Wireless MIMO Switching with Zero Forcing and Network Coding

Fanggang Wang, *Member, IEEE*, Soung Chang Liew, *Fellow, IEEE*, and Dongning Guo, *Senior Member, IEEE*

Abstract— A wireless relay with multiple antennas is called a multiple-input-multiple-output (MIMO) switch if it maps its input links to its output links using "precode-and-forward." Namely, the MIMO switch precodes the received signal vector in the uplink using some matrix for transmission in the downlink. This paper studies the scenario of K stations and a MIMO switch, which has full channel state information. The precoder at the MIMO switch is either a zero-forcing matrix or a network-coding matrix. With the zero-forcing precoder, each destination station receives only its desired signal with enhanced noise but no interference. With the network-coding precoder, each station receives not only its desired signal and noise, but possibly also self-interference, which can be canceled. Precoder design for optimizing the received signal-to-noise ratios at the destinations is investigated. For zero-forcing relaying, the problem is solved in closed form in the two-user case, whereas in the case of more users, efficient algorithms are proposed and shown to be close to what can be achieved by extensive random search. For network-coded relaying, we present efficient iterative algorithms that can boost the throughput further.

Index Terms—Beamforming, MIMO switching, network coding, relay, zero-forcing.

I. INTRODUCTION

R**ELAYS** in wireless networks can extend coverage as well as improve energy efficiency [1]. In this paper, we study a setup in which multiple single-antenna stations communicate with each other via a multi-antenna relay. In each uplink slot, the stations simultaneously transmit, then in the subsequent downlink slot, the relay precodes its received signal by a certain matrix before broadcasting to the stations. In the absence of noise, the multiple-input multiple-output (MIMO) system between the transmitters and the receivers can be viewed as a product of the downlink channel matrix, the precoder and the uplink channel matrix. In this work, we design a zero-forcing precoder so that the product channel is a desired permutation matrix, which forms a one-to-one mapping (or links) from the transmitters to the receivers.

F. Wang is with the State Key Lab of Rail Traffic Control and Safety, School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing, China (e-mail: fgwang@inc.cuhk.edu.hk).

S. Liew is with the Department of Information Engineering, the Chinese University of Hong Kong, HK SAR, China (e-mail: soung@ie.cuhk.edu.hk).

D. Guo is with the Department of Electrical Engineering & Computer Science, Northwestern University, Evanston, IL, USA (e-mail: dGuo@northwestern.edu).

Hence the technique is called MIMO switching. We also study a generalization referred to as network-coded precoding where the off-diagonal elements of the channel matrix form a permutation, and where the diagonal elements can be nonzero. Such nonzero diagonal elements cause self-interference, which can be fully canceled assuming the interference gains are available at the receivers. We study how to design the precoder to maximize the signal-to-noise ratios (SNRs) of the links.

To the best of our knowledge, this work is the first to treat unicast non-pairwise switching patterns. Prior work that investigates data exchange via a relay includes [1]–[6]. References [2], [4], [5] investigate the case of "full data exchange," in which all stations want to broadcast their data to all the other stations. A slotted system with a single-antenna relay is considered in [2] and the maximum throughput region is evaluated. Data transmissions in [4], [5] can be summarized as follows: In the first slot, all stations transmit to the relay simultaneously; subsequent slots are devoted to downlink transmissions; in each downlink slot, the relay multiplies the signal received in the first time slot by a different precoder, such that at the end of all downlink slots, all stations receive the broadcast data from all the other stations. By contrast, the framework investigated in this paper focuses on the unicast bijection case, in which each source is linked to one distinct destination and vice versa. Any general transmission pattern (unicast, multicast, broadcast, or a mixture of them) among the stations can be realized by scheduling a set of different unicast bijection transmissions, as has been pointed out by the authors in preliminary work [7]. A single-antenna relay with different forwarding strategies is considered in [1], which studies both full data exchange and "pairwise data exchange," in which stations form pairs to exchange data with each other only. It is a special case of unicast switching studied here. Reference [6] studies pairwise data exchange only, where the relay adopts the decode-and-forward strategy. The diversity-multiplexing tradeoff under reciprocal and non-reciprocal channels is also analyzed.

