Wireless MIMO Switching: Sum Rate Optimization

Fanggang Wang*, Xiaojun Yuan†, Soung Chang Liew†, Dongning Guo‡
*State Key Lab of Rail Traffic Control and Safety, EIE, Beijing Jiaotong Univ., Beijing, China
†Institute of Network Coding, The Chinese Univ. of Hong Kong, Hong Kong, China
‡Department of Electrical Engineering & Computer Science, Northwestern Univ., Evanston, IL, USA

Abstract— **This paper addresses relay design for a wireless multiple-input-multiple-output (MIMO) switching scheme that enables data exchange among multiple users. Here, a multi-antenna relay linearly precodes the received (uplink) signals from multiple users before forwarding the signal in the downlink, where the purpose of precoding is to let each user receive its desired signal with interference from other users suppressed. The problem of optimizing the precoder based on sum-rate maximization criteria is typically non-convex and difficult to solve. The main contribution of this paper is that we show the sum-rate maximization problem can be converted to an equivalent weighted sum-MSE minimization problem and can therefore be solved using an iterative algorithm proposed in our previous work. Asymptotic analysis reveals that, with properly chosen initial values, the proposed iterative algorithms are asymptotically optimal in both high and low signal-to-noise-ratio (SNR) regimes for MIMO switching, either with or without self-interference cancellation (a.k.a., physical-layer network coding). Numerical results show that the optimized MIMO switching scheme based on the proposed algorithms significantly outperforms existing approaches in the literature.**

Index Terms—**Beamforming, linear precoding, MIMO switching, minimum mean square error (MMSE), physical-layer network coding, relay.**

I. Introduction

Physical-layer network coding (PNC) has received much attention in recent years [1]. The simplest communication model for PNC is a two-way relay channel, in which two users accomplish bidirectional data exchange in two phases of transmission with the help of a relay. Significant progress has been made in approaching the ultimate capacity limit of two-way relay channel (see [1]–[6], [24] and the references therein).

More recently, multi-way relaying, in which multiple users exchange data via a single relay, has been studied [7]–[13]. In [7], the authors studied such a system where the relay is equipped with a single antenna. The use of multiple antennas at the relay provides extra spatial degrees of freedom that can boost throughput significantly. A multi-antenna relay that performs one-to-one mapping from the inputs to the outputs (i.e., that switches traffic in a one-to-one manner among the end users) is called a MIMO switch [8], [9]. Various traffic patterns have been studied in MIMO relaying, including pairwise data exchange [9], [10], [14], [15], where the users form pairs and data exchange is within each pair, and full data exchange, where each user broadcasts to all other users [11], [12], [16]. The authors in [8] further generalized pairwise data exchange to arbitrary unicast, in which each user sends data to one other user and could receive data from a different user. Arbitrary unicast is interesting because we can realize any traffic pattern, including unicast, multicast, broadcast, or any mixture of them, by scheduling a sequence of such unicast flows.

This paper is concerned with relay design for MIMO switching with arbitrary simutaneous unicast. Several pioneering works in this direction have been reported in the literature [8], [10], [11], [13], [14]. Zero-forcing relaying was first proposed in [10] to realize pairwise data exchange. In [13], the authors showed that zero-forcing relay with PNC, which employs self-interference cancellation, can improve system throughput considerably. However, zero-forcing involves channel inverse operations that could incur significant power penalties when the channel gain matrix is ill-conditioned. To alleviate power penalties, the authors of [10] and [11] proposed minimum-mean-square-error (MMSE) relaying, which achieves better performance in the practical signal-to-noise ratio (SNR) regime. In [17], an MMSE relaying scheme exploiting PNC was proposed to further improve the MSE performance. However, as a performance metric, throughput is arguably more directly related to user experience than MSE. Unfortunately, throughput maximization problems are often non-convex and difficult to solve. In this paper, we investigate a system throughput maximization problem and propose an efficient iterative algorithm to solve the problem.

In this work, the sum-rate maximization problem is converted to an equivalent weighted sum-MSE minimization problem, which admits an iterative solution. In the low and high SNR regimes, analytical results are provided for the properties of the asymptotically optimal solutions and the convergence conditions of the proposed iterative algorithms. Numerical results show that the proposed algorithms significantly outperform existing approaches in the literature [8], [10], [11], [13].

The remainder of the paper is organized as follows: Section II introduces the background of wireless MIMO switching. Sum-rate maximization is discussed in Section III. In Section IV, asymptotically optimal solutions are derived. Section V

This work was partially supported by the National Natural Science Foundation (No. 61201201); the Univ. Grants Committee of the Hong Kong (Nos. AoE/E-02/08, 414911 and 418712); the National Science Foundation (Nos. 0905572 and 1231806); the State Key Lab of Rail Traffic Control and Safety (No. RCS2011ZT011); the Fundamental Research Funds for the Central Universities (Nos. 2011JBM203 and 2010RC007).

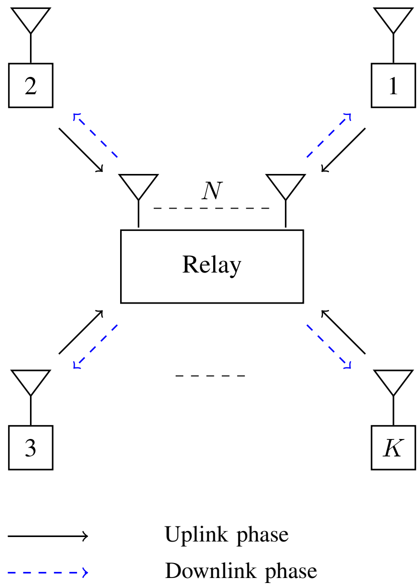

Fig. 1. Wireless MIMO switching.

presents simulation results. Section VI concludes this paper.

