Routing and Power Allocation for MIMO-Based Ad Hoc Networks with Dirty Paper Coding

Jia Liu*, Y. Thomas Hou*, and Hanif D. Sherali†
* Bradley Department of Electrical and Computer Engineering
† Grado Department of Industrial and Systems Engineering
Virginia Polytechnic Institute and State University, Blacksburg, VA 24061

***Abstract*—Recently, researchers showed that "dirty paper coding" (DPC) achieves the capacity region of MIMO Gaussian broadcast channels (MIMO-BC). So far, there has been little study on how this fundamental information-theoretic result will impact the cross-layer design for MIMO-based ad hoc networks. To fill this gap, we consider the problem of jointly optimizing DPC power allocation at the physical layer and multihop/multipath routing at the network layer for MIMO-based ad hoc networks. This optimization problem turns out to be a challenging non-convex problem. To address this difficulty, we transform the original problem to an equivalent problem by exploiting the uplink-downlink duality. For the transformed problem, we propose a solution procedure that integrates Lagrangian dual decomposition, conjugate gradient projection based on matrix differential calculus, and cutting-plane methods.**

I. Introduction

From network information theory perspective, the set of outgoing MIMO links from a node sharing a common communication spectrum can be modeled as a nondegraded MIMO Gaussian broadcast channel (MIMO-BC), for which the capacity region is a well-known hard problem [1]. Recently, significant progress has been made in this area. Most notably, Weigarten *et al.* proved in [2] that "dirty paper coding" (DPC) [3] is the optimal transmission strategy for MIMO-BC in the sense that the DPC rate region $\mathcal{C}_{\mathrm{DPC}}$ is equal to the broadcast channel's capacity region $\mathcal{C}_{\mathrm{BC}}$, i.e., $\mathcal{C}_{\mathrm{BC}} = \mathcal{C}_{\mathrm{DPC}}$.

However, this fundamental information-theoretic result is still not adequately exposed to the wireless networking research community. In current networking literature, most works considering links sharing a common communication spectrum are concerned with how to allocate frequency subbands/time-slots and schedule transmissions to share the common communication spectrum. Although time and frequency division schemes are simple and effective, it has been shown that they are sub-optimal [1]. So far, how to exploit DPC's benefits in the cross-layer design for MIMO-based wireless ad hoc networks has not been studied. The main objective of this paper is to fill this gap and to obtain a rigorous understanding of the impact of employing DPC in MIMO-based ad hoc networks.

However, applying DPC in MIMO-based ad hoc networks is far from trivial. For a K-user broadcast channel, there exist $K!$ encoding orders. Different encoding orders will have different impacts on system performance. Also, since DPC allows interference between links, optimal power allocation needs to be determined. Thus, the optimization for a single MIMO-BC with K users is itself a challenging *combinatorial non-convex* problem, not to mention the cross-layer design in a networking environment with multiple MIMO-BC.

In this paper, our goal is to solve the problem of jointly optimizing DPC power allocation at each node at the physical layer and multihop/multipath routing at the network layer for MIMO-based ad hoc networks. This optimization problem turns out to be a challenging non-convex problem. By exploiting uplink-downlink duality, we show that the original non-convex optimization problem can be transformed into an equivalent convex problem, thus paving the way to efficiently solve the physical layer subproblem. Based on the transformation, we develop an efficient solution procedure that integrates Lagrangian dual decomposition, conjugate gradient projection based on matrix differential calculus, and cutting-plane methods.

The remainder of this paper is organized as follows. In Section II, we discuss the network model and problem formulation. In Section III, we introduce the key components for solving the challenging physical layer subproblem in the Lagrangian decomposition. We provide numerical results in Section IV to illustrate the efficacy of our proposed algorithm. Section V concludes this paper.

II. Network Model

We first introduce notation for matrices, vectors, and complex scalars in this paper. We use boldface to denote matrices and vectors. For a matrix $\mathbf{A}$, $\mathbf{A}^{\dagger}$ denotes the conjugate transpose, $\mathrm{Tr}\{\mathbf{A}\}$ denotes the trace of $\mathbf{A}$, and $|\mathbf{A}|$ denotes the determinant of $\mathbf{A}$. $\mathrm{Diag}\{\mathbf{A}_1, \ldots, \mathbf{A}_n\}$ represents the block diagonal matrix with matrices $\mathbf{A}_1, \ldots, \mathbf{A}_n$ on its main diagonal. We denote $\mathbf{I}$ the identity matrix with dimension determined from the context. $\mathbf{A} \succeq 0$ represents that $\mathbf{A}$ is Hermitian and positive semidefinite (PSD). $\mathbf{1}$ and $\mathbf{0}$ denote vectors whose elements are all ones and zeros, respectively, and their dimensions are determined from the context. $(\mathbf{v})_m$ represents the m^{th} entry of vector $\mathbf{v}$. For a real vector $\mathbf{v}$ and a real matrix $\mathbf{A}$, $\mathbf{v} \geq \mathbf{0}$ and $\mathbf{A} \geq \mathbf{0}$ mean that all entries in $\mathbf{v}$ and $\mathbf{A}$ are nonnegative, respectively. We let $\mathbf{e}_i$ be the unit column vector where the i^{th} entry is 1 and all other entries are 0. The dimension of $\mathbf{e}_i$ is determined from the context as well. The operator "$\langle \cdot, \cdot \rangle$" represents the inner product operation for vectors or matrices.