In this paper, we consider both pairwise and non-pairwise switching, in which a multi-antenna relay works in precode-and-forward manner. We first study switching traffic among the stations using a zero-forcing MIMO relay, where each destination receives the desired signal with enhanced noise. We then study more general network-coded relaying that exploits physical-layer network coding to improve performance [8]–[11]. For fairness, we study how to design the precoder to maximize the minimum received SNR among all stations, which is referred to as the *maxmin problem*. Since the maxmin problem is NP-hard, we use a semidefinite relaxation technique to
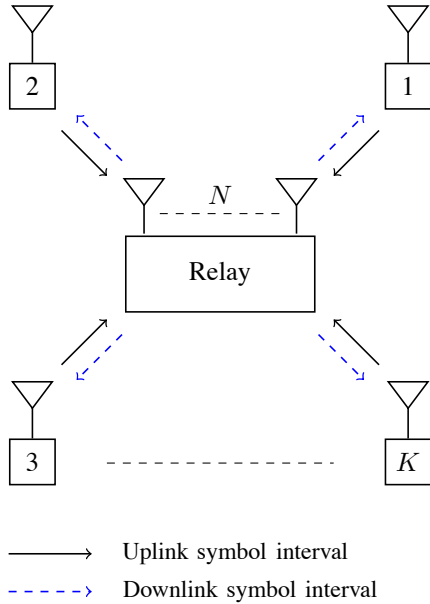
Fig. 1. Wireless MIMO switching.

compute an approximate solution. The problem can be further simplified if the required SNRs at all destinations are identical, and we call this formulation the *equal-SNR problem*. We derive conditions under which the maxmin and equal-SNR problems are equivalent. We find that the throughput gap between them is small, especially in the high SNR regime. That is, our numerical results suggest that we can use the equal-SNR problem to approximate the (NP-hard) maxmin problem. Furthermore, we show that network-coded relaying can noticeably improve the throughput performance over zero-forcing relaying.

The remainder of the paper is organized as follows: Section II introduces the scheme of wireless MIMO switching. In Section III (resp. Section IV), the maxmin (resp. equal-SNR) problem is investigated for both zero-forcing and network-coded relaying. Section V presents the simulation results. Section VI concludes this paper.

II. SYSTEM DESCRIPTION

Consider K stations, numbered $1, \ldots, K$, each with one antenna, as shown in Fig. 1. There is no direct link between any two stations and the stations communicate via a relay with N antennas. The precode-and-forward scheme applies under the condition of $K \leq N$, where the relay has enough degrees of freedom to switch all data streams at the same time. We assume $K = N$ throughout for simplicity. The analysis essentially also applies to the case of $K < N$, where all the matrix inverses in the paper shall be replaced by Moore-Penrose pseudo inverses [12]. Each transmission consists of one uplink symbol interval and one downlink symbol interval of equal duration. In particular, the two symbol intervals are two slots in a time-division system. The uplink symbol interval is for simultaneous uplink transmissions from the stations to the relay; the downlink symbol interval is for downlink transmissions from the relay to the stations. Each round

of uplink and downlink transmission realizes a switching permutation, as shall be described shortly.

Consider one transmission. Let $\boldsymbol{x} = [x_1, \cdots, x_N]^T$ be the vector representing the signals transmitted by the stations. Let $\boldsymbol{y} = [y_1, \cdots, y_N]^T$ be the received signals at the relay, and $\boldsymbol{u} = [u_1, \cdots, u_N]^T$ be the noise vector with independent identically distributed (i.i.d.) entries following circularly-symmetric complex Gaussian (CSCG) distribution, i.e., $u_n \sim \mathcal{N}_c(0, \gamma^2)$. Then