We adopt the following notational convention: boldface lower-case letters denote vectors and boldface upper-case letters denote matrices; diag$\{\boldsymbol{x}\}$ denotes a diagonal square matrix $\boldsymbol{X}$ whose diagonal consists of the elements of $\boldsymbol{x}$; diag$\{\boldsymbol{X}\}$ denotes a column vector $\boldsymbol{x}$ formed by the diagonal elements of $\boldsymbol{X}$. $[\boldsymbol{X}]_{\text{diag}}$ represents a diagonal matrix with the same diagonal elements as $\boldsymbol{X}$. Denote by $\mathcal{C}_{a,b}$ the covariance of two zero-mean random variables a and b, i.e., $\mathcal{C}_{a,b} = \mathbb{E}[ab^*]$. The operation $(\cdot)^\dagger$ denotes Moore-Penrose pseudo inverse [18]; and $\otimes$ denotes the Kronecker product.

II. System Description

The overall system is illustrated in Fig. 1. There are K users, numbered from 1 to K, each equipped with a single antenna. These users communicate via a relay with N antennas and there is no direct link between any two users. Throughout the paper, we focus on the pure unicast case, in which each user transmits to one other user only. Let $\pi(\cdot)$ specify a switching pattern, which can be represented as follows: user i transmits to $j = \pi(i)$ for every $i \in \{1, \cdots, K\}$. The pure unicast switching pattern can be equivalently represented by a permutation matrix $\boldsymbol{P}$.[1] Let $\boldsymbol{e}_j$ denote the jth column of an identity matrix. Then the ith column of $\boldsymbol{P}$ is equal to $\boldsymbol{e}_j$ if $\pi(i) = j$, i.e., $\boldsymbol{p}_i = \boldsymbol{e}_{\pi(i)} = \boldsymbol{e}_j$. If the diagonal elements of permutation $\boldsymbol{P}$ are all zero, it is also called a *derangement*. In particular, a symmetric derangement ($\boldsymbol{P} = \boldsymbol{P}^T$) realizes a pairwise data exchange. In general, any traffic flow pattern among the users can be realized by scheduling a set of different unicast traffic flows [8].

Each round of data exchange consists of one uplink phase and one downlink phase. The uplink phase sees simultaneous transmissions from the users to the relay; the downlink phase sees one transmission from the relay to the users. We assume that the two phases are of equal duration.

In the uplink phase, let $\boldsymbol{x} = [x_1, \cdots, x_K]^T$ be the vector representing the signals transmitted by the users. Let $\boldsymbol{y} = [y_1, \cdots, y_N]^T$ be the received signals at the relay's antennas, and $\boldsymbol{u} = [u_1, \cdots, u_N]^T$ be the noise vector with independent and identically distributed (i.i.d.) samples following the circularly-symmetric complex Gaussian (CSCG) distribution, denoted by $\mathcal{CN}(0, \gamma^2)$, where γ^2 is the noise variance at the relay. Then

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{u}, \tag{1}$$

where $\boldsymbol{H} \in \mathbb{C}^{N \times K}$ is the uplink channel matrix. We assume that all uplink signals are independent Gaussian with zero mean and unit average power, i.e., $x_i \sim \mathcal{CN}(0,1)$ and $\mathbb{E}\{x_i x_j\} = 0$, $\forall i, j, i \neq j$.

Upon receiving $\boldsymbol{y}$, the relay precodes $\boldsymbol{y}$ with matrix $\boldsymbol{G}$ and forward $\boldsymbol{G}\boldsymbol{y}$ in the downlink phase, where the transmit power of the relay is upper-bounded by P_r, i.e.

$$\text{Tr}\left[\boldsymbol{G}\left(\boldsymbol{H}\boldsymbol{H}^H + \gamma^2 \boldsymbol{I}\right)\boldsymbol{G}^H\right] \leq P_r. \tag{2}$$

The signals received by all users are collectively represented in the vector form as

$$\boldsymbol{r} = \boldsymbol{F}\boldsymbol{G}\boldsymbol{y} + \boldsymbol{w} = \boldsymbol{F}\boldsymbol{G}\boldsymbol{H}\boldsymbol{x} + \boldsymbol{F}\boldsymbol{G}\boldsymbol{u} + \boldsymbol{v}, \tag{3}$$

where $\boldsymbol{F}$ is the downlink channel matrix, and $\boldsymbol{v}$ is the noise vector at the receivers, with i.i.d. samples following the CSCG distribution, i.e., $v_k \sim \mathcal{CN}(0, \sigma^2)$, where σ^2 is the noise variance. The signal $r_{\pi(i)}$ received by user $\pi(i)$ is used to recover the message from user i. It is not difficult to see that interference-free switching can be achieved in the absence of noise by designing $\boldsymbol{G}$ to ensure $\boldsymbol{P}^T\boldsymbol{F}\boldsymbol{G}\boldsymbol{H}$ is diagonal. In the presence of noise, the MIMO precoder $\boldsymbol{G}$ shall be chosen to manage interference based on the knowledge of the uplink and downlink channels $\boldsymbol{H}$ and $\boldsymbol{F}$. Furthermore, physical-layer network coding techniques can be used if the users can cancel self-interference in the received signal. We refer to the precoding schemes without network coding as *non-PNC schemes*, and subsequent ones with network coding as *PNC schemes*.

III. Sum-Rate Maximization

In this section, we study sum-rate maximization, which is widely used in optimizing the performance of wireless networks. We show that the sum-rate maximization problem can be converted to an equivalent weighted sum-MSE minimization problem, and therefore, the results in [17] can be easily adopted here.

A. Problem Formulation

We now use a diagonal matrix $\boldsymbol{\Delta}$ to represent the weights of self-interference to be canceled. Note that letting $\boldsymbol{\Delta} = \boldsymbol{0}$ yields the conventional non-PNC problem. Then, after self-interference cancellation, the signal becomes $\boldsymbol{r} - \boldsymbol{\Delta}\boldsymbol{x}$. To simplify the index mapping of the received signal vector, let

$$\boldsymbol{z} = \boldsymbol{P}^T(\boldsymbol{r} - \boldsymbol{\Delta}\boldsymbol{x}). \tag{4}$$

[1] A square matrix $\boldsymbol{P}$ is a permutation matrix if it has one and only one nonzero element on each row and each column, which is equal to 1.

