slot, each R_i simply amplifies and transmits its received signal to D without decoding the message. At the end of the second time slot, D coherently combines all received signals to decode the message. In the rest of the paper, we classify MIMO-CN into two cases depending on whether or not the direct channel between S and D is strong enough to support communication. We refer to the case where the direct channel is absent as *pure relay* (PR) and refer to the opposite case as *cooperative relay* (CR). Although PR is a special case of CR, this categorization proves to be useful in that: 1) it is easier to obtain structural properties under the simpler PR case and these properties provide important insights to the more complex CR case; and 2) this categorization helps build a connection between PR and CR and deepens our understanding of cooperation benefits.

In a MIMO-CN, the received signal at the i-th relay, denoted as $\mathbf{y}_r^{[i]}$, can be written as $\mathbf{y}_r^{[i]} = \mathbf{H}_{sr}^{[i]}\mathbf{x}_s + \mathbf{n}_r^{[i]}$, $i = 1, 2, \ldots, M$, where $\mathbf{x}_s \in \mathbb{C}^{N_s}$ represents the transmission signal vector and $\mathbf{n}_r^{[i]} \in \mathbb{C}^{N_r}$ represents the zero-mean circularly symmetric Gaussian noise vector seen at R_i. We let $\mathbf{H}_{sr} \triangleq \left[\mathbf{H}_{sr}^{[1]\dagger}, \ldots, \mathbf{H}_{sr}^{[M]\dagger}\right]^\dagger \in \mathbb{C}^{MN_r \times N_s}$ represent the combined relay channel matrix between S and all R_i. Also, we let $\mathbf{n}_r \triangleq [\mathbf{n}_r^{[1]\dagger}, \ldots, \mathbf{n}_r^{[M]\dagger}]^\dagger \in \mathbb{C}^{MN_r}$. Then, the combined received signal of all R_i, defined as $\mathbf{y}_r \triangleq [\mathbf{y}_r^{[1]\dagger}, \ldots, \mathbf{y}_r^{[M]\dagger}]^\dagger \in \mathbb{C}^{MN_r}$, can be compactly written as $\mathbf{y}_r = \mathbf{H}_{sr}\mathbf{x}_s + \mathbf{n}_r$.

Recall that in a multi-relay AF-based MIMO-CN, each R_i amplifies the received signal. The amplification factor at R_i can be represented by a matrix $\mathbf{A}_i \in \mathbb{C}^{N_r \times N_r}$. Thus, the relay signal at R_i can be written as $\mathbf{x}_r^{[i]} = \mathbf{A}_i\mathbf{y}_r^{[i]}$. Since the structure of each $\mathbf{A}_i$ could significantly impact the performance of an AF-based MIMO-CN, one of the main goals in this paper is to understand the optimal structural property of each $\mathbf{A}_i$. Due to the *distributed nature* of ad hoc network environments, the relay nodes cannot share their received signals with each other. This can be mathematically modeled by introducing a *block diagonal constraint* on the overall amplification matrix $\mathbf{A}$ for all the relays, i.e., $\mathbf{A} = \text{Diag}\{\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_M\} \in \mathbb{C}^{MN_r \times MN_r}$. Using matrix $\mathbf{A}$, we can represent the overall relay signal as $\mathbf{x}_r = \mathbf{A}\mathbf{y}_r = \mathbf{A}(\mathbf{H}_{sr}\mathbf{x}_s + \mathbf{n}_r)$.

We use $\mathbf{H}_{rd} \triangleq \left[\mathbf{H}_{rd}^{[1]}, \ldots, \mathbf{H}_{rd}^{[M]}\right] \in \mathbb{C}^{N_d \times MN_r}$ to represent the combined channel gain matrix between all R_i and D. Under PR, the received signal at D can be written as:

$$\mathbf{y}_d = \mathbf{H}_{rd}\mathbf{x}_r + \mathbf{n}_d^{(2)} = \mathbf{H}_{rd}\mathbf{A}\mathbf{H}_{sr}\mathbf{x}_s + (\mathbf{H}_{rd}\mathbf{A}\mathbf{n}_r + \mathbf{n}_d^{(2)}), \quad (1)$$

where $\mathbf{n}_d^{(2)} \in \mathbb{C}^{N_d}$ is the zero-mean circularly symmetric Gaussian noise vector seen at D in the second time slot. Under CR, we can expand the received signal $\mathbf{y}_d$ in (1) as follows:

$$\mathbf{y}_d = \begin{bmatrix} \mathbf{H}_{rd}\mathbf{A}\mathbf{H}_{sr} \\ \mathbf{H}_{sd} \end{bmatrix}\mathbf{x}_s + \begin{bmatrix} \mathbf{H}_{rd}\mathbf{A} & \mathbf{I}_{N_d} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{N_d} \end{bmatrix}\begin{bmatrix} \mathbf{n}_r \\ \mathbf{n}_d^{(2)} \\ \mathbf{n}_d^{(1)} \end{bmatrix}, \quad (2)$$

where $\mathbf{n}_d^{(1)} \in \mathbb{C}^{N_d}$ denotes the zero-mean circularly symmetric Gaussian noise vector at D in the first time slot.

Under PR, the end-to-end achievable rate can be computed as [12], [13]:

$$I_{PR}(\mathbf{Q}, \mathbf{A}) = \frac{1}{2}\log_2 \left|\mathbf{I}_{N_d} + (\mathbf{H}_{rd}\mathbf{A}\mathbf{H}_{sr})\mathbf{Q}(\mathbf{H}_{rd}\mathbf{A}\mathbf{H}_{sr})^\dagger \times (\sigma_d^2\mathbf{I}_{N_d} + \sigma_r^2\mathbf{H}_{rd}\mathbf{A}\mathbf{A}^\dagger\mathbf{H}_{rd})^{-1}\right|, \quad (3)$$

where $\mathbf{Q} \triangleq \mathbb{E}\{\mathbf{x}\mathbf{x}^\dagger\}$ represents the input signal covariance matrix (i.e., the source power); and σ_r^2 and σ_d^2 denote the variances of $\mathbf{n}_r$ and $\mathbf{n}_d^{(2)}$, respectively. The factor $\frac{1}{2}$ in (3) accounts for the two time slots required to complete a transmission. By letting $\bar{\mathbf{H}}_{sd} \triangleq \mathbf{H}_{rd}\mathbf{A}\mathbf{H}_{sr}$ denote the equivalent end-to-end channel under PR and letting $\bar{\mathbf{R}} \triangleq \sigma_d^2\mathbf{I}_{N_d} + \sigma_r^2\mathbf{H}_{rd}\mathbf{A}\mathbf{A}^\dagger\mathbf{H}_{rd}^\dagger$ denote the equivalent noise power at D, we can compactly rewrite (3) as

$$I_{PR}(\mathbf{Q}, \mathbf{A}) = \frac{1}{2}\log_2 \left|\mathbf{I}_{N_d} + \bar{\mathbf{H}}_{sd}\mathbf{Q}\bar{\mathbf{H}}_{sd}^\dagger\bar{\mathbf{R}}^{-1}\right|. \quad (4)$$

Similarly, we can write the end-to-end achievable rate under CR as:

$$I_{CR}(\mathbf{Q}, \mathbf{A}) = \frac{1}{2}\log_2 \left|\mathbf{I}_{2N_d} + \mathbf{H}\mathbf{Q}\mathbf{H}^\dagger\mathbf{R}^{-1}\right|. \quad (5)$$

In (5), the equivalent end-to-end channel gain matrix $\mathbf{H}$ and noise power $\mathbf{R}$ are defined as

$$\mathbf{H} \triangleq \begin{bmatrix} \bar{\mathbf{H}}_{sd} \\ \mathbf{H}_{sd} \end{bmatrix} \quad \text{and} \quad \mathbf{R} \triangleq \begin{bmatrix} \bar{\mathbf{R}} & \mathbf{0} \\ \mathbf{0} & \sigma_d^2\mathbf{I}_{N_d} \end{bmatrix}, \quad (6)$$

respectively. Due to the maximum transmission power limit at S and each R_i, we have the following power constraints: $\text{Tr}(\mathbf{Q}) \le P_T$, $\text{Tr}(\mathbf{A}_i(\sigma_r^2\mathbf{I}_{N_r} + \mathbf{H}_{sr}^{[i]}\mathbf{Q}(\mathbf{H}_{sr}^{[i]})^\dagger)\mathbf{A}_i^\dagger) \le P_R$, $i = 1, \ldots, M$, where P_T and P_R represent the maximum transmission power of S and R_i, respectively. For convenience, we define two constraint sets as follows:

$$\Omega \triangleq \{\mathbf{Q} \,|\text{Tr}(\mathbf{Q}) \le P_T\},$$

$$\Psi \triangleq \left\{ \text{Diag}\{\mathbf{A}_1, \ldots, \mathbf{A}_M\} \middle| \begin{array}{l} \text{Tr}(\mathbf{A}_i(\sigma_{sr}^2\mathbf{I}_{N_r} + \mathbf{H}_{sr}^{[i]}\mathbf{Q}(\mathbf{H}_{sr}^{[i]})^\dagger) \\ \mathbf{A}_i^\dagger) \le P_R, \quad \forall i = 1, 2, \ldots, K \end{array} \right\}.$$