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

where $\boldsymbol{H}$ is the uplink channel gain matrix. The relay multiplies $\boldsymbol{y}$ by a precoding matrix $\boldsymbol{G}$ before relaying the signals. In this paper, we assume that the uplink channel and downlink channel are reciprocal, so that the downlink channel is $\boldsymbol{H}^T$. Thus, the received signals at the stations in vector form are

$$\boldsymbol{r} = \boldsymbol{H}^T \boldsymbol{G} \boldsymbol{y} + \boldsymbol{w} \tag{2}$$
$$= \boldsymbol{H}^T \boldsymbol{G} \boldsymbol{H} \boldsymbol{x} + \boldsymbol{H}^T \boldsymbol{G} \boldsymbol{u} + \boldsymbol{w}, \tag{3}$$

where $\boldsymbol{w}$ is the noise vector at the receiver, with the i.i.d. entries following CSCG distribution, i.e., $w_n \sim \mathcal{N}_c(0, \sigma^2)$.

In the following, we describe two precoding schemes.

A. Zero-forcing Relaying

We refer to an $N \times N$ matrix $\boldsymbol{P}$ that has one and only one nonzero element on each row and each column, which is equal to 1, as a *permutation* matrix. Evidently, $\boldsymbol{P}\boldsymbol{x}$ is a column vector consisting of the same elements as $\boldsymbol{x}$ but permuted in a certain order depending on $\boldsymbol{P}$. For example, if

$$\boldsymbol{P} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix},$$

then $\boldsymbol{P}[x_1, x_2, x_3]^T = [x_3, x_1, x_2]^T$. In the case where all diagonal elements of $\boldsymbol{P}$ are zero it is also called a *derangement*.

Suppose that the purpose of $\boldsymbol{G}$ is to realize a particular permutation represented by the permutation matrix $\boldsymbol{P}$, and to amplify the signals coming from the stations. That is,

$$\boldsymbol{H}^T \boldsymbol{G} \boldsymbol{H} = \boldsymbol{A} \boldsymbol{P}, \tag{4}$$

where $\boldsymbol{A} = \text{diag}\{a_1, \cdots, a_N\}$ is an "amplification" diagonal matrix. Each diagonal element is regarded as the gain of a link. Accordingly, the precoder can be calculated as

$$\boldsymbol{G} = \boldsymbol{H}^{-T} \boldsymbol{A} \boldsymbol{P} \boldsymbol{H}^{-1}. \tag{5}$$

Let the receivers compensate for the amplification to yield received signals expressed collectively as:

$$\hat{\boldsymbol{r}} = \boldsymbol{A}^{-1} \boldsymbol{r} = \boldsymbol{P}\boldsymbol{x} + \boldsymbol{v}, \tag{6}$$

where the *post-processing* noise is expressed as

$$\boldsymbol{v} = \boldsymbol{P}\boldsymbol{H}^{-1}\boldsymbol{u} + \boldsymbol{A}^{-1}\boldsymbol{w}. \tag{7}$$

Let us define

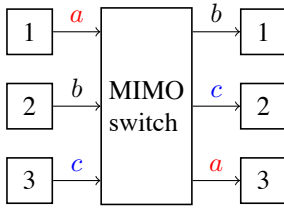$$\boldsymbol{Q} \triangleq \boldsymbol{I} + \gamma^2 \boldsymbol{P}\boldsymbol{H}^{-1}\boldsymbol{H}^{-H}\boldsymbol{P}^T. \tag{8}$$

Fig. 2.   A traffic demand among three stations.

The covariance of the post-processing noise v is written as

$$\mathbb{E}\{vv^H\} = \gamma^2 PH^{-1}H^{-H}P^T + \sigma^2 A^{-1}A^{-H} \qquad (9)$$
$$= Q - I + \sigma^2 A^{-1}A^{-H}. \qquad (10)$$