Using (3), the ith element of $\boldsymbol{z}$ is given by

$$z_i = \boldsymbol{p}_i^T \boldsymbol{F}\boldsymbol{G}\boldsymbol{h}_i x_i + \sum_{\ell \neq i} \boldsymbol{p}_i^T \boldsymbol{F}\boldsymbol{G}\boldsymbol{h}_\ell x_\ell + \boldsymbol{p}_i^T \boldsymbol{F}\boldsymbol{G}\boldsymbol{u} + \boldsymbol{p}_i^T \boldsymbol{v} - \boldsymbol{p}_i^T \boldsymbol{\Delta}\boldsymbol{x}, \tag{5}$$

where $\boldsymbol{h}_i$ is the ith column of $\boldsymbol{H}$. In this way, z_i is the received signal of user $\pi(i)$ for the recovery of the message from user i. Thus, the achievable rate of user i is given by

$$R_i = \frac{1}{2} \log \Bigg(1 + \frac{|\boldsymbol{p}_i^T \boldsymbol{F}\boldsymbol{G}\boldsymbol{h}_i|^2}{\|\boldsymbol{p}_i^T(\boldsymbol{F}\boldsymbol{G}\boldsymbol{H} - \boldsymbol{\Delta})\|^2 - |\boldsymbol{p}_i^T \boldsymbol{F}\boldsymbol{G}\boldsymbol{h}_i|^2 + \gamma^2 \|\boldsymbol{p}_i^T \boldsymbol{F}\boldsymbol{G}\|^2 + \sigma^2} \Bigg), \tag{6}$$

where the factor $1/2$ is due to the two-phase transmission. Our purpose is to maximize the sum rate under the power constraint of the relay. This optimization problem is formulated as

$$\underset{\boldsymbol{G},\boldsymbol{\Delta}}{\text{maximize}} \quad \sum_{i=1}^{K} R_i \tag{7a}$$

$$\text{subject to} \quad \text{Tr}\left[\boldsymbol{G} \left(\boldsymbol{H}\boldsymbol{H}^H + \gamma^2 \boldsymbol{I} \right) \boldsymbol{G}^H \right] \leq P_r. \tag{7b}$$

The problem (7) is non-convex w.r.t. $(\boldsymbol{G}, \boldsymbol{\Delta})$ and thus is difficult to solve directly.

B. Conversion to Weighted Sum-MSE Minimization

Recall that $x_i \sim \mathcal{CN}(0,1)$, $i = 1, \cdots, K$, i.e.,

$$p(x_i) = \frac{1}{\pi} e^{-|x_i|^2}, \quad i = 1, \cdots, K. \tag{8}$$

From the equivalent transmission (5), the conditional distribution can be readily obtained as

$$p(z_i|x_i) = \frac{1}{\pi \varSigma_i'} e^{-\frac{|x_i - \boldsymbol{p}_i^T \boldsymbol{F}\boldsymbol{G}\boldsymbol{h}_i z_i|^2}{\varSigma_i'}}, \quad i = 1, \cdots, K, \tag{9}$$

where

$$\varSigma_i' = \boldsymbol{p}_i^T \boldsymbol{F}\boldsymbol{G}(\boldsymbol{H}\boldsymbol{H}^H - \boldsymbol{h}_i \boldsymbol{h}_i^H)\boldsymbol{G}^H \boldsymbol{F}^H \boldsymbol{p}_i + \boldsymbol{p}_i^T \boldsymbol{\Delta}\boldsymbol{\Delta}^H \boldsymbol{p}_i + \gamma^2 \boldsymbol{p}_i^T \boldsymbol{F}\boldsymbol{G}\boldsymbol{G}^H \boldsymbol{F}^H \boldsymbol{p}_i + \sigma^2 \boldsymbol{p}_i^T \boldsymbol{p}_i. \tag{10}$$

With Bayes' rule, we can obtain *a posteriori* distribution $p(x_i|z_i)$, which is also Gaussian and is given by

$$p(x_i|z_i) = \frac{1}{\pi \varSigma_i} e^{-\frac{|x_i - \omega_i z_i|^2}{\varSigma_i}}, \quad i = 1, \cdots, K. \tag{11}$$

where w_i is a scaling coefficient to be determined, $\omega_i z_i$ represents the conditional mean, and $\varSigma_i$ represents the conditional variance. According to the definition of mutual information [19], the rate in (6) can be written as

$$R_i = \frac{1}{2} \mathbb{E} \left[\log \frac{p(x_i|z_i)}{p(x_i)} \right], \quad i = 1, \cdots, K, \tag{12}$$

where the expectation is take over the joint distribution of $\boldsymbol{x}$ and $\boldsymbol{z}$, i.e., $p(x_i)p(z_i|x_i)$. With (11) and (8), we can express the sum-rate (7a) as a function of $\boldsymbol{G}$, $\{\omega_i\}$ and $\{\varSigma_i\}$. Similar conversions have been previously used in [20], [21] for optimizing beamforming vectors in broadcast channels. Based on this conversion, we next establish a relation between the sum-rate maximization problem (7) and the weighted sum-MSE minimization problem in [17].

Substituting (11) and (8) into (12), we express the sum rate (7a) as a function of $\boldsymbol{G}$, $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma}$ as

$$\sum_{i=1}^{K} R_i = \frac{1}{2} \sum_{i=1}^{K} \mathbb{E} \bigg(\log \frac{1}{\varSigma_i} - \frac{|x_i - \omega_i z_i|^2}{\varSigma_i} + |x_i|^2 \bigg). \tag{13}$$

Plugging (4) into (13), the sum rate is rewritten in matrix form as

$$\sum_{i=1}^{K} R_i = -\frac{1}{2} \left\{ \mathbb{E} \left\| \boldsymbol{\Sigma}^{-\frac{1}{2}} [(\boldsymbol{P} + \boldsymbol{\Omega}\boldsymbol{\Delta})\boldsymbol{x} - \boldsymbol{\Omega}\boldsymbol{r}] \right\|^2 + \sum_{i=1}^{K} \log \varSigma_i - K \right\}, \tag{14}$$

where

$$\boldsymbol{\Sigma} = \text{diag}\{\varSigma_{\pi^{-1}(1)}, \cdots, \varSigma_{\pi^{-1}(K)}\}, \tag{15a}$$

$$\boldsymbol{\Omega} = \text{diag}\{\omega_{\pi^{-1}(1)}, \cdots, \omega_{\pi^{-1}(K)}\}. \tag{15b}$$