In this paper, our goal is to find optimal $\mathbf{Q}$ and $\mathbf{A}$ to maximize $I_{PR}(\mathbf{Q}, \mathbf{A})$ under PR or $I_{CR}(\mathbf{Q}, \mathbf{A})$ under CR. This can be formulated as the following joint source-relay power optimization (PO) problems under PR and CR, respectively:

$$\underline{\textbf{PO-PR}}\textbf{:} \begin{array}{ll} \text{Maximize} & I_{PR}(\mathbf{Q}, \mathbf{A}) \\ \text{subject to} & \mathbf{Q} \in \Omega, \mathbf{A} \in \Psi, \end{array} \quad (7)$$

$$\underline{\textbf{PO-CR}}\textbf{:} \begin{array}{ll} \text{Maximize} & I_{CR}(\mathbf{Q}, \mathbf{A}) \\ \text{subject to} & \mathbf{Q} \in \Omega, \mathbf{A} \in \Psi. \end{array} \quad (8)$$

It is easy to see that both PO-PR and PO-CR can be naturally *decomposed* into two parts as follows:

$$\max_{\mathbf{Q} \in \Omega}\left(\max_{\mathbf{A} \in \Psi} I_j(\mathbf{Q}, \mathbf{A})\right), \quad (9)$$

where $j \in \{PR, CR\}$. Hence, solving PO-PR and PO-CR reduces to *iteratively* solving an inner subproblem with respect to $\mathbf{A}$ (with $\mathbf{Q}$ fixed) and an outer main program with respect to $\mathbf{Q}$ (with $\mathbf{A}$ fixed). Due to the complex matrix variables and operations, directly tackling such problems is intractable in general. For PO-PR and PO-CR, however, it turns out that we can exploit the inherent special structure to significantly reduce the problem complexity. In what follows, we will first

study the optimal structural properties of $\mathbf{A}$ and $\mathbf{Q}$. Based on these properties, we will reformulate PO-PR and PO-CR. We will start with the relatively simpler PO-PR problem. It will soon be clear that the results under PR provide useful insights for the more complex problem PO-CR.

IV. Optimal Power Allocation Structure: The Pure Relay Case

In Section IV-A, we will investigate the structural properties of an optimal $\mathbf{A}$ for a given $\mathbf{Q}$ (the inner problem in (9)). Then, we will study the structural properties of an optimal $\mathbf{Q}$ in Section IV-B (the outer problem in (9)). Based on these results, we will reformulate Problem PO-PR in Section IV-C.

A. The Structure of an Optimal Amplification Matrix

For now, we assume that $\mathbf{Q}$ is given. We let $\tilde{\mathbf{H}}_{rd} \triangleq \frac{\sigma_r}{\sigma_d}\mathbf{H}_{rd}\mathbf{A}$ and $\tilde{\mathbf{H}}_{sr} \triangleq \frac{1}{\sigma_r}\mathbf{H}_{sr}\mathbf{Q}^{\frac{1}{2}}$. We let $\tilde{\mathbf{\Lambda}}_{sr} \in \mathbb{C}^{MN_r \times MN_r}$ and $\tilde{\mathbf{U}}_{sr} \in \mathbb{C}^{MN_r \times MN_r}$ be the real diagonal matrix and the orthonormal matrix obtained from the eigenvalue decomposition of $\tilde{\mathbf{H}}_{sr}\tilde{\mathbf{H}}_{sr}^{\dagger}$, i.e., $\tilde{\mathbf{H}}_{sr}\tilde{\mathbf{H}}_{sr}^{\dagger} = \tilde{\mathbf{U}}_{sr}\tilde{\mathbf{\Lambda}}_{sr}\tilde{\mathbf{U}}_{sr}^{\dagger}$. To study the structural properties of an optimal $\mathbf{A}$, we first need the following result to simplify $I_{PR}(\mathbf{Q}, \mathbf{A})$:

Lemma 1. *The achievable rate expression in (3) is equivalent to the following expression:*

$$I_{PR}(\mathbf{Q}, \mathbf{A}) = \frac{1}{2}\log_2 \Big|\mathbf{I}_{MN_r} + \tilde{\mathbf{\Lambda}}_{sr}(\mathbf{I}_{MN_r} - (\mathbf{I}_{MN_r} + \tilde{\mathbf{U}}_{sr}^{\dagger}\tilde{\mathbf{H}}_{rd}^{\dagger}\tilde{\mathbf{H}}_{rd}\tilde{\mathbf{U}}_{sr})^{-1})\Big|. \quad (10)$$

We note that Lemma 1 was also used in [19] but without proof. Since this result is non-obvious, we provide a proof of Lemma 1 in [11] to make this paper self-contained. Lemma 1 implies that maximizing (3) over $\mathbf{A}$ can be equivalently done by maximizing (10), which is relatively easier due to its *diagonal* structure. According to the Hadamard inequality [21], the right-hand-side (RHS) of (10) is maximized when the matrix inside the determinant is diagonal. Note that every matrix term in the determinant is already diagonal except for $(\mathbf{I}_{MN_r} + \tilde{\mathbf{U}}_{sr}^{\dagger}\tilde{\mathbf{H}}_{rd}^{\dagger}\tilde{\mathbf{H}}_{rd}\tilde{\mathbf{U}}_{sr})^{-1}$. Hence, for the matrix inside the determinant to be diagonal, it suffices that $\tilde{\mathbf{U}}_{sr}^{\dagger}\tilde{\mathbf{H}}_{rd}^{\dagger}\tilde{\mathbf{H}}_{rd}\tilde{\mathbf{U}}_{sr}$ is diagonal. This leads to the following result:

Lemma 2. *Let $\mathbf{V}_{rd}$ be the eigenvector matrix for $\mathbf{H}_{rd}^{\dagger}\mathbf{H}_{rd}$ i.e., $\mathbf{H}_{rd}^{\dagger}\mathbf{H}_{rd} = \mathbf{V}_{rd}\mathbf{\Lambda}_{rd}\mathbf{V}_{rd}^{\dagger}$. If $\mathbf{V}_{rd}^{\dagger}\mathbf{A}\tilde{\mathbf{U}}_{sr}$ is diagonal, then $\tilde{\mathbf{U}}_{sr}^{\dagger}\tilde{\mathbf{H}}_{rd}^{\dagger}\tilde{\mathbf{H}}_{rd}\tilde{\mathbf{U}}_{sr}$ is diagonal.*

Proof: We rewrite $\tilde{\mathbf{U}}_{sr}^{\dagger}\tilde{\mathbf{H}}_{rd}^{\dagger}\tilde{\mathbf{H}}_{rd}\tilde{\mathbf{U}}_{sr}$ in the following form: $\tilde{\mathbf{U}}_{sr}^{\dagger}\tilde{\mathbf{H}}_{rd}^{\dagger}\tilde{\mathbf{H}}_{rd}\tilde{\mathbf{U}}_{sr} = (\sigma_r^2/\sigma_d^2)\tilde{\mathbf{U}}_{sr}^{\dagger}\mathbf{A}^{\dagger}\mathbf{H}_{rd}^{\dagger}\mathbf{H}_{rd}\mathbf{A}\tilde{\mathbf{U}}_{sr} = (\sigma_r^2/\sigma_d^2)\tilde{\mathbf{U}}_{sr}^{\dagger}\mathbf{A}^{\dagger}\mathbf{V}_{rd}\mathbf{\Lambda}_{rd}\mathbf{V}_{rd}^{\dagger}\mathbf{A}\tilde{\mathbf{U}}_{sr}$. Since $\mathbf{\Lambda}_{rd}$ is diagonal, in order for $\tilde{\mathbf{U}}_{sr}^{\dagger}\tilde{\mathbf{H}}_{rd}^{\dagger}\tilde{\mathbf{H}}_{rd}\tilde{\mathbf{U}}_{sr}$ to be diagonal, it suffices for $\mathbf{V}_{rd}^{\dagger}\mathbf{A}\tilde{\mathbf{U}}_{sr}$ to be diagonal. ∎