Suppose all uplink transmissions are independent and use unit average power, i.e., $\mathbb{E}\{x_i^2\} = 1$, $i = 1, \cdots, N$. The problem is to design the precoder G to minimize the variance of the post-processing noise subject to a power constraint for the relay:

$$\mathbb{E}\{y^H G^H Gy\} \leq p. \qquad (11)$$

For notational convenience, let the entries of an $N \times N$ matrix S be given by

$$s_{ij} \triangleq Q_{ji}[(H^*)^{-1}H^{-T}]_{ij}. \qquad (12)$$

From (1), the relay's transmit power can be evaluated as

$$\mathbb{E}[x^H H^H G^H GHx + u^H G^H Gu]$$
$$= \text{Tr}[GHH^H G^H + \gamma^2 GG^H] \qquad (13)$$
$$= \text{Tr}[H^{-T}A(I + \gamma^2 PH^{-1}H^{-H}P^T)A^H(H^*)^{-1}] \qquad (14)$$
$$= \text{Tr}[AQA^H(H^*)^{-1}H^{-T}] \qquad (15)$$
$$= a^H Sa, \qquad (16)$$

where we have used (5) and (8), and $a = [a_1, \cdots, a_N]^T$ is the gain vector with the diagonal elements of A. The power constraint on the relay is thus expressed as

$$a^H Sa \leq p. \qquad (17)$$

B. Network-coded Relaying

The MIMO switch described in Section II.A makes use of zero-forcing relaying, by which data are switched based on a permutation matrix, whose diagonal elements are all zero. If the kth diagonal element is nonzero, it means that the relay forwards the signal from station k back to itself. There is no need to force a diagonal element to zero because the self-interference is known and can be removed. This is the basic idea behind physical-layer network coding (PNC) [11], which underlies many other works, e.g., [13]–[16].

In general, allowing PNC improves the performance. Even though the self-interference costs the relay some energy, removing the constraint on the diagonal of the derangement enlarges the feasible set of the optimization problem, and thereby yields a better optimal objective. This can also be seen from an example of multi-way relaying, in which three single-antenna stations communicate with the help of a three-antenna relay. The traffic switching pattern among the three
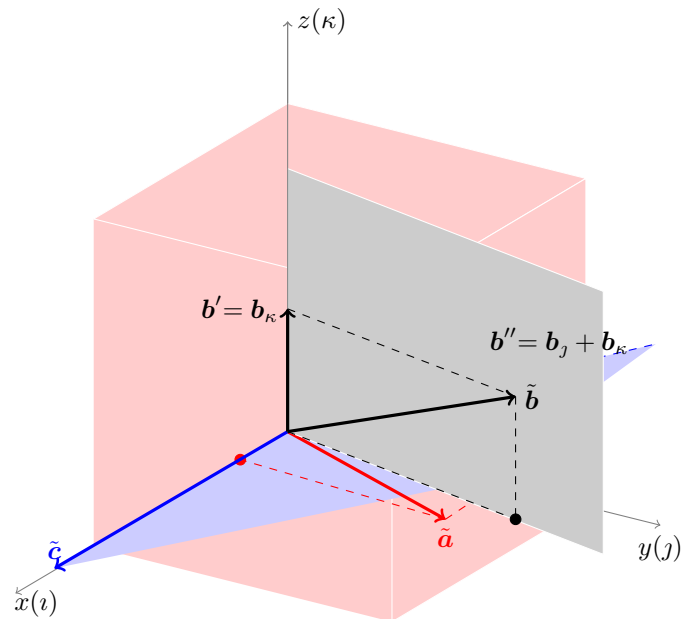

Fig. 3.   Assume the channel outputs of three users' signal vectors: the back vector $\tilde{b}$ is the desired signal; the red vector $\tilde{a}$ is the self-interference; the blue vector $\tilde{c}$ is the other interference signal.