The expectation taken over $p(x_i)p(z_i|x_i)$ is equivalent to that taken over $p(x_i)p(u_{\pi^{-1}(i)})p(v_{\pi^{-1}(i)})$ since the signal and noise terms are independent. In (14), $\mathbb{E}\|\boldsymbol{\Sigma}^{-\frac{1}{2}}((\boldsymbol{P} + \boldsymbol{\Omega}\boldsymbol{\Delta})\boldsymbol{x} - \boldsymbol{\Omega}\boldsymbol{r})\|^2$ is the same as the weighted sum-MSE in [17] by letting

$$\boldsymbol{W} = \boldsymbol{\Sigma}^{-1}, \quad \boldsymbol{B} = \boldsymbol{\Omega}\boldsymbol{\Delta}, \quad \text{and} \quad \boldsymbol{C} = \boldsymbol{\Omega}, \tag{16}$$

where $\boldsymbol{W}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ are well defined in [17], i.e., $\boldsymbol{W}$ is the weights of the sum-MSE minimization problem in [17], $\boldsymbol{B}$ is the diagonal weights of the self-interferences, $\boldsymbol{C}$ is the combination of the scaling factors of the receivers. For fixed scaled $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma}$, the optimization of $\boldsymbol{G}$ is exactly the same as that in [17]. For fixed scaled $\boldsymbol{G}$, the optimal $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma}$ can be determined by the MMSE estimator of the transmission in (5), which will be presented explicitly in the following subsection.

C. Iterative Algorithm

For fixed $\boldsymbol{\Delta}$ and $\boldsymbol{\Sigma}$, the optimization problem with respect to $\boldsymbol{G}$ is convex, and it can be solved numerically with softwares. We define

$$\bar{\boldsymbol{G}} \triangleq \alpha^{-1}\boldsymbol{G}, \quad \bar{\boldsymbol{\Delta}} = \alpha^{-1}\boldsymbol{\Delta}, \quad \text{and} \quad \bar{\boldsymbol{\Omega}} \triangleq \alpha\boldsymbol{\Omega}, \tag{17}$$

where α represents the scaling factor to meet the relay power constraint, which will be further discussed at the end of this section. With (17) and the received signal vector at the users $\boldsymbol{r}$ in (3), the sum rate in (14) is expanded as

$$\mathcal{R}(\bar{\boldsymbol{G}}, \alpha, \bar{\boldsymbol{\Delta}}, \bar{\boldsymbol{\Omega}}, \boldsymbol{\Sigma}) = -\frac{1}{2} \text{Tr} \bigg[\boldsymbol{\Sigma}^{-1} \bigg(\boldsymbol{I} + \bar{\boldsymbol{\Omega}}\bar{\boldsymbol{\Delta}}\bar{\boldsymbol{\Delta}}^H \bar{\boldsymbol{\Omega}}^H - 2\Re\{\bar{\boldsymbol{\Omega}}\boldsymbol{F}\bar{\boldsymbol{G}}\boldsymbol{H}(\boldsymbol{P} + \bar{\boldsymbol{\Omega}}\bar{\boldsymbol{\Delta}})^H\} + \bar{\boldsymbol{\Omega}}\boldsymbol{F}\bar{\boldsymbol{G}}(\boldsymbol{H}\boldsymbol{H}^H + \gamma^2 \boldsymbol{I})\bar{\boldsymbol{G}}^H \boldsymbol{F}^H \bar{\boldsymbol{\Omega}}^H + \sigma^2 \alpha^{-2} \bar{\boldsymbol{\Omega}}\bar{\boldsymbol{\Omega}}^H \bigg) \bigg] - \frac{1}{2} \sum_{i=1}^{K} \varSigma_i + K. \tag{18}$$

Then the problem (7) can be equivalently expressed as

$$\underset{\bar{\boldsymbol{G}},\alpha,\bar{\boldsymbol{\Delta}},\bar{\boldsymbol{\Omega}},\boldsymbol{\Sigma}}{\text{maximize}} \quad \mathcal{R}(\bar{\boldsymbol{G}},\alpha,\bar{\boldsymbol{\Delta}},\bar{\boldsymbol{\Omega}},\boldsymbol{\Sigma}) \tag{19a}$$

$$\text{subject to} \quad \alpha^2 \operatorname{Tr}\left[\bar{\boldsymbol{G}}\left(\boldsymbol{H}\boldsymbol{H}^H+\gamma^2\boldsymbol{I}\right)\bar{\boldsymbol{G}}^H\right] \leq P_r. \tag{19b}$$

Noting the similarity to the sum-MSE minimization problem, we develop an iterative algorithm to solve problem (19) as follows.

1) Optimizing $(\bar{\boldsymbol{G}},\alpha)$ *for fixed* $(\bar{\boldsymbol{\Delta}},\bar{\boldsymbol{\Omega}},\boldsymbol{\Sigma})$*:* For fixed $(\bar{\boldsymbol{\Delta}},\bar{\boldsymbol{\Omega}},\boldsymbol{\Sigma})$, the problem in (19) is the same as the one in [17] by letting $\boldsymbol{W}=\boldsymbol{\Sigma}^{-1}$, $\boldsymbol{B}=\bar{\boldsymbol{\Omega}}\bar{\boldsymbol{\Delta}}$ and $\bar{\boldsymbol{C}}=\bar{\boldsymbol{\Omega}}$ except for some additive constants. Thus, from results in [17], the optimal precoder can be immediately written as $\boldsymbol{G}^{opt}=\alpha\bar{\boldsymbol{G}}^{opt}$ with