Based on Lemma 2, we now analyze the diagonality of $\mathbf{V}_{rd}^{\dagger}\mathbf{A}\tilde{\mathbf{U}}_{sr}$. Recalling that $\mathbf{A} = \mathrm{Diag}\{\mathbf{A}_1, \ldots, \mathbf{A}_M\}$, we have that

$$\mathbf{V}_{rd}^{\dagger}\mathbf{A}\tilde{\mathbf{U}}_{sr} = \sum_{k=1}^{M}(\mathbf{V}_{rd}^{[k]})^{\dagger}\mathbf{A}_k\tilde{\mathbf{U}}_{sr}^{[k]}, \quad (11)$$

where $\mathbf{V}_{rd}^{[k]} \in \mathbb{C}^{N_r \times MN_r}$ and $\tilde{\mathbf{U}}_{sr}^{[k]} \in \mathbb{C}^{N_r \times MN_r}$ represent the submatrices in $\mathbf{V}_{rd}$ and $\tilde{\mathbf{U}}_{sr}$ starting from the $((k-1)N_r+1)$-st row to the kN_r-th row, respectively. From (11), we obtain the following key lemma:

Lemma 3. *$\mathbf{V}_{rd}^{\dagger}\mathbf{A}\tilde{\mathbf{U}}_{sr}$ is diagonal if and only if $(\mathbf{V}_{rd}^{[k]})^{\dagger}\mathbf{A}_k\tilde{\mathbf{U}}_{sr}^{[k]}$ is diagonal for all $k = 1, \ldots, M$.*

Proof: The "if" part follows immediately from (11). The "only if" part can be proved by first assuming that not all $(\mathbf{V}_{rd}^{[k]})^{\dagger}\mathbf{A}_k\tilde{\mathbf{U}}_{sr}^{[k]}$ are diagonal. Then, by analyzing a homogeneous linear equation system associated with $\mathbf{V}_{rd}^{\dagger}\mathbf{A}\tilde{\mathbf{U}}_{sr}$, we can reach a contradiction. We relegate the proof details of Lemma 3 to [11]. ∎

Using Lemma 3, we have the first main result in this paper.

Theorem 1. *For a given $\mathbf{Q}$, the optimal relay amplification matrix $\mathbf{A}_k$ at R_k has the following structure:*

$$\mathbf{A}_k = ((\mathbf{V}_{rd}^{[k]})^{\dagger})^{\mathrm{LI}}\mathbf{\Phi}_k(\tilde{\mathbf{U}}_{sr}^{[k]})^{\mathrm{RI}}, \quad k = 1, 2, \ldots, M, \quad (12)$$

where $(\cdot)^{\mathrm{LI}}$ and $(\cdot)^{\mathrm{RI}}$ represent the left and right inverses, respectively; and $\mathbf{\Phi}_k \in \mathbb{R}^{MN_r \times MN_r}$ is a real diagonal matrix.

Proof: From Lemma 3, we have that $(\mathbf{V}_{rd}^{[k]})^{\dagger}\mathbf{A}_k\tilde{\mathbf{U}}_{sr}^{[k]}$ is diagonal for all k. Thus, we let $(\mathbf{V}_{rd}^{[k]})^{\dagger}\mathbf{A}_k\tilde{\mathbf{U}}_{sr}^{[k]} = \mathbf{\Phi}_k$, where $\mathbf{\Phi}_k$ is diagonal as stated in the theorem. Since $(\mathbf{V}_{rd}^{[k]})^{\dagger}$ and $\tilde{\mathbf{U}}_{sr}^{[k]}$ are tall-skinny and short-fat, there exist left and right inverses for $(\mathbf{V}_{rd}^{[k]})^{\dagger}$ and $\tilde{\mathbf{U}}_{sr}^{[k]}$, respectively. Then, Eq.(12) follows from multiplying $((\mathbf{V}_{rd}^{[k]})^{\dagger})^{\mathrm{LI}}$ and $(\tilde{\mathbf{U}}_{sr}^{[k]})^{\mathrm{RI}}$ on the left- and right- hand sides of $(\mathbf{V}_{rd}^{[k]})^{\dagger}\mathbf{A}_k\tilde{\mathbf{U}}_{sr}^{[k]} = \mathbf{\Phi}_k$, respectively. ∎

Remark 1. We can see from (12) that $\mathbf{A}_k$ contains two parts: $(\tilde{\mathbf{U}}_{sr}^{[k]})^{\mathrm{RI}}$ matches with the eigen-directions of the joint source-relay channel, and $(\mathbf{V}_{rd}^{[k]})^{\mathrm{LI}}$ matches with the eigen-directions of the joint relay-destination channel. With such direction matchings, the power allocation is solely determined by the diagonal entries of $\mathbf{\Phi}_k$. We point out that by imposing per-node power constraints, the matching direction of $\mathbf{A}_k$ is completely *different* from that in [19] (cf. [19, Eq. (12)]), where all relays are subject to a sum power constraint.

Remark 2. Theorem 1 implies that each R_i needs the knowledge of the joint channels $\mathbf{H}_{sr}$ and $\mathbf{H}_{rd}$. This means that, although the relays do not share received signals, they do need to share CSI, which is the price to pay in order to achieve better performance. This CSI requirement could potentially impose some challenges on practical implementations, and the so-called limited CSI feedback techniques could be desirable to employ in practice (see [22] and references therein). Another remedy to this CSI issue is to let each R_i send its respective observed channel gain $\mathbf{H}_{sr}^{[i]}$ to D. Using this information and the channel $\mathbf{H}_{rd}$ observed by D itself, D can compute optimal amplification matrices for each R_i and then feed back the solutions to each R_i. Under this approach, the large amount of CSI sharing among the relays can be avoided.

B. The Structure of Optimal Source Covariance Matrix

As mentioned earlier, under an appropriate direction matching, the value of the achievable rate depends on the diagonal

entries in $\tilde{\mathbf{\Lambda}}_{sr}$. Hence, following a similar argument as in [17], [18], it can be shown that an optimal $\mathbf{Q}$ should match with the dominant right singular matrix of $\mathbf{H}_{sr}$. We state the result in the following proposition and omit its proof in this paper.

Proposition 1. *An optimal* $\mathbf{Q}$ *can be decomposed as* $\mathbf{Q} = \mathbf{V}_{sr}\mathbf{\Lambda}\mathbf{V}_{sr}^{\dagger}$*, where* $\mathbf{\Lambda}$ *is real diagonal and* $\mathbf{V}_{sr}$ *represents the right singular matrix of* $\mathbf{H}_{sr}$ *(i.e.,* $\mathbf{H}_{sr} = \mathbf{U}_{sr}\mathbf{\Sigma}_{sr}\mathbf{V}_{sr}^{\dagger}$*, where* $\mathbf{\Sigma}_{sr}$ *is the diagonal singular value matrix and* $\mathbf{U}_{sr}$ *is the left singular matrix).*

C. Problem Reformulation

Based on the diagonal decomposition results for $\mathbf{A}$ and $\mathbf{Q}$, we can simplify the PO-PR problem to a scalar form. We start with the objective function in (10). After some algebraic manipulations from Theorem 1, we obtain that

$$I_{PR}(\mathbf{Q},\mathbf{A}) = \frac{1}{2}\sum_{j=1}^{MN_r}\log_2\left[1+\frac{\tilde{\lambda}_{sr}^{(j)}\lambda_{rd}^{(j)}(\sum_{i=1}^{M}d_j^{(i)})^2}{\lambda_{rd}^{(j)}(\sum_{i=1}^{M}d_j^{(i)})^2+\frac{\sigma_d^2}{\sigma_r^2}}\right]. \quad (13)$$