stations are defined in Fig. 2. The desired signal of station 1 is the black signal b; the red signal a is its self-interference; the blue signal c is the other interference signal. We assume the three signals and the channel gains are real-valued for simpler illustration. A sketch of the three signals after passing through the channel is shown in Fig. 3, i.e., $\tilde{a}$, $\tilde{b}$ and $\tilde{c}$, which are all three-dimensional vectors due to three antennas at the relay. As shown in Fig. 3, we assume that $\tilde{c}$ is along x-axis; $\tilde{a}$ is in xy-plane; and $\tilde{b} = b_\iota + b_\jmath + b_\kappa$. In the case of zero-forcing, the desired signal of station 1, $\tilde{b}$, should be projected to z-axis, which is perpendicular to xy-plane spanned by the two signals $\tilde{a}$ and $\tilde{c}$. In this way both the two interference signals are zeroed out and we get a post-processing signal $b' = b_\kappa$. However, with network coding, we do not have to zero out the self-interference. By dropping this constraint, we only need to project the desired signal to the yz-plane which is perpendicular to the interference signal $\tilde{c}$, then we obtain the projected signal $b'' = b_\jmath + b_\kappa$. Obviously, the projection b'' with network coding is stronger than b' by zero-forcing.

With PNC, we rewrite (4) as

$$H^T GH = A(P + B), \qquad (18)$$

where $B = \text{diag}\{b_1, \cdots, b_N\}$ is a diagonal matrix to be determined.

As an example of a symmetric derangement, the corresponding network-coding switch matrix has the pattern:

$$P_1 + B = \begin{bmatrix} b_1 & 0 & 0 & 1 \\ 0 & b_2 & 1 & 0 \\ 0 & 1 & b_3 & 0 \\ 1 & 0 & 0 & b_4 \end{bmatrix}. \qquad (19)$$

This switching pattern corresponds to two pairwise data exchanges, in which stations 1 and 4 are one pair and stations 2 and 3 are the other pair. Hence, the network-coded MIMO

switching can construct multiple parallel two-way relay transmissions. An example of asymmetric derangement may have the following switch matrix:

$$\boldsymbol{P}_2 + \boldsymbol{B} = \begin{bmatrix} b_1 & 0 & 1 & 0 \\ 1 & b_2 & 0 & 0 \\ 0 & 0 & b_3 & 1 \\ 0 & 1 & 0 & b_4 \end{bmatrix}. \tag{20}$$

This generalizes the traditional physical-layer network coding setting as presented in [11] because the data exchange is not pairwise. For both symmetric and asymmetric switch matrices, we shall refer to the corresponding matrices with nonzero diagonal as network-coding switch matrix, and the associated setup as MIMO switching with network-coded relaying.

Once the receivers compensate for the amplification and remove self-interference, the resulting signals form this vector:

$$\hat{\boldsymbol{r}} = \boldsymbol{P}\boldsymbol{x} + \boldsymbol{v}', \tag{21}$$

where the post-processing noise $\boldsymbol{v}'$ is expressed as

$$\boldsymbol{v}' = (\boldsymbol{P} + \boldsymbol{B})\boldsymbol{H}^{-1}\boldsymbol{u} + \boldsymbol{A}^{-1}\boldsymbol{w}. \tag{22}$$

The covariance of $\boldsymbol{v}'$ is written as

$$\mathbb{E}\{\boldsymbol{v}'\boldsymbol{v}'^H\} = \boldsymbol{R} - \boldsymbol{I} + \sigma^2 \boldsymbol{A}^{-1}\boldsymbol{A}^{-H}, \tag{23}$$

where

$$\boldsymbol{R} \triangleq \boldsymbol{I} + \gamma^2 (\boldsymbol{P} + \boldsymbol{B})\boldsymbol{H}^{-1}\boldsymbol{H}^{-H}(\boldsymbol{P} + \boldsymbol{B})^H. \tag{24}$$

The constraint of the relay power consumption is rewritten as

$$\Omega(\boldsymbol{A}, \boldsymbol{B})$$
$$\triangleq \text{Tr}[\boldsymbol{H}^{-T}\boldsymbol{A}(\boldsymbol{R} + \boldsymbol{B}\boldsymbol{P}^T + \boldsymbol{P}\boldsymbol{B}^H + \boldsymbol{B}\boldsymbol{B}^H)\boldsymbol{A}^H(\boldsymbol{H}^*)^{-1}] \tag{25}$$
$$= \text{Tr}[\boldsymbol{A}(\boldsymbol{R} + \boldsymbol{B}\boldsymbol{P}^T + \boldsymbol{P}\boldsymbol{B}^H + \boldsymbol{B}\boldsymbol{B}^H)\boldsymbol{A}^H(\boldsymbol{H}^*)^{-1}\boldsymbol{H}^{-T}] \tag{26}$$
$$\leq p. \tag{27}$$