A. MIMO Gaussian Broadcast Channels and DPC Rate Region

In network information theory, a communication system where a single MIMO-based transmitter sends independent information to multiple *uncoordinated* MIMO-based receivers is referred to as a MIMO broadcast channel (MIMO-BC). Let $\mathbf{H} = [\mathbf{H}_1, \ldots, \mathbf{H}_K]^T$ be the collection of K channel gain matrices in a MIMO-BC, and $\mathbf{\Gamma} = [\mathbf{\Gamma}_1 \ldots, \mathbf{\Gamma}_K]$ be the collection of K input covariance matrices. From the encoding process of DPC, the achievable DPC rate in a MIMO-BC can be computed as follows [4]:

$$R_{\pi(i)}(\mathbf{\Gamma}) = \log_2 \frac{\left|\mathbf{I} + \rho_{\pi(i)}\mathbf{H}_{\pi(i)}\left(\sum_{j\geq i}\mathbf{\Gamma}_{\pi(j)}\right)\mathbf{H}_{\pi(i)}^{\dagger}\right|}{\left|\mathbf{I} + \rho_{\pi(i)}\mathbf{H}_{\pi(i)}\left(\sum_{j> i}\mathbf{\Gamma}_{\pi(j)}\right)\mathbf{H}_{\pi(i)}^{\dagger}\right|}, \quad (1)$$

where $\pi(\cdot)$ denotes a permutation of the set $\{1, \ldots, K\}$, $\rho_{\pi(i)}$ captures the path-loss effect on link $\pi(i)$, $\mathbf{H}_{\pi(i)}$ is the channel gain matrix of link $\pi(i)$. It is evident that (1) is a non-convex function of the input covariance matrices $\mathbf{\Gamma}_i$. Also, the choice of $\pi(\cdot)$ will have a significant impact on $R_{\pi(i)}$.

The dirty paper rate region $\mathcal{C}_{\mathrm{DPC}}(P, \mathbf{H})$ is defined as the convex hull of the union of all such rate vectors over all positive semidefinite covariance matrices $\mathbf{\Gamma}_1, \ldots, \mathbf{\Gamma}_K$ satisfying $\mathrm{Tr}\{\sum_{i=1}^K \mathbf{\Gamma}_i\} \leq P$ (the maximum transmit power constraint at the transmitter) and over all $K!$ permutations, i.e.,

$$\mathcal{C}_{\mathrm{DPC}}(P, \mathbf{H}) \triangleq Conv\left(\cup_{\pi,\mathbf{\Gamma}}\mathbf{R}(\pi, \mathbf{\Gamma})\right),$$

where $Conv(\cdot)$ represents the convex hull.

B. Routing

In this paper, the topology of a MIMO-based ad hoc network is represented by a directed graph, denoted by $\mathcal{G} = \{\mathcal{N}, \mathcal{L}\}$, where $\mathcal{N}$ and $\mathcal{L}$ are the set of nodes and all possible MIMO-based links, respectively. By saying "possible" we mean the distance between a pair of nodes is less than or equal to the maximum transmission range $D_{\max}$, i.e., $\mathcal{L} = \{(i, j) : D_{ij} \leq D_{\max}, i, j \in \mathcal{N}, i \neq j\}$, where D_{ij} represents the distance between node i and node j. $D_{\max}$ can be determined by a node's maximum transmission power. Without loss of generality, we assume that $\mathcal{G}$ is always connected. Suppose that the cardinalities of the sets $\mathcal{N}$ and $\mathcal{L}$ are $|\mathcal{N}| = N$ and $|\mathcal{L}| = L$, respectively. For convenience, we index the links numerically (e.g., link $1, 2, \ldots, L$) rather than using node pairs (i, j).

The network topology of $\mathcal{G}$ can be represented by a *node-arc incidence matrix* (NAIM) [5] $\mathbf{A} \in \mathbb{R}^{N\times L}$, whose entry a_{nl} associating with node n and arc l is defined as

$$a_{nl} = \begin{cases} 1 & \text{if } n \text{ is the transmitting node of arc } l \\ -1 & \text{if } n \text{ is the receiving node of arc } l \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

We define $\mathcal{O}(n)$ and $\mathcal{I}(n)$ as the sets of links that are outgoing from and incoming to node n, respectively. We use a multicommodity flow model for the routing of data packets across the network. In this model, several nodes send independent data to their corresponding destinations, possibly through *multipath* and *multihop* routing. We assume that flow conservation at each node is satisfied, i.e., the network is a flow-balanced system.

Suppose that there is a total of F sessions in the network, representing F different commodities. The source and destination nodes of session f, $1 \leq f \leq F$, are denoted as $\mathrm{src}(f)$ and $\mathrm{dst}(f)$, respectively. For the supply and demand of each session, we define a *source-sink vector* $\mathbf{s}_f \in \mathbb{R}^N$, whose entries, other than at the positions of $\mathrm{src}(f)$ and $\mathrm{dst}(f)$, are all zeros. In addition, for flow conservation, we have $(\mathbf{s}_f)_{\mathrm{src}(f)} = -(\mathbf{s}_f)_{\mathrm{dst}(f)}$. Without loss of generality, we let $(\mathbf{s}_f)_{\mathrm{src}(f)} \geq 0$ and simply denote it as a scalar s_f. Therefore, we can further write the source-sink vector of flow f as

$$\mathbf{s}_f = s_f \left[\begin{array}{ccccc} \cdots & 1 & \cdots & -1 & \cdots \end{array}\right]^T, \quad (3)$$

where the dots represent zeros, and 1 and -1 are in the positions of $\mathrm{src}(f)$ and $\mathrm{dst}(f)$, respectively. Note that 1 does not necessarily appear before -1 as in (3), which is only for an illustrative purpose. Using the notation "$=_{x,y}$" to represent the component-wise equality of a vector except at the x^{th} and the y^{th} entries, we have $\mathbf{s}_f =_{\mathrm{src}(f),\mathrm{dst}(f)} \mathbf{0}$. In addition, using the matrix $\mathbf{S} \triangleq \left[\begin{array}{cccc} \mathbf{s}_1 & \mathbf{s}_2 & \ldots & \mathbf{s}_F \end{array}\right] \in \mathbb{R}^{N\times F}$ to denote the collection of all source-sink vectors, we further have

$$\mathbf{S}\mathbf{e}_f =_{\mathrm{src}(f),\mathrm{dst}(f)} \mathbf{0}, \quad 1 \leq f \leq F, \quad (4)$$

$$\langle \mathbf{1}, \mathbf{S}\mathbf{e}_f\rangle = 0, \quad 1 \leq f \leq F, \quad (5)$$

$$(\mathbf{S}\mathbf{e}_f)_{\mathrm{src}(f)} = s_f, \quad 1 \leq f \leq F, \quad (6)$$

where $\mathbf{e}_f$ is the f^{th} unit column vector.