Here, the eigenvalues $\{\tilde{\lambda}_{sr}^{(j)}\}_{j=1}^{MN_r}$ and $\{\lambda_{rd}^{(j)}\}_{j=1}^{MN_r}$ are sorted in nonincreasing order based on an argument similar to [13, Appendix IV]. Also, the source power constraint can be re-written as $\sum_{j=1}^{MN_r}\tilde{\lambda}_{sr}^{(j)} \le P_T$. By using $\mathrm{Tr}(\mathbf{MXNX}^{\dagger}) = \mathbf{x}^{\dagger}(\mathbf{M}\circ\mathbf{N}^T)\mathbf{x}$, where $\mathbf{M},\mathbf{N}$ are square and $\mathbf{X}$ is a diagonal matrix with $\mathbf{x}$ on its main diagonal [19], [23], we have that

$$\mathrm{Tr}(\mathbf{A}_i(\sigma_r^2\mathbf{I}_{N_r}+\mathbf{H}_{sr}^{[i]}\mathbf{Q}(\mathbf{H}_{sr}^{[i]})^{\dagger})\mathbf{A}_i^{\dagger}) \le P_R \Rightarrow \mathbf{d}_i^{\dagger}\mathbf{S}_i\mathbf{d}_i \le P_R,$$

where $\mathbf{S}_i \in \mathbb{C}^{MN_r\times MN_r}$ is defined as

$$\mathbf{S}_i = \left[(((\mathbf{V}_{rd}^{[i]})^{\dagger})^{\mathrm{LI}})^{\dagger}((\mathbf{V}_{rd}^{[i]})^{\dagger})^{\mathrm{LI}}\right]\circ \left[(\mathbf{U}_{sr}^{[i]})^{\mathrm{RI}}(\sigma_r^2\mathbf{I}_{N_r}+\mathbf{H}_{sr}\mathbf{Q}\mathbf{H}_{sr}^{\dagger})((\mathbf{U}_{sr}^{[i]})^{\mathrm{RI}})^{\dagger}\right]^T. \quad (14)$$

Note that $\mathbf{S}_i$ is symmetric and PSD because it is a Hadamard product of two PSD matrices [23]. Note also that $\tilde{\lambda}_{sr}^{(j)}\lambda_{rd}^{(j)} = 0$ for $j = D+1,\ldots,MN_r$, where $D \triangleq \min\{rank(\tilde{\mathbf{H}}_{sr}), rank(\mathbf{H}_{rd})\}$. Therefore, the last $MN_r - D$ terms on the RHS of (13) can be ignored. For convenience, we let $\delta_j \triangleq \sum_{i=1}^{M}d_j^{(i)}$. It then follows that Problem PO-PR can be simplified to the following equivalent problem:

PO-PR-SIM:

$$\text{Max}\ \ \frac{1}{2}\sum_{j=1}^{D}\log_2\left[1+\frac{\tilde{\lambda}_{sr}^{(j)}\lambda_{rd}^{(j)}\delta_j^2}{\lambda_{rd}^{(j)}\delta_j^2+\frac{\sigma_d^2}{\sigma_r^2}}\right] \quad (15)$$

$$\text{s.t.}\ \ \sum_{j=1}^{MN_r}(\mathbf{S}_i)_{jj}(d_j^{(i)})^2+\sum_{j=1}^{MN_r}\sum_{k=1,\neq j}^{N_r}(\mathbf{S}_i)_{jk}d_j^{(i)}d_k^{(i)} \le P_R, \forall i,$$

$$\delta_j - \sum_{i=1}^{M}d_j^{(i)} = 0,\ j=1,\ldots,D, \quad \sum_{j=1}^{MN_r}\tilde{\lambda}_{sr}^{(j)} \le P_T,$$

where the decision variables are δ_j, $d_j^{(i)}$, and $\tilde{\lambda}_{sr}^{(j)}$, $\forall i,j$. Note that matrix variables no longer appear in PO-PR-SIM, which significantly reduces the computational complexity. This simplified formulation will serve as a foundation for us to design a global optimization algorithm in Section VI.

V. Optimal Power Allocation Structure: The Cooperative Relay Case

In this section, we turn our attention to the more complex case of CR. The major difference between CR and PR is the presence of the direct link $\mathbf{H}_{sd}$. In Section V-A, we will first study the impacts of the direct link and AF relays and what roles they play under CR. Based on this, we will investigate the optimal structures of $\mathbf{A}$ and $\mathbf{Q}$ under CR in Section V-B.

A. The Roles of the Direct Link and AF Relays

Recall that $I_{CR}(\mathbf{Q},\mathbf{A})$ can be written as in (5) and the equivalent end-to-end channel gain matrix $\mathbf{H}$ and noise power $\mathbf{R}$ can be written as in (6). To compute the determinant in (5), we substitute (6) into (5), which yields:

$$\begin{aligned}&\left|\mathbf{I}_{2N_d}+\mathbf{H}\mathbf{Q}\mathbf{H}^{\dagger}\mathbf{R}^{-1}\right| \\ &= \left|\mathbf{I}_{2N_d}+\begin{bmatrix}\bar{\mathbf{H}}_{sd}\\ \mathbf{H}_{sd}\end{bmatrix}\mathbf{Q}\begin{bmatrix}\bar{\mathbf{H}}_{sd}^{\dagger} & \mathbf{H}_{sd}^{\dagger}\end{bmatrix}\begin{bmatrix}\bar{\mathbf{R}}^{-1} & \mathbf{0}\\ \mathbf{0} & \frac{1}{\sigma_d^2}\mathbf{I}_{N_d}\end{bmatrix}\right| \\ &= \begin{vmatrix}\mathbf{I}_{N_d}+\bar{\mathbf{H}}_{sd}\mathbf{Q}\bar{\mathbf{H}}_{sd}^{\dagger}\bar{\mathbf{R}}^{-1} & \frac{1}{\sigma_d^2}\bar{\mathbf{H}}_{sd}\mathbf{Q}\mathbf{H}_{sd}^{\dagger}\\ \mathbf{H}_{sd}\mathbf{Q}\bar{\mathbf{H}}_{sd}^{\dagger}\bar{\mathbf{R}}^{-1} & \mathbf{I}_{N_d}+\frac{1}{\sigma_{sd}^2}\mathbf{H}_{sd}\mathbf{Q}\mathbf{H}_{sd}^{\dagger}\end{vmatrix}. \quad (16)\end{aligned}$$

Based on the determinant in (16), $I_{CR}(\mathbf{Q},\mathbf{A})$ can be decomposed into two parts as follows:

$$I_{CR}(\mathbf{Q},\mathbf{A}) = \underbrace{\frac{1}{2}\log_2\left|\mathbf{I}_{N_d}+\frac{1}{\sigma_d^2}\mathbf{H}_{sd}\mathbf{Q}\mathbf{H}_{sd}^{\dagger}\right|}_{\text{Direct link without relays}}+ \underbrace{\frac{1}{2}\log_2\left|\mathbf{I}_{N_d}+\bar{\mathbf{H}}_{sd}\left(\mathbf{Q}^{-1}+\frac{1}{\sigma_{sd}^2}\mathbf{H}_{sd}^{\dagger}\mathbf{H}_{sd}\right)^{-1}\bar{\mathbf{H}}_{sd}^{\dagger}\bar{\mathbf{R}}^{-1}\right|}_{\text{Gain from AF relays}}. \quad (17)$$

Upon a closer look at the RHS of (17), it is easy to recognize that the first term is exactly the MIMO capacity expression for the direct link without the relays. Hence, the second term represents the rate gain due to the cooperation from the relays. The derivation of (17) is based on a "backward" use of the Sherman-Morrison-Woodbury matrix inversion lemma [24] and is similar to deriving the mutual information expression for three-node MIMO AF-relay channel in [13]. We relegate the derivation details to [11]. By expanding $\bar{\mathbf{H}}_{sd}$ and $\bar{\mathbf{R}}$ in (17), we arrive at the following result:

Proposition 2. *Under CR, the end-to-end achievable rate gain due to the AF relay links, denoted by* ΔC_{AF}*, is given by*

$$\Delta C_{AF} = \frac{1}{2}\log_2\left|\mathbf{I}_{N_d}+\left(\sum_{i=1}^{M}\mathbf{H}_{rd}^{[i]}\mathbf{A}_i\mathbf{H}_{sr}^{[i]}\right)\left(\mathbf{Q}^{-1}+\frac{1}{\sigma_d^2}\mathbf{H}_{sd}^{\dagger}\mathbf{H}_{sd}\right)^{-1}\left(\sum_{i=1}^{M}\mathbf{H}_{rd}^{[i]}\mathbf{A}_i\mathbf{H}_{sr}^{[i]}\right)^{\dagger}\bar{\mathbf{R}}^{-1}\right|.$$

Remark 3. Proposition 2 can also be interpreted in an asymptotic manner: as $\mathbf{H}_{sd} \to \mathbf{0}$ (i.e., the direct link gets weaker), the capacity of the direct link in (17) approaches zero and $\Delta C_{AF} \to I_{PR}(\mathbf{Q},\mathbf{A})$, which makes intuitive sense.