$$\bar{\boldsymbol{G}}^{opt} = \left(\frac{\sigma^2}{P_r}\operatorname{Tr}[\boldsymbol{\Sigma}^{-1}\bar{\boldsymbol{\Omega}}\bar{\boldsymbol{\Omega}}^H]\boldsymbol{I}+\boldsymbol{F}^H\bar{\boldsymbol{\Omega}}^H\boldsymbol{\Sigma}^{-1}\bar{\boldsymbol{\Omega}}\boldsymbol{F}\right)^{-1}\times$$
$$\times\boldsymbol{F}^H\bar{\boldsymbol{\Omega}}^H\boldsymbol{\Sigma}^{-1}(\boldsymbol{P}+\bar{\boldsymbol{\Omega}}\bar{\boldsymbol{\Delta}})\boldsymbol{H}^H\left(\boldsymbol{H}\boldsymbol{H}^H+\gamma^2\boldsymbol{I}\right)^{-1} \tag{20a}$$

$$\alpha^{opt} = P_r^{\frac{1}{2}}\left(\operatorname{Tr}\left[\bar{\boldsymbol{G}}^{opt}\left(\boldsymbol{H}\boldsymbol{H}^H+\gamma^2\boldsymbol{I}\right)(\bar{\boldsymbol{G}}^{opt})^H\right]\right)^{\frac{1}{2}}. \tag{20b}$$

2) Optimizing $(\bar{\boldsymbol{\Delta}},\bar{\boldsymbol{\Omega}},\boldsymbol{\Sigma})$ *for fixed* $(\bar{\boldsymbol{G}},\alpha)$*:* For fixed $(\bar{\boldsymbol{G}},\alpha)$, we aim to find the optimal 3-tuple $(\bar{\boldsymbol{\Delta}},\bar{\boldsymbol{\Omega}},\boldsymbol{\Sigma})$ that maximizes $\mathcal{R}(\bar{\boldsymbol{\Delta}},\bar{\boldsymbol{\Omega}},\boldsymbol{\Sigma})$. We first determine the optimal $\bar{\boldsymbol{\Delta}}$. From (16) and (17), we see that (18) is equivalent to the sum-MSE minimizaion problem in [17] by replacing $\boldsymbol{W}$ with $\boldsymbol{\Sigma}^{-1}$, $\boldsymbol{B}$ with $\bar{\boldsymbol{\Omega}}\bar{\boldsymbol{\Delta}}$ and $\bar{\boldsymbol{C}}$ with $\bar{\boldsymbol{\Omega}}$. Together with the fact that the optimal $\boldsymbol{B}$ in [17] is $\boldsymbol{B}^{opt}=[\bar{\boldsymbol{C}}^{opt}\boldsymbol{F}\bar{\boldsymbol{G}}\boldsymbol{H}]_{\text{diag}}$, the optimal $\bar{\boldsymbol{\Delta}}$ is given by

$$\bar{\boldsymbol{\Delta}}^{opt}=[\boldsymbol{F}\bar{\boldsymbol{G}}\boldsymbol{H}]_{\text{diag}}. \tag{21}$$

With (21), we obtain $\boldsymbol{\Delta}^{opt}=\alpha\bar{\boldsymbol{\Delta}}^{opt}=[\boldsymbol{F}\boldsymbol{G}\boldsymbol{H}]_{\text{diag}}$ which consists of the self-interference weights in the received signal $\boldsymbol{r}$. This means that the self-interference is perfectly canceled at the receiver ends. (Since self-interference is known precisely, it is rather obvious it should be completely canceled before detection.)

We next determine the optimal $\bar{\boldsymbol{\Omega}}$ and $\boldsymbol{\Sigma}$. Recall that $\bar{\boldsymbol{\Omega}}$ and $\boldsymbol{\Sigma}$ specify the means and variances of the Gaussian distributions $p(x_i|z_i)$ in (11). Given $(\bar{\boldsymbol{G}},\alpha,\bar{\boldsymbol{\Delta}})$, z_i and x_i are linearly related by noting $\boldsymbol{z}=\boldsymbol{P}^T(\boldsymbol{r}-\boldsymbol{\Delta}\boldsymbol{x})$. Thus, from [22], the *a posteriori* mean and variance are respectively given by

$$\mathbb{E}[x_i|z_i]=\mathcal{C}_{x_iz_i}\mathcal{C}_{z_iz_i}^{-1}z_i, \tag{22a}$$
$$\mathcal{C}_{x_ix_i|z_i}=\mathcal{C}_{x_ix_i}-\mathcal{C}_{x_iz_i}\mathcal{C}_{z_iz_i}^{-1}\mathcal{C}_{z_ix_i}, \tag{22b}$$

where the involved covariances are given by

$$\mathcal{C}_{x_ix_i} = 1, \tag{23a}$$
$$\mathcal{C}_{z_ix_i} = \alpha\boldsymbol{p}_i^T\boldsymbol{F}\bar{\boldsymbol{G}}\boldsymbol{h}_i, \tag{23b}$$
$$\mathcal{C}_{z_iz_i} = \alpha^2\boldsymbol{p}_i^T\boldsymbol{F}\bar{\boldsymbol{G}}\left(\boldsymbol{H}\boldsymbol{H}^H+\gamma^2\boldsymbol{I}\right)\bar{\boldsymbol{G}}^H\boldsymbol{F}^H\boldsymbol{p}_i -\alpha^2\boldsymbol{p}_i^T\bar{\boldsymbol{\Delta}}\boldsymbol{\Delta}^H\boldsymbol{p}_i+\sigma^2. \tag{23c}$$

Therefore, for fixed $(\bar{\boldsymbol{G}},\alpha,\bar{\boldsymbol{\Delta}})$, the optimal $\bar{\boldsymbol{\Omega}}$ and $\boldsymbol{\Sigma}$ can be written as

$$\bar{\omega}_i^{opt}=\alpha\mathcal{C}_{x_iz_i}\mathcal{C}_{z_iz_i}^{-1}, \quad \Sigma_i^{opt}=\mathcal{C}_{x_ix_i|z_i},\ i=1,\cdots,K \tag{24}$$

3) The iterative algorithm: The sum-rate optimization problem (7) can be solved by iteratively solving the above two subproblems. The procedure is outlined in the following algorithm.