Denote $t_l^{(f)} \geq 0$ be the amount of flow of session f on link l. Define $\mathbf{t}^{(f)} \in \mathbb{R}^L$ the *flow vector* for session f. At node n, components of the flow vector and source-sink vector for the same commodity satisfy the flow conservation as follows: $\sum_{l\in\mathcal{O}(n)} t_l^{(f)} - \sum_{l\in\mathcal{I}(n)} t_l^{(f)} = (\mathbf{s}_f)_n$, $1 \leq n \leq N$, $1 \leq f \leq F$. With NAIM, the flow conservation for the entire network can be written as $\mathbf{A}\mathbf{t}^{(f)} = \mathbf{s}_f$, $\ 1 \leq f \leq F$. Denote matrix $\mathbf{T} \triangleq \left[\begin{array}{cccc} \mathbf{t}^{(1)} & \mathbf{t}^{(2)} & \ldots & \mathbf{t}^{(F)} \end{array}\right] \in \mathbb{R}^{L\times F}$ the collection of all flow vectors. With $\mathbf{T}$ and $\mathbf{S}$, the flow conservation can be further written as $\mathbf{A}\mathbf{T} = \mathbf{S}$.

Since the total amount of flow on each link l cannot exceed its capacity limit, we must have $\sum_{f=1}^F t_l^{(f)} \leq R_l(\mathbf{\Gamma})$, $\forall l$, where $R_l(\mathbf{\Gamma}) \in \mathcal{C}_{\mathrm{DPC}}^{(n)}(P_{\max}^{(n)}, \mathbf{H}^{(n)})$ is the DPC rate of link l. This can be further compactly written using matrix-vector notations as $\langle \mathbf{1}, \mathbf{T}^T\mathbf{e}_l\rangle \leq R_l(\mathbf{\Gamma})$, $\forall l$.

C. Problem Formulation

In this paper, we focus on how to jointly optimize routing at the network layer and power allocation at the physical layer. We suppose that each node in the network has been assigned an orthogonal frequency band. There is a vast amount of literature on how to perform channel assignments (see [6] and references therein) and its discussion is beyond the scope of this paper. We adopt the proportional fairness utility function, i.e., $\ln(s_f)$ for flow f. The objective is to maximize the sum of utilities of all sessions. Putting together the physical layer constraints in Section II-A and the network layer constraints in Section II-B, we have the problem formulation as follows:

$$\begin{array}{lll} \text{Maximize} & \sum_{f=1}^{F} \ln(s_f) & \\ \text{subject to} & \mathbf{AT} = \mathbf{S} & \\ & \mathbf{T} \geq \mathbf{0} & \\ & \mathbf{Se}_f =_{\mathrm{src}(f),\mathrm{dst}(f)} \mathbf{0} & \forall f \\ & \langle \mathbf{1}, \mathbf{Se}_f \rangle = 0 & \forall f \\ & (\mathbf{Se}_f)_{\mathrm{src}(f)} = s_f & \forall f \\ & \langle \mathbf{1}, \mathbf{T}^T \mathbf{e}_l \rangle \leq R_l(\mathbf{\Gamma}) & \forall l \\ & R_l(\mathbf{\Gamma}) \in \mathcal{C}_{\mathrm{DPC}}^{(n)}(P_{\max}^{(n)}, \mathbf{H}^{(n)}) & \forall l \in \mathcal{O}(n) \\ & \sum_{l \in \mathcal{O}(n)} \mathrm{Tr}\{\mathbf{\Gamma}_l\} \leq P_{\max}^{(n)} & \forall n \\ & \mathbf{\Gamma}_l \succeq 0 & \forall l \\ & \text{Variables: } \mathbf{S}, \mathbf{T}, \mathbf{\Gamma}, & \end{array} \tag{7}$$

It is evident that (7) is a non-convex optimization problem since DPC rate equation in (1) is a non-convex function. However, in what follows, we will show that (7) can be reformulated as an equivalent convex optimization problem.

D. Reformulation

In this paper, we employ an important concept called *uplink-downlink duality* from network information theory to reformulate problem (7). Due to space limitation, we refer readers to [4] for details on how to construct a dual MIMO multiple access channel (MIMO-MAC) from a MIMO-BC. In essence, uplink-downlink duality theorem says that the DPC rate region of a MIMO-BC channel with maximum power constraint P is equal to the capacity region of its dual MIMO-MAC with the same sum power, i.e., $\mathcal{C}_{\mathrm{DPC}}(P, \mathbf{H}) = \mathcal{C}_{\mathrm{MAC}}(P, \mathbf{H}^\dagger)$. Thus, we can replace $\mathcal{C}_{\mathrm{DPC}}(\cdot)$ in (7) by $\mathcal{C}_{\mathrm{MAC}}(\cdot)$ and problem (7) becomes

$$\begin{array}{lll} \text{Maximize} & \sum_{f=1}^{F} \ln(s_f) & \\ \text{subject to} & \mathbf{AT} = \mathbf{S} & \\ & \mathbf{T} \geq \mathbf{0} & \\ & \mathbf{Se}_f =_{\mathrm{src}(f),\mathrm{dst}(f)} \mathbf{0} & \forall f \\ & \langle \mathbf{1}, \mathbf{Se}_f \rangle = 0 & \forall f \\ & (\mathbf{Se}_f)_{\mathrm{src}(f)} = s_f & \forall f \\ & \langle \mathbf{1}, \mathbf{T}^T \mathbf{e}_l \rangle \leq R_l(\mathbf{Q}) & \forall l \\ & R_l(\mathbf{Q}) \in \mathcal{C}_{\mathrm{MAC}}^{(n)}(P_{\max}^{(n)}, \mathbf{H}^{\dagger(n)}) & \forall l \in \mathcal{O}(n) \\ & \sum_{l \in \mathcal{O}(n)} \mathrm{Tr}\{\mathbf{Q}_l\} \leq P_{\max}^{(n)} & \forall n \\ & \mathbf{Q}_l \succeq 0 & \forall l \\ & \text{Variables: } \mathbf{S}, \mathbf{T}, \mathbf{Q} & \end{array} \tag{8}$$

The benefit of these replacements is that $\mathcal{C}_{\mathrm{MAC}}(\cdot)$ is convex with respect to the input covariance matrices $\mathbf{Q}_1, \ldots, \mathbf{Q}_K$ in the dual MIMO-MAC. After solving (8), we can recover the corresponding MIMO-BC covariance matrices $\mathbf{\Gamma}^*$ from the optimal solution $\mathbf{Q}^*$ of (8) by the MAC-to-BC mapping provided in [4].