B. The Structure of the Optimal $\mathbf{A}$ and $\mathbf{Q}$ under CR

We can see from (17) that when $\mathbf{Q}$ is fixed, an optimal $\mathbf{A}$ can be found by maximizing the second term in (17). Comparing this term to (4), we can see that they are of the same structure. This means that the optimal structure of $\mathbf{A}$ does not change under CR. We state this result as follows:

Proposition 3. *For a given $\mathbf{Q}$ under CR, the structural property of $\mathbf{A}$ in (12) continues to hold. However, $\tilde{\mathbf{U}}_{sr}^{[k]}$ needs to be replaced by the corresponding submatrix block of the left singular matrix of $\frac{1}{\sigma_r}\mathbf{H}_{sr}(\mathbf{Q}^{-1}+\frac{1}{\sigma_d^2}\mathbf{H}_{sd}^{\dagger}\mathbf{H}_{sd})^{-\frac{1}{2}}$.*

Compared to the case of PR, identifying the optimal structure of $\mathbf{Q}$ under CR is more involved. Due to the first term in (17), we need to simultaneously take $\mathbf{H}_{sd}$ and $\mathbf{H}_{sr}$ into consideration. This means that the optimal structure of $\mathbf{Q}$ depends on $\mathbf{A}$, which is different from the case of PR where the optimal $\mathbf{Q}$ is solely dictated by $\mathbf{H}_{sr}$. We first rewrite the expression in (5) as $I_{CR}(\mathbf{Q},\mathbf{A}) = \frac{1}{2}\log_2\left|\mathbf{I}_{2N_d}+\hat{\mathbf{H}}\mathbf{Q}\hat{\mathbf{H}}^{\dagger}\right|$, where $\hat{\mathbf{H}} \triangleq \mathbf{R}^{-\frac{1}{2}}\mathbf{H}$. We can see that this expression is equivalent to a point-to-point MIMO channel with the channel gain matrix being $\hat{\mathbf{H}}$. Thus, the proposition below follows from the classical water-filling result [25]:

Proposition 4. *The optimal $\mathbf{Q}$ under CR can be decomposed as $\mathbf{Q}=\hat{\mathbf{V}}\mathbf{\Lambda}\hat{\mathbf{V}}^{\dagger}$, where $\mathbf{\Lambda}$ is real diagonal and $\hat{\mathbf{V}}$ represents the right singular matrix of $\hat{\mathbf{H}}$ (i.e., $\hat{\mathbf{H}}=\hat{\mathbf{U}}\hat{\mathbf{\Sigma}}\hat{\mathbf{V}}^{\dagger}$, where $\hat{\mathbf{\Sigma}}$ is the diagonal singular value matrix and $\hat{\mathbf{U}}$ is the left singular matrix).*

Propositions 3 and 4 indicate that, even though the channels under CR become more complex, the same matching structures of matrices $\mathbf{A}$ and $\mathbf{Q}$ continue to hold.

VI. Optimization Algorithm

Based on the structural results in Sections IV and V, we are now in a position to optimize $\mathbf{Q}$ and $\mathbf{A}$. According to Propositions 3 and 4, it suffices to only consider designing an algorithm for PR. Recall that PO-PR can be decomposed into an inner subproblem with $\mathbf{Q}$ being fixed and an outer main program with respect to $\mathbf{Q}$. Also, it is not difficult to see that (15) is concave with respect to the $\tilde{\lambda}_{sr}^{(j)}$-variables. Thus, $\mathbf{Q}$ can be readily solved by standard convex programming tools once we know how to determine $\mathbf{A}$. Hence, we would only focus on the inner subproblem in the remainder of this section.

A. Convexity Property of PO-PR-SIM

Before solving the inner subproblem of PO-PR-SIM (i.e., with $\tilde{\lambda}_{sr}^{(j)}$ fixed), we first examine its convexity property. In PO-PR-SIM, the second constraint is linear. The first constraint is a Schur product involving a PSD matrix, which means it is convex as well. However, the objective function (15) can be written as a difference of two concave functions: $\frac{1}{2}\sum_{j=1}^{D}\log_2\left((1+\tilde{\lambda}_{sr}^{(j)})\lambda_{rd}^{(j)}\delta_j^2+\frac{\sigma_d^2}{\sigma_r^2}\right) - \frac{1}{2}\sum_{j=1}^{D}\log_2\left(\lambda_{rd}^{(j)}\delta_j^2+\frac{\sigma_d^2}{\sigma_r^2}\right)$, which implies that it is not concave. In fact, by checking the positive definiteness of the Hessian of (15), we can show that PO-PR-SIM is convex if the following condition is satisfied:

$$\left(\sum_{i=1}^{M} d_j^{(i)}\right)^2 \geq \frac{\sigma_d^2}{6\lambda_{rd}^{(j)}\sigma_c^2}\left[\sqrt{\left(\frac{\tilde{\lambda}_{sr}^{(j)}+2}{\tilde{\lambda}_{sr}^{(j)}+1}\right)^2+\left(\frac{12}{\tilde{\lambda}_{sr}^{(j)}+1}\right)} - \left(\frac{\tilde{\lambda}_{sr}^{(j)}+2}{\tilde{\lambda}_{sr}^{(j)}+1}\right)\right],\ j=1,\ldots,D. \quad (18)$$

The derivation of (18) is similar to that given in [19] and is thus omitted here for brevity. By observing (18), we can see that as $\tilde{\lambda}_{sr}^{(j)}\to\infty$ or $\lambda_{rd}^{(j)}\to\infty$, the RHS approaches zero. Thus, we have the following result:

Proposition 5. *The threshold value on $\left(\sum_{i=1}^{M} d_j^{(i)}\right)^2$ in (18) approaches zero as $\tilde{\lambda}_{sr}^{(j)}\to\infty$ or $\lambda_{rd}^{(j)}\to\infty$.*

Remark 4. Since $\tilde{\lambda}_{sr}^{(j)}$ and $\lambda_{rd}^{(j)}$ represent the quality of the source-relay and relay-destination channels, Proposition 5 implies that PO-PR-SIM is more likely to be convex if the channels $\mathbf{H}_{sr}$ and $\mathbf{H}_{rd}$ are strong.

Due to the nonconvexity of PO-PR-SIM, determining a global optimal solution remains challenging even though the problem has been reduced to a scalar form. Rather than settling with a heuristic or local optimal solution, we propose a global optimization approach based on the branch-and-bound framework coupled with a custom-desgined convex programming relaxation (BB/CPR) [26]–[28] that exploits the special structure in PO-PR-SIM. To the best of our knowledge, this work is the first that considers solving nonconvex MIMO-CN problems using a global optimization approach.

B. A Global Optimization Approach

The basic idea of BB/CPR: Here, we provide an overview on using BB to solve PO-PR-SIM. We refer readers to [29] for a comprehensive understanding of the BB procedure. The BB procedure proceeds iteratively as follows. During the initial step, an upper bound on the objective value is obtained by solving a convex programming relaxation (CPR) of PO-PR-SIM. The obtained solution from CPR serves to provide an incumbent solution to PO-PR-SIM and a global lower bound on the objective value, denoted as LB. Next, we partition the problem into two subproblems at the CPR solution. The CPR for each of these two subproblems is then solved to obtain two sets of lower and upper bounds. LB is then updated to the maximum of the two new lower bounds if this value is greater than the current LB. This step completes an iteration.

After an iteration, if the gap between LB and the largest upper bound (among all the subproblems) is larger than some predefined desired error ϵ, we perform another iteration on the subproblem having the largest upper bound. Also, during each iteration, we can remove those subproblems whose upper bounds have a gap less than ϵ compared to LB (since further branching on these subproblems could not yield improved feasible solutions beyond the ϵ-tolerance), thus controlling the increase of the number of subproblems. The BB iterations continue until the largest upper bound is within ϵ of the LB. Therefore, the best feasible solution is $(1-\epsilon)$-optimal. A more detailed description of BB/CPR is shown in Algorithm 1.