Algorithm 1.

1: **Init**: $\bar{\boldsymbol{\Delta}}=\boldsymbol{\Delta}_0$, $\bar{\boldsymbol{\Omega}}=\boldsymbol{\Omega}_0$, $\boldsymbol{\Sigma}=\boldsymbol{\Sigma}_0$;
2: **while** the sum rate can be improved by more than δ **do**
3: Compute $\bar{\boldsymbol{G}}$ and α using (20);
4: Compute $\bar{\boldsymbol{\Delta}}$, $\bar{\boldsymbol{\Omega}}$ and $\boldsymbol{\Sigma}$ using (21) and (24);
5: **end while**

Algorithm 1 converges, as the sum rate is bounded and monotonically increases in the iterative process. The convergence point depends on the initial point $(\bar{\boldsymbol{\Omega}}_0,\boldsymbol{\Sigma}_0,\bar{\boldsymbol{\Delta}}_0)$. We will discuss the initialization in Section IV.

With Algorithm 1, we could obtain one local optimal solution with given initial values. If α is not introduced, we could also iteratively optimize $\boldsymbol{G}$ and $(\boldsymbol{\Delta},\boldsymbol{\Omega},\boldsymbol{\Sigma})$ and achieve a different local optimal solution. Which of the two local optimal solutions is better highly depends on the initial values. In Section IV, we discuss the initial setup of the proposed iterative algorithm. In addition, by introducing the auxiliary variable α, closed-form solutions are available in each iterative step. In contrast, if α is not introduced, numerical calculations are needed when solving convex subproblems.

IV. Asymptotic Analysis

Algorithm 1 only guarantee local optima of the sum-rate maximization problem. In this section, we carry out asymptotic analysis and show that, with proper initialization, the proposed iterative algorithm is asymptotically optimal in the low and high SNR regimes.

A. Low-SNR Analysis

We start with the low-SNR case. For convenience, we focus on the limit where the noise levels σ^2 and γ^2 tend to infinity, i.e., $\sigma^2,\gamma^2\to+\infty$. The main result is summarized as follows: the proof is given in our technical report [23] and is omitted here due to space limit.

Proposition 1. *In the limit of* $\sigma^2,\gamma^2\to+\infty$*, the asymptotically optimal precoders for sum-rate maximization in* (7)*, with and without PNC, are identical and can be expressed as*

$$\boldsymbol{G}^0=\alpha\bar{\boldsymbol{G}}^0, \tag{25}$$

where $\bar{\boldsymbol{G}}^0$ *is such that* $vec(\bar{\boldsymbol{G}}^0)$ *is an eigenvector corresponding to the maximum eigenvalue of*

$$\boldsymbol{\Psi} = \sum_{\ell=1}^{K}(\boldsymbol{h}_\ell\boldsymbol{h}_\ell^H)^T\otimes(\boldsymbol{F}^H\boldsymbol{p}_\ell\boldsymbol{p}_\ell^T\boldsymbol{F}), \tag{26}$$

and scalar α *is such that the precoder* $\boldsymbol{G}^0$ *satisfies the power constraint with equality at the relay.*

At low SNR, the optimal precoder is identical with and without PNC in the limit. This is not surprising because when the noise dominates the received signal, the benefit of self interference cancellation is marginal.

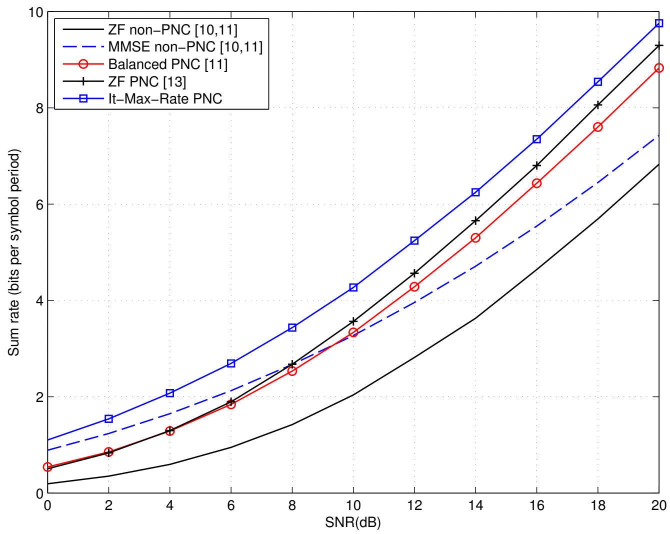

Fig. 2. Throughput comparison of different relaying schemes in the cases of four users.

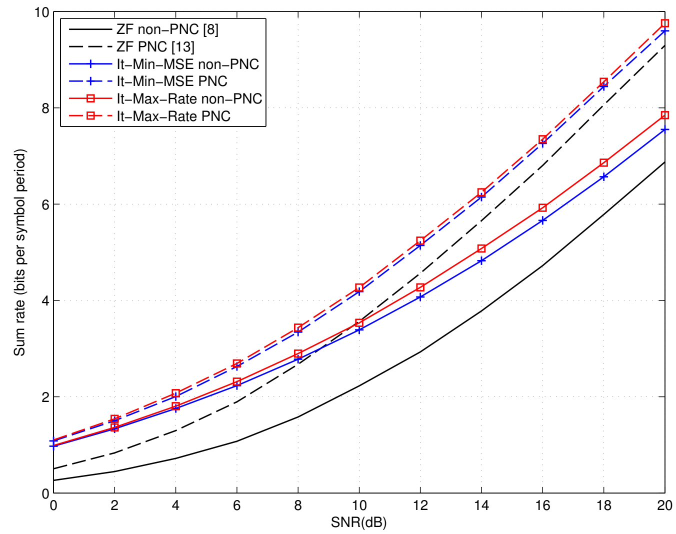

Fig. 3. Throughput comparison of network-coded relaying schemes and their non-PNC relaying schemes in the cases of four users.