III. Solution Procedure

Due to the convexity of (8), we can solve it by solving its Lagrangian dual problem. Introducing Lagrangian multipliers u_i to the link capacity coupling constraints $\langle \mathbf{1}, \mathbf{T}^T \mathbf{e}_l \rangle \leq R_l(\mathbf{Q})$, we can write the Lagrangian as

$$\Theta(\mathbf{u}) = \sup_{\mathbf{S},\mathbf{T},\mathbf{Q}} \left\{ L(\mathbf{S},\mathbf{T},\mathbf{Q},\mathbf{u}) | (\mathbf{S},\mathbf{T},\mathbf{Q}) \in \Psi \right\}, \tag{9}$$

where $L(\mathbf{S},\mathbf{T},\mathbf{Q},\mathbf{u}) = \sum_f \ln(s_f) + \sum_l u_l (R_l(\mathbf{Q}) - \langle \mathbf{1}, \mathbf{T}^T \mathbf{e}_l \rangle)$ and Ψ is defined as

$$\Psi \triangleq \left\{ (\mathbf{S},\mathbf{T},\mathbf{Q}) \left| \begin{array}{ll} \mathbf{AT} = \mathbf{S} & \\ \mathbf{T} \geq \mathbf{0} & \\ \mathbf{Se}_f =_{\mathrm{src}(f),\mathrm{dst}(f)} \mathbf{0} & \forall f \\ \langle \mathbf{1}, \mathbf{Se}_f \rangle = 0 & \forall f \\ (\mathbf{Se}_f)_{\mathrm{src}(f)} = s_f & \forall f \\ \sum_{l \in \mathcal{O}(n)} \mathrm{Tr}\{\mathbf{Q}_l\} \leq P_{\max}^{(n)} & \forall n \\ \mathbf{Q}_l \succeq 0 & \forall l \\ R_l(\mathbf{Q}) \in \mathcal{C}_{\mathrm{MAC}}(P_{\max}^{(n)}, \mathbf{H}^{\dagger(n)}) & \forall n \end{array} \right. \right\}.$$

The Lagrangian dual problem of (8) can be written as:

$$\mathbf{D}: \quad \begin{array}{ll} \text{Minimize} & \Theta(\mathbf{u}) \\ \text{subject to} & \mathbf{u} \geq \mathbf{0}. \end{array}$$

For a given $\mathbf{u}$, the Lagrangian in (9) can be rearranged and separated into two parts:

$$\Theta(\mathbf{u}) = \Theta_{\mathrm{net}}(\mathbf{u}) + \Theta_{\mathrm{phy}}(\mathbf{u}),$$

where, for a given Lagrangian multiplier $\mathbf{u}$, Θ_{net} and Θ_{phy} correspond to network layer and physical layer variables, respectively:

$$\begin{array}{rll} \Theta_{\mathrm{net}}(\mathbf{u}) \triangleq & \text{Maximize} \sum_f \ln(s_f) & \\ & \quad - \sum_l u_l \langle \mathbf{1}, \mathbf{T}^T \mathbf{e}_l \rangle & \\ \text{subject to} & \mathbf{AT} = \mathbf{S} & \\ & \mathbf{T} \geq \mathbf{0} & \\ & \mathbf{Se}_f =_{\mathrm{src}(f),\mathrm{dst}(f)} \mathbf{0} & \forall f \\ & \langle \mathbf{1}, \mathbf{Se}_f \rangle = 0 & \forall f \\ & (\mathbf{Se}_f)_{\mathrm{src}(f)} = s_f & \forall f \\ \text{Variables:} & \mathbf{S}, \mathbf{T} & \end{array}$$

$$\begin{array}{rll} \Theta_{\mathrm{phy}}(\mathbf{u}) \triangleq & \text{Maximize} \sum_l u_l R_l(\mathbf{Q}) & \\ \text{subject to} & \sum_{l \in \mathcal{O}(n)} \mathrm{Tr}\{\mathbf{Q}_l\} \leq P_{\max}^{(n)} & \forall n \\ & \mathbf{Q}_l \succeq 0 & \forall l \\ & R_l(\mathbf{Q}) \in \mathcal{C}_{\mathrm{MAC}}(P_{\max}^{(n)}, \mathbf{H}^{\dagger(n)}), & \\ & \quad \forall l \in \mathcal{O}(n), n \in N & \\ \text{Variables:} & \mathbf{Q} & \end{array}$$

The Lagrangian dual problem can thus be written as the following master dual problem:

$$\mathbf{MD}^{\mathrm{CRPA-E}}: \quad \begin{array}{ll} \text{Minimize} & \Theta_{\mathrm{net}}(\mathbf{u}) + \Theta_{\mathrm{phy}}(\mathbf{u}) \\ \text{subject to} & \mathbf{u} \geq \mathbf{0} \end{array}$$

From the structure of $\Theta_{\mathrm{net}}(\mathbf{u})$, it is clear that it can be readily solved by many polynomial time convex programming methods. However, solving $\Theta_{\mathrm{phy}}(\mathbf{u})$ is substantially more challenging. Note that $\Theta_{\mathrm{phy}}(\mathbf{u})$ can be further decomposed on a node-by-node basis as follows:

$$\Theta_{\mathrm{phy}}(\mathbf{u}) = \max \sum_l u_l R_l(\mathbf{Q})$$

$$= \sum_{n=1}^{N} \left(\max \sum_{l \in \mathcal{O}(n)} u_l R_l(\mathbf{Q}) \right) = \sum_{n=1}^{N} \Theta_{\mathrm{phy}}^{(n)}(\mathbf{u}^{(n)}). \tag{10}$$

We can see that $\Theta_{\text{phy}}^{(n)}(\mathbf{u}^{(n)}) \triangleq \max\sum_{l\in\mathcal{O}(n)} u_l R_l(\mathbf{Q})$ is a maximum weighted sum rate problem of the dual MIMO-MAC for some given dual variables $\mathbf{u}^{(n)}$ as weights. Without loss of generality, suppose that node n has K outgoing links, which are indexed as $1,\ldots,K$ and are associated with dual variables $u_1,\ldots,u_K$, respectively.