Algorithm 1 BB/CPR Solution Procedure

Initialization:

1. Let the optimal solution $\psi^* = \emptyset$ and the initial lower bound $LB = -\infty$.
2. Determine partitioning variables (variables associated with CPR) and derive their initial bounding intervals.
3. Let the initial problem list contain only the original problem, denoted by P_1.
4. Construct CPR based on the partitioning variables. Denote the solution to CPR as ψ_1 and its objective value as the upper bound UB_1.

Main Loop:

5. Select problem P_z that has the largest upper bound among all problems in the problem list.
6. Find a feasible solution ψ_z by solving Problem P_z. Denote the objective value of ψ_z by LB_z.
7. If $LB_z > LB$ then let $\psi^* = \psi_z$ and $LB = LB_z$. If $(1+\epsilon)LB \geq UB$ then stop with the $(1-\epsilon)$-optimal solution ψ^*; else, remove all problems $P_{z'}$ having $(1+\epsilon)LB_{z'} \geq UB$ from the problem list.
8. Compute relaxation error for each partitioning variable.
9. Select a partitioning variable having the maximum relaxation error and divide its bounding interval into two new intervals by partitioning at its value in ψ_z.
10. Remove the selected problem P_z from the problem list, and construct two new problems P_{z1} and P_{z2} based on the two partitioned intervals.
11. Compute two new upper bounds UB_{z1} and UB_{z2} by solving the CPR for P_{z1} and P_{z2}, respectively.
12. If $(1+\epsilon)LB < UB_{z1}$ then add problem P_{z1} to the problem list. If $(1+\epsilon)LB < UB_{z2}$ then add problem P_{z2} to the problem list.
13. If the problem list is empty, stop with the $(1-\epsilon)$-optimal solution ψ^*. Otherwise, go to Step 5.

In what follows, we will develop the corresponding key components in the BB/CPR framework that are problem-specific.

Further reformulation of PO-PR-SIM. Note that in PO-PR-SIM, the nonconvexity stems from the objective function. Therefore, we start by rewriting (15) as

$$\tfrac{1}{2}\sum_{j=1}^{D}\log_2\left[1+\frac{\tilde{\lambda}_{sr}^{(j)}\lambda_{rd}^{(j)}g_j}{\lambda_{rd}^{(j)}g_j+\sigma^2}\right], \tag{19}$$

where $\sigma^2 \triangleq \frac{\sigma_d^2}{\sigma_r^2}$ and $g_j \triangleq \delta_j^2 \geq 0$. Note that, although not obvious, Eq. (19) is convex with respect to $g_j \geq 0$, $\forall j = 1,\ldots,D$. This can be readily verified by checking the second order condition of (19). With this change of variables, a quadratic equality constraint $g_j = \delta_j^2$ needs to be appended to the inner problem. Although this constraint remains nonconvex, we have shifted the nonconvex part away from the objective function. As we shall see immediately, with this reformulation, we can apply a powerful technique called the *Reformulation-Linearization Technique* (RLT) [26], [27].

Convexifying the quadratic equality constraints. The basic idea in our convexification is to approximate the quadratic equality constraint $g_j = \delta_j^2$ using its *convex hull relaxation*. We first observe that if we replace $g_j = \delta_j^2$ by $g_j \geq \delta_j^2$, then the latter is convex. However, since PO-PR-SIM is a maximization problem in g_j, we need an additional constraint to bound g_j from above because otherwise, PO-PR-SIM will be unbounded. To this end, we can apply RLT as follows. First, we note that

$$\delta_j^L \leq \delta_j \leq \delta_j^U, \tag{20}$$

where δ_j^L and δ_j^U are some appropriate lower and upper bounds for δ_j, respectively (more on the values of δ_j^L and δ_j^U later). From (20), we derive the following *bound-factor constraint*: $(\delta_j - \delta_j^L)(\delta_j - \delta_j^U) \leq 0$, which, upon using the substitution $g_j = \delta_j^2$, yields: $g_j - (\delta_j^L + \delta_j^U)\delta_j + \delta_j^L\delta_j^U \leq 0$.

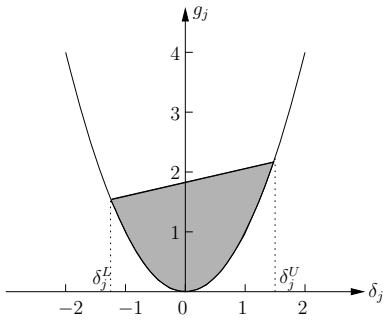

Fig. 4. The quadratic curve $g_j = \delta_j^2$ is approximated by its convex hull as expressed by the RLT constraints in (21) and (22).

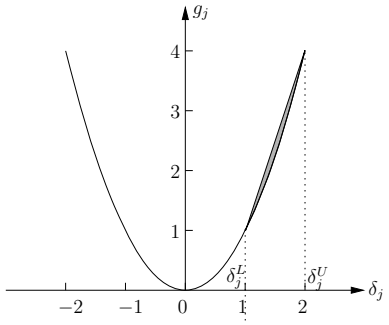

Fig. 5. The quadratic curve $g_j = \delta_j^2$ can be well approximated by its convex hull when the interval $[\delta_j^L, \delta_j^U]$ is small.

Observe now that the nonconvex constraint $g_j = \delta_j^2$ has been relaxed into two convex constraints in g_j and δ_j:

$$\delta_j^2 - g_j \leq 0 \qquad \forall j = 1,\ldots,D, \tag{21}$$

$$g_j - (\delta_j^L + \delta_j^U)\delta_j + \delta_j^L\delta_j^U \leq 0 \qquad \forall j = 1,\ldots,D. \tag{22}$$

Geometrically, Constraints (21) and (22) represent the convex hull of the set defined by $g_j = \delta_j^2$, $\delta_j \in [\delta_j^L,\ \delta_j^U]$, as shown in Fig. 4. When the bounds δ_j^L and δ_j^U are far from each other (i.e., the interval $[\delta_j^L, \delta_j^U]$ is large), the convex hull relaxation may not be a tight approximation of the quadratic curve. However, as δ_j^L and δ_j^U get closer, as shown in Fig. 5, we can see that the convex hull quickly becomes an excellent approximation of the quadratic curve. With the above convex relaxation, we obtain the following CPR:

R-PO-PR-SIM:

$$\text{Max}\ \ \frac{1}{2}\sum_{j=1}^{D}\log_2\left[1+\frac{\tilde{\lambda}_{sr}^{(j)}\lambda_{rd}^{(j)}g_j}{\lambda_{rd}^{(j)}g_j+\frac{\sigma_d^2}{\sigma_r^2}}\right]$$

$$\text{s.t.}\ \ \sum_{j=1}^{MN_r}(\mathbf{S}_i)_{jj}(d_j^{(i)})^2 + \sum_{j=1}^{MN_r}\sum_{k=1,\neq j}^{MN_r}(\mathbf{S}_i)_{jk}d_j^{(i)}d_k^{(i)} \leq P_R, \forall i,$$

$$\delta_j^2 - g_j \leq 0,\ g_j - (\delta_j^L + \delta_j^U)\delta_j + \delta_j^L\delta_j^U \leq 0,\ j = 1,\ldots,D,$$

$$\delta_j - \sum_{i=1}^{M} d_j^{(i)} = 0,\ j = 1,\ldots,D.$$

Upper and lower bounds for partitioning variables. The partitioning variables in BB/CPR are those that are involved in nonconvex terms, for which we have defined new variables and whose bounding intervals will need to be partitioned

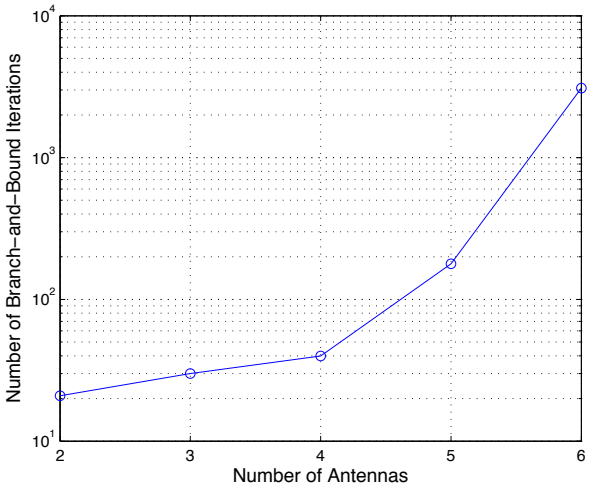

Fig. 6. The scaling of the number of BB iterations with respect to the number of per-node antennas.