Lemma 1. *As $\sigma^2, \gamma^2 \to +\infty$, Algorithms 1 converges to the same global optimum $\boldsymbol{G}^0$ given in (25).*

According to Lemma 1, the point of convergence of Algorithms 1 and is not sensitive to the initial condition in the low SNR regime.

B. High-SNR Analysis

In the high SNR regime, we are interested in the limit of $\sigma^2, \gamma^2 \to 0$. The asymptotically optimal precoders are described as follows, where the proof is given in [23].

Proposition 2. *Suppose $N \geq K$. In the limit $\sigma^2, \gamma^2 \to 0$, the asymptotically optimal precoders for sum-rate maximization, with and without PNC, are identical and can be expressed as*

$$\boldsymbol{G}^\infty = \boldsymbol{F}^\dagger \boldsymbol{C}^{-1}(\boldsymbol{P}+\boldsymbol{B})\boldsymbol{H}^\dagger, \tag{27}$$

where $\boldsymbol{C} \in \mathbb{C}^{K\times K}$ is diagonal, and $\boldsymbol{B} \in \mathbb{C}^{K\times K}$ is an all-zero matrix in the non-PNC case and is a diagonal matrix in the PNC case.

It is obvious that the precoder in (27) forces all the interference to zero, and hence is referred to as the zero-forcing precoder. Proposition 2 reveals that zero-forcing precoding is asymptotically optimal when the relay has no fewer antennas than the number of users ($N \geq K$). Otherwise, the relay does not have enough degree of freedom to force interference seen by all users to be zero. In this case, we may schedule fewer users to allow zero forcing. Then, Proposition 2 can be applied to yield the optimal precoder.

Lemma 2. *Suppose $N \geq K$. As $\sigma^2, \gamma^2 \to 0$, Algorithms 1 converges to $\boldsymbol{G} = \boldsymbol{F}^\dagger \boldsymbol{C}_0^{-1}(\boldsymbol{P}+\boldsymbol{B}_0)\boldsymbol{H}^\dagger$ for any initial values $\boldsymbol{B} = \boldsymbol{B}_0$ and $\boldsymbol{C} = \boldsymbol{C}_0$.*

Lemma 2 suggests that, in the high SNR regime, the convergence point of Algorithms 1 highly depends on the initial conditions. Therefore, it is necessary to carefully choose $\boldsymbol{B}_0$ and $\boldsymbol{C}_0$ in the high SNR regime. At high SNR, the sum-rate maximization problem reduces to a convex problem, which can be easily solved. The details are in [23].

To summarize, our asymptotic analysis reveals that the proposed iteratives in Section III converge to the asymptotically optimal solution in (25) at low SNR, and that this convergence is insensitive to the initial conditions. At high SNR, Algorithms 1 converges to the asymptotically optimal zero-forcing form in (27), but could perform poorly depending on the initial conditions of $\boldsymbol{B}$ and $\boldsymbol{C}$. Therefore, in implementation, we set the initial value of Algorithms 1 to the high-SNR optimal/suboptimal solutions of $\boldsymbol{B}$ and $\boldsymbol{C}$ (cf. Propositions $5-8$ in [23]). The performance of these algorithms are presented in the next section.

V. Numerical Results

We evaluate the sum MSE and the sum rate of the proposed MIMO switching schemes. In our simulations, we assume unit transmit power for each user and the relay, i.e., $P_r = 1$, and the noise levels at the relay and at the users are the same, i.e., $\sigma^2 = \gamma^2$. The channel SNR is then defined as $\text{SNR} = 1/\sigma^2 = 1/\gamma^2$. Numerical results not presented here indicate that the system performance is not sensitive to the switching pattern. Thus, we only present the numerical results for a specific permutation $\boldsymbol{P} = [\boldsymbol{e}_2, \boldsymbol{e}_1, \boldsymbol{e}_4, \boldsymbol{e}_3]$. We assume Rayleigh fading, i.e., the elements of $\boldsymbol{H}$ and $\boldsymbol{F}$ are independently drawn from $\mathcal{CN}(0,1)$. Each simulation point in the presented figures is obtained by averaging over 10^5 random channel realizations.

The key findings are summarized as two observations below.

Observation 1: *The iterative sum-rate maximization (It-Rate-Max) algorithm achieves significant throughput gains over the existing relaying schemes, such as ZF/MMSE relaying [10], [11] and the network-coded relaying [11], [14].*

Fig. 2 illustrates the throughput performance of various approaches including the proposed iterative sum-rate maximization scheme with PNC (It-Max-Rate PNC), the zero-forcing scheme without PNC (ZF non-PNC) in [10], [11], the MMSE scheme without PNC (MMSE non-PNC) in [10], [11], the balanced PNC scheme proposed in [11],[2] and the zero-forcing scheme with PNC (ZF PNC) in [13]. From Fig. 2, the proposed It-Rate-Max PNC algorithm significantly outperforms the other schemes throughout the SNR range of interest. Specifically, the proposed algorithm outperforms the MMSE non-PNC scheme, especially in the high SNR regime, since the former utilizes the PNC technique and jointly optimizes the precoder and the receive filter. The proposed algorithm also outperforms the zero-forcing schemes in [10], [11], [13], since the latters suffer from noise enhancement. Furthermore, we also see that the proposed iterative max-rate scheme has roughly 1.5 dB gain over the balanced PNC scheme in [11] throughout the whole SNR range of interest.

Observation 2: *The PNC schemes achieve considerably higher throughputs than their corresponding non-PNC schemes, especially at medium and high SNR.*

Fig. 3 illustrates the PNC gain for the proposed It-Rate-Max approaches, as well as for the zero-forcing relaying schemes in [8], [13] and the iterative sum-MSE minimization scheme (It-Min-MSE) in [17]. At low SNR, the proposed It-Rate-Max algorithms with and without PNC, exhibit roughly the same throughput performance, which numerically verifies Proposition 1. At high SNR, the proposed schemes with PNC achieve about 6 dB gain over the best non-PNC schemes (i.e., the It-Rate-Max scheme without PNC) at the sum rate of 8 bits per symbol period.