Now, we consider the maximum weighted sum rate problem of the dual MIMO-MAC. It can be shown that we do not have to enumerate all $K!$ successive decoding order when computing the maximum sum rate of the dual MIMO-MAC. Instead, the maximum sum rate problem can be simplified as follows:

Theorem 1. *Suppose that each link rate R_i in MIMO-MAC is associated with a non-negative weight u_i, $i=1,\ldots,K$. The maximum weighted sum rate $\max\sum_{i=1}^{K} u_i R_i(\mathbf{Q})$ can be solved by the following convex optimization problem:*

$$\begin{array}{ll}\text{Maximize} & \sum_{i=1}^{K}(u_{\pi(i)}-u_{\pi(i-1)})\times \\ & \log\left|\mathbf{I}+\sum_{j=i}^{K}\rho_{\pi(j)}\mathbf{H}_{\pi(j)}^{\dagger}\mathbf{Q}_{\pi(j)}\mathbf{H}_{\pi(j)}\right| \\ \text{subject to} & \sum_{i=1}^{K}\mathrm{Tr}(\mathbf{Q}_i)\le P_{\max} \\ & \mathbf{Q}_i \succeq 0,\ i=1,\ldots,K,\end{array} \tag{11}$$

where $u_{\pi(0)} \triangleq 0$, $\pi(i), i=1,\ldots,K$ is a permutation on $\{1,\ldots,K\}$ such that $u_{\pi(1)}\le\ldots\le u_{\pi(K)}$.

Due to space limitation, we refer readers to [7] for the proof of Theorem 1.

Although $\Theta_{\text{phy}}^{(n)}(\mathbf{u}^{(n)})$ is a convex problem, generic convex programming methods are not efficient because of the complex structure of the objective function and constraints in (11). In the next subsection, we propose a custom-designed method to solve $\Theta_{\text{phy}}^{(n)}(\mathbf{u}^{(n)})$.

A. Solving the Physical Layer Subproblem

Our proposed method is based on conjugate gradient projection (CGP). CGP utilizes an important concept called Hessian conjugate, which deflects the gradient direction appropriately to achieve an asymptotic superlinear convergence rate [8]. In each iteration, CGP projects the conjugate gradient direction to find an improving feasible direction. The framework of CGP for solving (11) is shown in Algorithm 1. Due to the complexity of the objective function, performing an exact line search is onerous as it calls for excessive evaluations of the objective function. Therefore, we adopt the "Armijo rule" inexact line search method [8], which offers provable convergence. For convenience, we use $F(\mathbf{Q})$ to represent the objective function in (11), where $\mathbf{Q}=(\mathbf{Q}_1,\ldots,\mathbf{Q}_K)$ denotes the set of covariance matrices at a node. In our numerical study, we choose $s_k=1$. According to Armijo Rule, in the k^{th} iteration, α_k can be computed as $\alpha_k=\beta^{m_k}$, where $0<\beta<1$ is a fixed scalar, and m_k is the first non-negative integer m that satisfies

Algorithm 1 Conjugate Gradient Projection Method

Initialization:
Choose the initial conditions $\mathbf{Q}^{(0)}=[\mathbf{Q}_1^{(0)},\mathbf{Q}_2^{(0)},\ldots,\mathbf{Q}_K^{(0)}]^T$. Let $k=0$.

Main Loop:

1. Calculate the conjugate gradients $\mathbf{G}_i^{(k)}$, $i=1,2,\ldots,K$.
2. Choose an appropriate step size s_k. Let $\mathbf{Q}_i^{'(k)}=\mathbf{Q}_i^{(k)}+s_k\mathbf{G}_i^{(k)}$, for $i=1,2,\ldots,K$.
3. Let $\bar{\mathbf{Q}}^{(k)}$ be the projection of $\mathbf{Q}^{'(k)}$ onto $\Omega_+(P_{\max}^{(n)})$.
4. Choose an appropriate step size α_k. Let $\mathbf{Q}_i^{(k+1)}=\mathbf{Q}_i^{(k)}+\alpha_k(\bar{\mathbf{Q}}_i^{(k)}-\mathbf{Q}_i^{(k)})$, $i=1,2,\ldots,K$.
5. $k=k+1$. If the maximum absolute value of the elements in $\mathbf{Q}_i^{(k)}-\mathbf{Q}_i^{(k-1)}<\epsilon$, for $i=1,2,\ldots,L$, then stop; else go to Step 1.

$$\begin{aligned}&F(\mathbf{Q}^{(k+1)})-F(\mathbf{Q}^{(k)})\ge\sigma\beta^m\langle\mathbf{G}^{(k)},\bar{\mathbf{Q}}^{(k)}-\mathbf{Q}^{(k)}\rangle\\&=\sigma\beta^m\sum_{i=1}^{K}\mathrm{Tr}\left[\mathbf{G}_i^{\dagger(k)}\left(\bar{\mathbf{Q}}_i^{(k)}-\mathbf{Q}_i^{(k)}\right)\right],\end{aligned} \tag{12}$$

where $0<\sigma<1$ is a fixed scalar. Since $\frac{\partial\ln|\mathbf{A}+\mathbf{BXC}|}{\partial\mathbf{X}}=\left[\mathbf{C}(\mathbf{A}+\mathbf{BXC})^{-1}\mathbf{B}\right]^T$ [9], [10], we have

$$\begin{aligned}\bar{\mathbf{G}}_{\pi(j)}=2\rho_{\pi(j)}\mathbf{H}_{\pi(j)}\Bigg[\sum_{i=1}^{j}\left(u_{\pi(i)}-u_{\pi(i-1)}\right)\times\\ \left(\mathbf{I}+\sum_{k=i}^{K}\rho_{\pi(k)}\mathbf{H}_{\pi(k)}^{\dagger}\mathbf{Q}_{\pi(k)}\mathbf{H}_{\pi(k)}\right)^{-1}\Bigg]\mathbf{H}_{\pi(j)}^{\dagger}.\end{aligned} \tag{13}$$