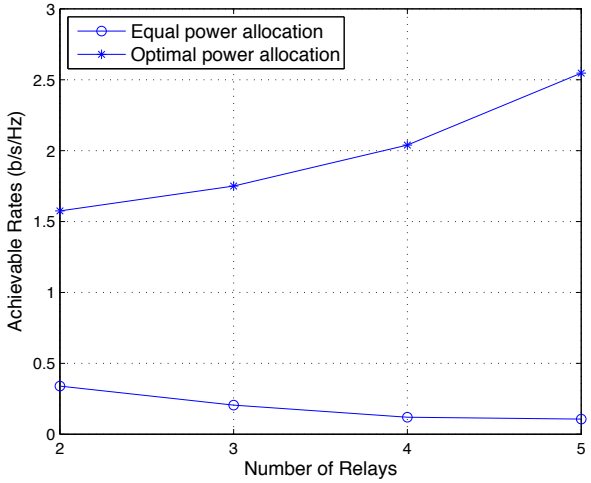

Fig. 8. The achievable rate comparison between equal power allocation and optimal power allocation as the number of relays changes from 2 to 5.

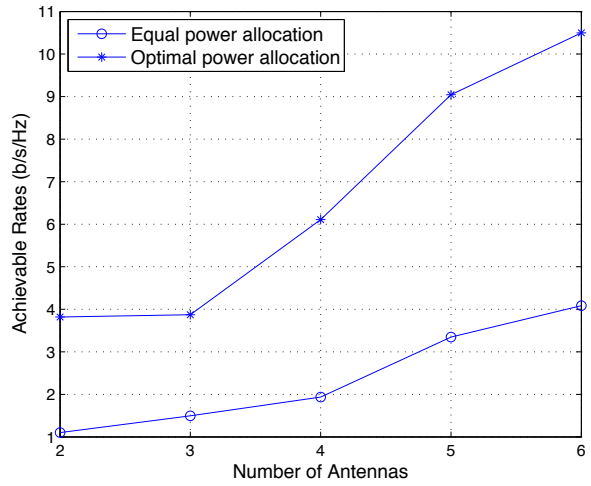

Fig. 7. Achievable rate comparison between equal power allocation and optimal power allocation as the number of per-node antennas changes from 2 to 6.

during BB [26]–[28]. In R-PO-PR-SIM, these partitioning variables are $\delta_j, j = 1, \ldots, D$. To derive tight upper and lower bounds for δ_j, we let $\bar{\mathbf{S}}_i^{(j)} \in \mathbb{R}^{(MN_r-1)\times MN_r}$ be the matrix obtained by taking the real parts of $\mathbf{S}_i$ and then deleting the j-th row. Also, let $\bar{\mathbf{v}}_i^{(j)} = [\bar{v}_{i,1}^{(j)} \cdots \bar{v}_{i,MN_r}^{(j)}]^T$ be the right singular vector corresponding to the zero singular value of $\bar{\mathbf{S}}_i^{(j)}$. Further, define $\hat{\mathbf{v}}_i^{(j)}$ as the scaled version of $\bar{\mathbf{v}}_i^{(j)}$ obtained by $\hat{\mathbf{v}}_i^{(j)} = \frac{1}{(\bar{\mathbf{v}}_i^{(j)})_j}\bar{\mathbf{v}}_i^{(j)}$. Then, we have the following closed-form result:

Lemma 4. *The upper and lower bounds of δ_j can be respectively computed as:*

$$\delta_j^L = -\sum_{i=1}^{M}\sqrt{\frac{P_R}{(\hat{\mathbf{v}}_i^{(j)})^\dagger \mathbf{S}_i \hat{\mathbf{v}}_i^{(j)}}},\ \delta_j^U = \sum_{i=1}^{M}\sqrt{\frac{P_R}{(\hat{\mathbf{v}}_i^{(j)})^\dagger \mathbf{S}_i \hat{\mathbf{v}}_i^{(j)}}}. \quad (23)$$

Lemma 4 can be proved by studying $\max d_j^{(i)}$ because $\delta_j^U = \max \sum_{i=1}^M d_j^{(i)} \le \sum_{i=1}^M \max d_j^{(i)}$ and $\delta_j^L = \min \sum_{i=1}^M d_j^{(i)} \ge -\sum_{i=1}^M \max d_j^{(i)}$. $\max d_j^{(i)}$ can be obtained by solving

$$\begin{aligned} &\text{Max} \quad d_{j'}^{(i)} \quad (24)\\ &\text{s.t.} \quad \sum_{j=1}^{MN_r}(\mathbf{S}_i)_{jj}(d_j^{(i)})^2 + \sum_{j=1}^{MN_r}\sum_{k=1,k\neq j}^{MN_r}(\mathbf{S}_i)_{jk}d_j^{(i)}d_k^{(i)} \le P_R. \end{aligned}$$

Since Problem (24) is convex and the Slater condition holds, the result in Lemma 4 follows from analyzing the KKT system of (24) (see [24]). Due to space limitation, we relegate the proof details of Lemma 4 to [11].

C. Numerical Results

We present some numerical results to demonstrate the efficacy of our proposed BB/CPR method. First, we study the scaling of BB iterations as the numbers of relays and per-node antennas grow. In this paper, the number of BB iterations depends on the number of δ-variables, which is in turn determined by the number of per-node antennas. Thus, we expect that the number of per-node antennas has a more direct impact than the number of relays does. We first fix the number of relays to 4 and vary the number of per-node antennas from 2 to 6, which is the practical range in real systems. All other simulation settings remain the same as in the previous example. In this case, the scaling of the number of BB iterations is shown in Fig. 6 (in log-scale). Each data point in Fig. 6 is averaged over 10 randomly generated network instances. The averaged number of iterations changes from 21 to 3090, which increases roughly exponentially. This confirms that the number of BB iterations depends on the number of per-node antennas. The exponential increase is an expected phenomenon when searching for global optimal solutions for NP-hard nonconvex optimization problems. However, we note that the BB process converges reasonably fast for practical numbers of antennas. On the other hand, we also evaluate the scaling of the number of BB iterations with respect to the number of relays. We fix the number of per-node antennas to 2 and vary the number of relays from 2 to 6. For each setting, the result is again averaged over 10 randomly generated network

instances. The number of BB iterations vary as 16, 25, 20, 18, and 19, respectively, which shows that the solution effort is relatively insensitive to the number of relays.

It is also interesting to study how much rate gain can be obtained through an optimal power allocation. In Fig. 7, we compare the achievable rates under optimal power allocation and the naive equal power allocation as the number of per-node antennas increases. We can see that the achievable rate gain is significant under optimal power allocation. For these five settings, the average ratios between equal power allocation and optimal power allocation are 0.2849, 0.3914, 0.3171, 0.3699, and 0.3891, respectively. In other words, the achievable rates under equal power allocation are no more than 40% of that under an optimal power allocation. In Fig. 8, we compare the achievable rates under optimal power allocation and under equal power allocation as the number of relays increases. Again, we can see that the achievable rate gain is significant under optimal power allocation. For these four settings, the average ratios between equal power allocation and optimal power allocation are 0.2154, 0.1173, 0.0586, and 0.0419, respectively. We note that, under optimal power allocation, the achievable rate increases as the number of relays grows. In contrast, under equal power allocation, the achievable rate actually decreases. This highlights the importance of power allocation optimization in multi-relay MIMO-CN.

VII. Conclusion

In this paper, we have investigated the structural properties of optimal power allocation in multi-relay MIMO-CN and designed algorithms to solve the optimal power allocation problem. Our contributions include generalizing the matching result under the basic three-node setting to the multi-relay setting with per-node power constraints, for which the optimal power allocation has been heretofore unknown. We also quantified the performance gain due to cooperative relay and established the connection between cooperative relay and pure relay. Based on the derived structural insights, we reduced the MIMO-CN rate maximization problem to an equivalent scalar form, which allowed us to develop an efficient global optimization algorithm by using a branch-and-bound framework coupled with a custom-designed convex programming relaxation. Our results in this paper offer important analytical tools and insights to fully exploit the potential of AF-based MIMO-CN with multiple relays. More importantly, our proposed global optimization approach overcomes the fundamental nonconvex difficulty encountered in large-scale MIMO cooperative networks, which we hope will benefit future research in this area.