VI. Conclusion

In this paper, we have proposed an approach to iteratively solve the sum-rate maximization problem for the wireless MIMO switching networks with and without PNC. We proved that although the proposed algorithms are suboptimal (local optimal) in general, they can converge to asymptotically optimal solution in the low SNR regime regardless of the initial conditions, and near optimal solution in the high SNR regime with properly setting initial conditions. Numerical results show that the proposed iterative algorithms significantly outperform the existing ZF and MMSE relaying schemes for MIMO switching for all SNR.

This paper makes several assumptions to simplify the design and analysis of the MIMO switching schemes. For example, we assume that each user has a single antenna, and full channel state information is available to the relay. It will be of theoretical and practical interest to investigate the impact of relaxing there assumptions on the MIMO switching design. We will look into these issues in future work.

[2] A PNC scheme was proposed in [14] as well, which used the same block-diagonalization technique as that in [11]. However, the scheme in [11] induced an extra step to balance the channel gain of each user, which outperformed the scheme in [14]. Thus, we show the result of [11] in Fig. 2 only.

References

[1] S. Zhang, S. C. Liew, and P. P. Lam, “Hot topic: Physical-layer network coding,” in *Proc. of ACM Mobicom*, 2006.

[2] R. Zhang, Y.-C. Liang, C. C. Chai, and S. Cui, “Optimal beamforming for two-way multi-antenna relay channel with analogue network coding,” *IEEE J. Select. Areas Commun.*, vol. 27, no. 5, pp. 699–712, Jun. 2009.

[3] Z. Ding, I. Krikidis, J. Thompson, and K. Leung, “Physical layer network coding and precoding for the two-way relay channel in cellular systems,” *IEEE Trans. Signal Process.*, vol. 59, no. 2, 2011.

[4] R. Wang and M. Tao, “Joint source and relay precoding designs for MIMO two-way relaying based on MSE criterion,” *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1352–1365, Mar. 2012.

[5] T. Yang, X. Yuan, L. Ping, I. B. Collings, and J. Yuan, “A new eigen-direction alignment algorithm for physical layer network coding in MIMO two-way relay channels,” in *Proc. IEEE Int'l Symp. Inform. Theory*, Jul. 2011.

[6] C. Sun, C. Yang, Y. Li, and B. Vucetic, “Transceiver design for multi-user multi-antenna two-way relay cellular systems,” *to appear in IEEE Trans. Commun.*, 2012.

[7] D. Gündüz, A. Yener, A. J. Goldsmith, and H. V. Poor, “The multi-way relay channel,” *submitted to IEEE Trans. Inform. Theory, CoRR*, vol. abs/1004.2434, 2010.

[8] F. Wang, S. C. Liew, and D. Guo, “Wireless MIMO switching with zero-forcing relaying,” in *49th Annual Allerton Conf. on Commun., Control, and Computing*, Sept. 2011, pp. 551–558.

[9] Y. Mohasseb, H. Ghozlan, G. Kramer, and H. El Gamal, “The MIMO wireless switch: Relaying can increase the multiplexing gain,” in *Proc. IEEE Int'l Symp. Inform. Theory*, 2009, pp. 1448–1552.

[10] J. Joung and A. Sayed, “Multiuser two-way amplify-and-forward relay processing and power control methods for beamforming systems,” *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1833–1846, Mar. 2010.

[11] A. U. T. Amah and A. Klein, “Beamforming-based physical layer network coding for non-regenerative multi-way relaying,” *EURASIP J. Wireless Commun. and Network*, vol. 2010, p. 12, 2010.

[12] Y. E. Sagduyu, R. A. Berry, and D. Guo, “Throughput and stability for relay-assisted wireless broadcast with network coding,” *IEEE J. Select. Areas Commun.*, 2013.

[13] F. Wang, S. C. Liew, and D. Guo, “Wireless MIMO switching with zero forcing and network coding,” *IEEE J. Select. Areas Commun.*, vol. 30, no. 8, pp. 1452–1463, Sep. 2012.

[14] E. Yilmaz, R. Zakhour, D. Gesbert, and R. Knopp, “Multi-pair two-way relay channel with multiple antenna relay station,” in *Proc. IEEE Int'l Conf. on Commun.*, May 2010, pp. 1–5.

[15] M. Tao and R. Wang, “Linear precoding for multi-pair two-way mimo relay systems with max-min fairness,” *IEEE Trans. Signal Proc.*, vol. 60, no. 10, pp. 5361–5370, Oct. 2012.

[16] T. Cui, T. Ho, and J. Kliewer, “Space-time communication protocols for N-way relay networks,” in *Proc. IEEE GLOBECOM*, 2008, pp. 1–5.

[17] F. Wang, S. C. Liew, and D. Guo, “Wireless MIMO switching with MMSE relaying,” in *Proc. IEEE Int'l Symp. Inform. Theory*, Jul. 2012.

[18] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1990.

[19] T. M. Cover and J. A. Thomas, *Elements of Information Theory, 2nd Edition*. Wiley-Interscience, 2006.

[20] S. Christensen, R. Agarwal, E. Carvalho, and J. Cioffi, “Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design,” *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4792–4799, Dec. 2008.

[21] X. Wang and X.-D. Zhang, “Linear transmission for rate optimization in MIMO broadcast channels,” *IEEE Trans Wireless Commun.*, vol. 9, no. 10, pp. 3247–3257, Oct. 2010.

[22] D. P. Bertsekas, *Nonlinear Programming: 2nd Edition*. Athena Scientific, 1999.

[23] F. Wang, X. Yuan, S. C. Liew, and D. Guo, “Wireless MIMO switching: Sum mean square error and sum rate optimization,” *CoRR*, vol. abs/1207.6706, 2012.

[24] S. C. Liew, S. Zhang, and L. Lu, “Physical-layer network coding: Tutorial, survey, and beyond,” *Invited Paper; Elsevier Phycom, Special Issue of Physical Communication on “Network Coding and Its Applications to Wireless Communications”*, May 2012. [Online]. Available: http://dx.doi.org/10.1016/j.phycom.2012.05.002