The conjugate gradient direction in the m^{th} iteration can be computed as $\mathbf{G}_{\pi(j)}^{(m)}=\bar{\mathbf{G}}_{\pi(i)}^{(m)}+\kappa_m\mathbf{G}_{\pi(i)}^{(m-1)}$. We adopt the Fletcher and Reeves' choice of deflection [8], which can be computed as

$$\kappa_m=\frac{\|\bar{\mathbf{G}}_{\pi(j)}^{(m)}\|^2}{\|\bar{\mathbf{G}}_{\pi(j)}^{(m-1)}\|^2}. \tag{14}$$

The purpose of deflecting the gradient using (14) is to find the Hessian-conjugate direction that tends to reduce the "zigzagging" phenomenon encountered in the conventional gradient projection method; achieve an asymptotic K-step superlinear convergence rate under certain regulation conditions [8]; and without actually storing a large Hessian approximation matrix as in quasi-Newton methods.

Noting that $\mathbf{G}_{\pi(j)}$ is Hermitian, we have that $\mathbf{Q}_{\pi(j)}^{'(k)}=\mathbf{Q}_{\pi(j)}^{(k)}+s_k\mathbf{G}_{\pi(j)}^{(k)}$ is Hermitian as well. Then, the projection problem becomes how to simultaneously project K Hermitian matrices onto the set $\Omega_+(P_{\max})\triangleq\{\mathbf{Q}_l:\sum_l\mathrm{Tr}\{\mathbf{Q}_l\}\le P_{\max},\mathbf{Q}_l\succeq0,\ l=1,\ldots,K\}$. We construct a block diagonal matrix $\mathbf{D}=\mathrm{Diag}\{\mathbf{Q}_{\pi(1)}\ldots\mathbf{Q}_{\pi(K)}\}\in\mathbb{C}^{(K\cdot n_r)\times(K\cdot n_r)}$. It is easy to recognize that $\mathbf{Q}_{\pi(j)}\in\Omega_+(P_{\max})$, $j=1,\ldots,K$, if and only if $\mathrm{Tr}(\mathbf{D})=\sum_{j=1}^{K}\mathrm{Tr}\left(\mathbf{Q}_{\pi(j)}\right)\le P_{\max}$ and $\mathbf{D}\succeq0$. In our projection, given a block diagonal matrix $\mathbf{D}_n$, we wish to find a matrix $\tilde{\mathbf{D}}_n\in\Omega_+(P_{\max})$ such that $\tilde{\mathbf{D}}_n$ minimizes $\|\tilde{\mathbf{D}}_n-\mathbf{D}_n\|_F$, where $\|\cdot\|_F$ denotes Frobenius norm. For more convenient algebraic manipulations, we instead study the following equivalent optimization problem:

$$\begin{array}{ll}\text{Minimize} & \frac{1}{2}\|\tilde{\mathbf{D}}-\mathbf{D}\|_F^2\\ \text{subject to} & \mathrm{Tr}(\tilde{\mathbf{D}})\le P_{\max},\ \tilde{\mathbf{D}}\succeq0.\end{array} \tag{15}$$

Note that this problem is a convex minimization problem and we can solve this minimization problem by solving its Lagrangian dual. Associating Hermitian matrix $\mathbf{\Pi}$ to the constraint $\tilde{\mathbf{D}} \succeq 0$ and μ to the constraint $\mathrm{Tr}(\tilde{\mathbf{D}}) \le P_{\max}$, we can write the Lagrangian as $g(\mathbf{\Pi}, \mu) = \min_{\tilde{\mathbf{D}}}\{(1/2)\|\tilde{\mathbf{D}} - \mathbf{D}\|_F^2 - \mathrm{Tr}(\mathbf{\Pi}^\dagger \tilde{\mathbf{D}}) + \mu(\mathrm{Tr}(\tilde{\mathbf{D}}) - P_{\max})\}$.

Since $g(\mathbf{\Pi}, \mu)$ is an unconstrained convex quadratic minimization problem, we can compute the minimizer of the Lagrangian by simply setting its first derivative (with respect to $\tilde{\mathbf{D}}$) to zero, i.e., $(\tilde{\mathbf{D}} - \mathbf{D}) - \mathbf{\Pi}^\dagger + \mu \mathbf{I} = 0$. Noting that $\mathbf{\Pi}^\dagger = \mathbf{\Pi}$, we have $\tilde{\mathbf{D}} = \mathbf{D} - \mu\mathbf{I} + \mathbf{\Pi}$. Substituting $\tilde{\mathbf{D}}$ back into the Lagrangian and after some algebraic simplifications, we can rewrite the Lagrangian dual problem as

$$\begin{array}{ll} \text{Maximize} & -\frac{1}{2}\|\mathbf{D} - \mu\mathbf{I} + \mathbf{\Pi}\|_F^2 - \mu P_{\max} + \frac{1}{2}\|\mathbf{D}\|^2 \\ \text{subject to} & \mathbf{\Pi} \succeq 0, \mu \ge 0. \end{array} \quad (16)$$

Eq. (16) belongs the class of so-called *matrix nearness problems*, which are not easy to solve in general (see [11] and references therein). However, based on the special structure in (16), we are able to design a polynomial time algorithm to solve (16). Due to space limitation, we only give the pseudocode in Algorithm 2 and refer readers to [12] for more details.

Algorithm 2 Positive Semidefinite Cone Projection

Initiation:
1. Construct a block diagonal matrix $\mathbf{D}$. Perform eigenvalue decomposition $\mathbf{D} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\dagger$, sort the eigenvalues in non-increasing order.
2. Let $\lambda_0 = \infty$ and $\lambda_{K\cdot n_t+1} = -\infty$. Let $\hat{I} = 0$. Let the endpoint objective value $\psi_{\hat{I}}(\lambda_0) = 0$, $\phi^* = \psi_{\hat{I}}(\lambda_0)$, and $\mu^* = \lambda_0$.