Acknowledgements

The authors thank the editors of this JSAC special issue and the anonymous reviewers for their constructive comments, which have helped improve the quality of this work. The authors also thank the FABE department and OARDC at The Ohio State University for additional funding support.

References

[1] J. N. Laneman, D. N. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.

[2] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity – Part I and Part II," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1927–1948, Nov. 2003.

[3] E. C. van der Meulen, "Three-terminal communication channels," *Advanced Applied Probability*, vol. 3, pp. 120–154, 1971.

[4] T. M. Cover and A. A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 5, pp. 572–584, Sep. 1979.

[5] E. M. Yeh and R. A. Berry, "Throughput optimal control of cooperative relay networks," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3827–3833, Oct. 2007.

[6] T. Himsoon, W. P. Siriwongpairat, Z. Han, and K. J. R. Liu, "Lifetime maximization via cooperative nodes and relay deployment in wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 2, pp. 306–317, Feb. 2007.

[7] S. Sharma, Y. Shi, Y. T. Hou, H. D. Sherali, and S. Kompella, "Cooperative communications in multi-hop wireless networks: Joint flow routing and relay node assignment," in *Proc. IEEE INFOCOM*, San Diego, CA, Mar. 15-19, 2010, pp. 1–9.

[8] P. Liu, Z. Tao, S. Narayanan, T. Korakis, and S. S. Panwar, "CoopMAC: A cooperative MAC for wireless LANs," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 2, pp. 340–354, Feb. 2007.

[9] Y. Shi, S. Sharma, Y. T. Hou, and S. Kompella, "Optimal relay assignment for cooperative communications," in *Proc. ACM MobiHoc*, Hong Kong SAR, China, May 2008, pp. 3–12.

[10] J. Cai, X. Shen, J. W. Mark, and A. S. Alfa, "Semi-distributed user relaying algorithm for amplify-and-forward wireless relay networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 4, pp. 1348–1357, Apr. 2008.

[11] J. Liu, N. B. Shroff, and H. D. Sherali, "Optimal power allocation in multi-relay MIMO cooperative networks: Theory and algorithms," *Technical Report, Deptment of ECE, Ohio State University*, Jul. 2010. [Online]. Available: http://www2.ece.ohio-state.edu/~liu/publications/AF_MIMO_MRELAY_TR.pdf

[12] X. Tang and Y. Hua, "Optimal design of non-regenerative MIMO wireless relays," *IEEE Trans. Wireless Commun.*, vol. 6, no. 4, pp. 1398–1407, Apr. 2007.

[13] O. Muñoz-Medina, J. Vidal, and A. Agustín, "Linear transceiver design in nonregenerative relays with channel state information," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2593–2604, Jun. 2007.

[14] W. Guan, H. Luo, and W. Chen, "Linear relaying scheme for MIMO relay system with QoS requirements," *IEEE Signal Process. Lett.*, vol. 15, pp. 697–700, 2008.

[15] A. Behbahani, R. Merched, and A. M. Eltawil, "Optimizations of a MIMO relay network," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 5062–5073, Oct. 2008.

[16] Y. Rong, X. Tang, and Y. Hua, "A unified framework for optimizing linear nonregenerative multicarrier MIMO relay communication systems," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4837–4851, Dec. 2009.

[17] B. Khoshnevis, W. Yu, and R. Adve, "Grassmannian beamforming for MIMO amplify-and-forward relaying," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1397–1407, Oct. 2008.

[18] Z. Fang, Y. Hua, and J. D. Koshy, "Joint source and relay optimization for non-regenerative MIMO relay," in *Proc. IEEE Workshop Sensor Array Multi-Channel Signal Processing*, Waltham, WA, Jul. 12-14, 2006, pp. 239–243.

[19] Y. Fu, L. Yang, W.-P. Zhu, and Z. Yang, "A convex optimization design of relay precoder for two-hop MIMO relay networks," in *Proc. IEEE ICC*, Cape Town, South Africa, May 23-27, 2010.

[20] Y. Rong, "Non-regenerative multicarrier MIMO relay communications based on minimization of mean-squared error," in *Proc. IEEE ICC*, Dresden, Germany, Jun. 14-18, 2009, pp. 1–5.

[21] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: John Wiley & Sons, Inc., 1991.

[22] D. J. Love, R. W. Heath, W. Santipach, and M. Honig, "What is the value of limited feedback for MIMO channels?" *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 54–59, Oct. 2004.

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

[24] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Nonlinear Programming: Theory and Algorithms*, 3rd ed. New York, NY: John Wiley & Sons Inc., 2006.

[25] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecomm.*, vol. 10, no. 6, pp. 585–596, Nov. 1999.
[26] H. D. Sherali and W. P. Adams, *A Reformulation-Linearization-Technique for Solving Discrete and Continuous Nonconvex Problems*. Boston, MA: Kluwer Academic Publishing, 1999.
[27] H. D. Sherali and C. H. Tuncbilek, "A global optimization algorithm for polynomial programming problems using a reformulation-linearization technique," *Journal of Global Optimization*, vol. 2, no. 1, pp. 101–112, 1992.
[28] H. D. Sherali and H. Wang, "Global optimization of nonconvex factorable programming problems," *Mathematical Programming*, vol. 89, no. 3, pp. 459–478, 2001.
[29] G. L. Nemmhauser and L. A. Wolsey, *Integer and Combinatorial Optimization*, 2nd ed. New York: Wiley-Interscience Publication, 1999.

Jia Liu (S'03–M'10) received his B.S. and M.S. degrees from South China University of Technology in 1996 and 1999, respectively, both in Electrical Engineering. He also received his B.S. degree in Computer Science from South China University of Technology in 1996. He received his Ph.D. degree in the Bradley Department of Electrical and Computer Engineering at Virginia Tech, Blacksburg, VA in 2010. Since April 2010, he has been working as a Postdoctoral Researcher in the Department of Electrical and Computer Engineering at The Ohio State University. His research focus is in the areas of optimization of communication systems and networks, theoretical foundations of cross-layer optimization on wireless networks, design of algorithms, and information theory. Prior to joining Virginia Tech, Dr. Liu was with Bell Labs, Lucent Technologies in Beijing China from 1999 to 2003.

Dr. Liu is a member of IEEE and SIAM. He is a regular reviewer for major IEEE conferences and journals. He was a recipient of the Bell Labs President Gold Award in 2001, Chinese Government Award for Outstanding Ph.D. Students Abroad in 2008, and IEEE ICC 2008 Best Paper Award. His paper was also the only Runner-up paper of IEEE INFOCOM 2011 Best Paper Award. He served as a TPC member for IEEE INFOCOM 2010 and 2011.

Ness B. Shroff (S'91–M'93–SM'01–F'07) is currently the Ohio Eminent Scholar of Networking and Communications, and Professor of ECE and CSE at The Ohio State University. Previously, he was a Professor of ECE at Purdue University and the director of the Center for Wireless Systems and Applications (CWSA), a university-wide center on wireless systems and applications. His research interests span the areas of wireless and wireline communication networks, where he investigates fundamental problems in the design, performance, pricing, and security of these networks. Dr. Shroff has received numerous awards for his networking research, including the NSF CAREER award, the best paper awards for IEEE INFOCOM 06 and IEEE INFOCOM08, the best paper award for IEEE IWQoS06, the best paper of the year award for the Computer Networks journal, and the best paper of the year award for the Journal of Communications and Networks (JCN) (his IEEE INFOCOM05 paper was one of two runner-up papers). He also currently serves as a guest chaired professor of Wireless Communications in the department of Electronic Engineering, at Tsinghua University, in Beijing, China.

Hanif D. Sherali is a University Distinguished Professor and the W. Thomas Rice Chaired Professor of Engineering in the Industrial and Systems Engineering Department at Virginia Polytechnic Institute and State University. His areas of research interest are in analyzing problems and designing algorithms for specially structured linear, nonlinear, and integer programs arising in various applications, global optimization methods for non-convex programming problems, location and transportation theory and applications, and economic and energy mathematical modeling and analysis. He has published over 290 refereed articles in various Operations Research journals, has (co-) authored eight books in this area, and serves on the editorial board of eight journals. He is a Fellow of INFORMS and IIE, and an elected member of the National Academy of Engineering.