Main Loop:
1. If $\hat{I} > K \cdot n_r$, go to the Final Step; else let $\mu^*_{\hat{I}} = (\sum_{j=1}^{\hat{I}} \lambda_j - P)/\hat{I}$.
2. If $\mu^*_{\hat{I}} \in [\lambda_{\hat{I}+1}, \lambda_{\hat{I}}] \cap \mathbb{R}_+$, then let $\mu^* = \mu^*_{\hat{I}}$ and go to the final step.
3. Compute $\psi_{\hat{I}}(\lambda_{\hat{I}+1})$. If $\psi_{\hat{I}}(\lambda_{\hat{I}+1}) < \phi^*$, then go to the final step; else let $\mu^* = \lambda_{\hat{I}+1}$, $\phi^* = \psi_{\hat{I}}(\lambda_{\hat{I}+1})$, $\hat{I} = \hat{I} + 1$ and continue.

Final Step: Compute $\tilde{\mathbf{D}}$ as $\tilde{\mathbf{D}} = \mathbf{U}\left(\mathbf{\Lambda} - \mu^*\mathbf{I}\right)_+ \mathbf{U}^\dagger$.

B. Solving the Master Dual Problem

In this paper, we propose a cutting-plane algorithm based on outer-linearization to solve the master dual problem. Compared to the widely-used subgradient approach, the attractive feature of the cutting-plane method is its robustness, efficiency, and simplicity in recovering optimal primal feasible solutions. We briefly introduce the basic idea of cutting-plane method as follows. Letting $z = \Theta(\mathbf{u})$, the dual problem is equivalent to

$$\begin{array}{ll} \text{Minimize} & z \\ \text{subject to} & z \ge \sum_f \ln(s_f) + \sum_l u_l \left(R_l(\mathbf{Q}) - \langle \mathbf{1}, \mathbf{T}^T \mathbf{e}_l \rangle\right) \\ & \mathbf{u} \ge 0, \end{array} \quad (17)$$

where $(\mathbf{S}, \mathbf{T}, \mathbf{Q}) \in \Psi$. Although (17) is a linear program with infinite number of constraints not known explicitly, we can consider the following *approximating* problem:

$$\begin{array}{ll} \text{Minimize} & z \\ \text{subject to} & z \ge \sum_f \ln(s_f^{(j)}) + \sum_l u_l \left(R_l(\mathbf{Q}^{(j)}) - \right. \\ & \left. \langle \mathbf{1}, \mathbf{T}^{(j)T} \mathbf{e}_l \rangle\right) \\ & \mathbf{u} \ge 0, \end{array} \quad (18)$$

where the points $(\mathbf{S}^{(j)}, \mathbf{T}^{(j)}, \mathbf{Q}^{(j)}) \in \Psi$, $j = 1, \dots, k-1$. The problem in (18) is a linear program with a finite number of constraints and can be solved efficiently. Let $(z^{(k)}, \mathbf{u}^{(k)})$ be an optimal solution to the approximating problem, which we refer to as the *master program*. If the solution is feasible to (17), then it is an optimal solution to the Lagrangian dual problem. To check the feasibility, we consider the following *subproblem*:

$$\begin{array}{ll} \text{Maximize} & \sum_f \ln(s_f) + \sum_l u_l^{(k)} \left(R_l(\mathbf{Q}) - \langle \mathbf{1}, \mathbf{T}^T \mathbf{e}_l \rangle\right) \\ \text{subject to} & (\mathbf{S}, \mathbf{T}, \mathbf{Q}) \in \Psi. \end{array} \quad (19)$$

Suppose that $(\mathbf{S}^{(k)}, \mathbf{T}^{(k)}, \mathbf{Q}^{(k)})$ is an optimal solution to the subproblem (19) and $\Theta^*(\mathbf{u}^{(k)})$ is the corresponding optimal objective value. If $z_k \ge \Theta^*(\mathbf{u}^{(k)})$, then $\mathbf{u}^{(k)}$ is an optimal solution to the Lagrangian dual problem. Otherwise, for $\mathbf{u} = \mathbf{u}^{(k)}$, the inequality constraint in (17) is not satisfied for $(\mathbf{S}^{(j)}, \mathbf{T}^{(j)}, \mathbf{Q}^{(j)})$. Thus, we can add the constraint

$$z \ge \sum_f \ln\left(s_f^{(k)}\right) + \sum_l u_l \left(R_l(\mathbf{Q}^{(k)}) - \langle \mathbf{1}, \mathbf{T}^{(k)T} \mathbf{e}_l \rangle\right) \quad (20)$$

to (18), and solve the master linear program again. Obviously, $(z^{(k)}, \mathbf{u}^{(k)})$ violates (20) and will be cut off by (20). We summarize the cutting plane algorithm in Algorithm 3.

Algorithm 3 Cutting Plane Algorithm for Solving $\mathbf{D}^{\mathrm{CRPA}}$

Initialization:
Find a point $(\mathbf{S}^{(0)}, \mathbf{T}^{(0)}, \mathbf{Q}^{(0)}) \in \Psi$. Let $k = 1$.

Main Loop:
1. Solve the master program in (18). Let $(z^{(k)}, \mathbf{u}^{(k)})$ be an optimal solution.
2. Solve the subproblem in (19). Let $(\mathbf{S}^{(k)}, \mathbf{T}^{(k)}, \mathbf{Q}^{(k)})$ be an optimal point, and let $\Theta^*(\mathbf{u}^{(k)})$ be the corresponding optimal objective value.
3. If $z^{(k)} \ge \Theta(\mathbf{u}^{(k)})$, then stop with $\mathbf{u}^{(k)}$ as the optimal dual solution. Otherwise, add the constraint (20) to the master program, replace k by $k+1$, and go to step 1.

IV. Numerical Results

In this section, we present some numerical results through simulations to provide further insights. As shown in Fig. 1, we have 15 nodes uniformly distributed in a square region of 1200m $\times$ 1200m. Each node is equipped with two antennas. There are three sessions in the network: node 14 to node 1, node 6 to node 10, and node 5 to node 4.

The convergence process for the cutting-plane method is illustrated in Fig. 2. The optimal objective value for this 15-node example with DPC is 6.72. The optimal flows for sessions N14 to N1, N6 to N10, and N5 to N4 are 9.17 bps/Hz, 9.30 bps/Hz, and 9.93 bps/Hz, respectively. It can be observed that the cutting-plane algorithm efficiently converges with approximately 160 cuts. As expected, the duality gap is zero because the convexity of the transformed equivalent problem based on dual MIMO-MAC.

For comparison, we plot the convergence process of the subgradient approach for the same 15-node example in Fig. 3. The step size selection is $\lambda_k = 0.1/k$. The subgradient method also achieves the same optimal solution and objective value

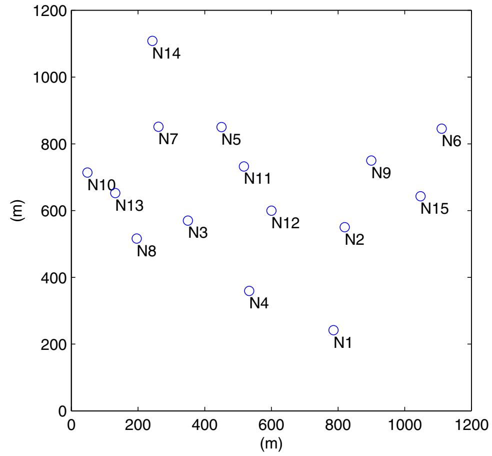

Fig. 1. Topology of a 15-node example network.

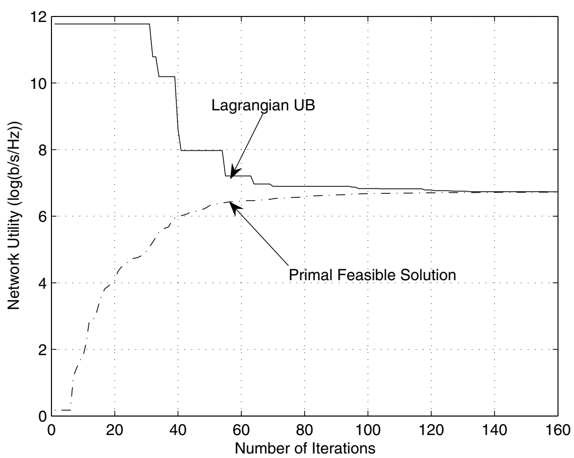

Fig. 2. Convergence behavior of cutting-plane method.

when it converges. However, it is seen that the subgradient algorithm takes approximately 1600 iterations to converge, which is much slower than the cutting-plane method. This is due to the heuristic nature in step size selection (cannot be too large or too small at each step).

V. CONCLUSION

In this paper, we investigated the problem of cross-layer optimization of multi-path routing and power allocation for MIMO-based ad hoc networks with dirty paper coding (DPC). We developed a mathematical solution procedure, which integrates Lagrangian decomposition, gradient projection, and cutting-plane methods. We provided theoretical insights for our proposed solution and gave some numerical results. The decomposable structure of the Lagrangian dual problem and the efficiency of the CGP algorithm make our solution an attractive approach for optimizing MIMO-based ad hoc networks with DPC.

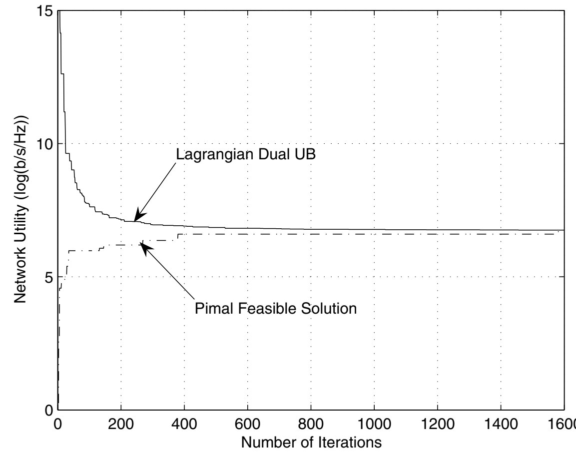

Fig. 3. Convergence behavior of the subgradient method

ACKNOWLEDGEMENTS

The work of Y.T. Hou and J. Liu has been supported in part by the National Science Foundation (NSF) under Grant CNS-0721421 and Office of Naval Research (ONR) under Grant N00014-08-1-0084. The work of H.D. Sherali has been supported in part by NSF Grant CMMI-0552676.

REFERENCES

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

[2] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3936–3964, Sep. 2006.

[3] M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[4] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of MIMO broadcast channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.

[5] M. S. Bazaraa, J. J. Jarvis, and H. D. Sherali, *Linear Programming and Network Flows*, 3rd ed. New York: John Wiley & Sons Inc., 2005.

[6] R. Vedantham, S. Kakumanu, S. Lakshmanan, and R. Sivakumar, "Component based channel assignment in single radio, multi-channel ad hoc networks," in *Proc. ACM Mobicom*, Las Angeles, CA, USA, Sep.23-26, 2006, pp. 378–389.

[7] J. Liu, Y. T. Hou, and H. D. Sherali, "Maximum weighted sum rate of MIMO gaussian broadcast channels," *Technical Report, Deptment of ECE, Virginia Tech*, Nov. 2006. [Online]. Available: http://www.ece.vt.edu/thou/Publications/publications.html

[8] 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.

[9] S. Ye and R. S. Blum, "Optimized signaling for MIMO interference systems with feedback," *IEEE Trans. Signal Process.*, vol. 51, no. 11, pp. 2839–2848, Nov. 2003.

[10] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Economics*. New York: Wiley, 1999.

[11] J. Malick, "A dual approach to semidefinite least-squares problems," *SIAM Journal on Matrix Analysis and Applications*, vol. 26, no. 1, pp. 272–284, Sep. 2005.

[12] J. Liu, Y. T. Hou, and H. D. Sherali, "On the maximum weighted sum-rate of MIMO Gaussian broadcast channels," in *Proc. IEEE ICC*, Beijing, China, May19-23, 2008.