We have thus established another framework for MIMO switching by network-coded relaying method, in which $\boldsymbol{P} + \boldsymbol{B}$ is a switch matrix. This framework can be generalized to the case where the switch matrix realizes a general transmission pattern. For example, if there are two nonzero non-diagonal elements in a column of the switch matrix, then a multicast connection is being realized within one switch matrix. In fact, by scheduling a set of switch matrices, each realizing a permutation, we can satisfy arbitrary user traffic patterns.

III. THE MAXMIN PROBLEM

In this section, we formulate a *maxmin problem*, in which the minimum received SNR among all the stations is maximized. According to (9), the post-processing noise power of receiver i is

$$\epsilon_i = q_i - 1 + \frac{\sigma^2}{|a_i|^2}, \tag{28}$$

where $q_i \triangleq \boldsymbol{Q}_{ii}$. Thus, the received SNR is $1/\epsilon_i$. Since the system is half-duplex with uplink and downlink of equal duration, the throughput achieved by Gaussian signaling is

$$c_i = \frac{1}{2} \log_2\left(1 + \frac{1}{\epsilon_i}\right), \tag{29}$$

in bits per symbol period.

We first study the maxmin problem without PNC in Section III.A, and then allow PNC in Section III.B.

A. Zero-forcing Relaying

Let ϵ denote the maximum post-processing noise power among the stations. An optimization problem is formulated as follows:

$$\min_{\boldsymbol{a}} \quad \epsilon \tag{30a}$$

$$\text{s.t.} \quad |a_i|^2 \geq \frac{\sigma^2}{\epsilon + 1 - q_i}, \quad i = 1, \cdots, N, \tag{30b}$$

$$\boldsymbol{a}^H \boldsymbol{S} \boldsymbol{a} \leq p, \tag{30c}$$

$$\epsilon \geq 0. \tag{30d}$$

Lemma 1: Every optimal solution for the optimization problem (30) must satisfy the relay power constraint (30c) with equality.

Proof: Let $\tilde{\boldsymbol{a}} = [\tilde{a}_1, \cdots, \tilde{a}_N]^T$ denote the optimal solution for the problem (30) with the optimal objective $\tilde{\epsilon}$. Suppose $\tilde{\boldsymbol{a}}$ satisfies the constraint (30c) with strict inequality. Then there exists $\tau > 0$ such that

$$\boldsymbol{a}^H \boldsymbol{S} \boldsymbol{a} < p, \tag{31}$$

for all $\boldsymbol{a}$ with $|a_i| \in (|\tilde{a}_i| - \tau, |\tilde{a}_i| + \tau)$, $i = 1, \cdots, N$. Let $|\tilde{a}_i'| = |\tilde{a}_i| + \frac{\tau}{2}$, $i = 1, \cdots, N$, and let

$$\Xi_i(a_i) \triangleq q_i + \frac{\sigma^2}{|a_i|^2}. \tag{32}$$

Then

$$\Xi_i(\tilde{a}_i') < \Xi_i(\tilde{a}_i), \quad i = 1, \cdots, N. \tag{33}$$

Since each $\Xi_i(\tilde{a}_i')$ is smaller than $\Xi_i(\tilde{a}_i)$ for all i, the maxmin objective becomes smaller with the solution $\tilde{\boldsymbol{a}}' = [a_1', \cdots, a_N']^T$. Thus, $\tilde{\boldsymbol{a}}$ is not the optimal solution, and contradiction arises. ∎

With this lemma, we could eliminate the feasible solutions with which the relay's power consumption is less than p.

Proposition 1: The maxmin problem[1] (30) is equivalent to the following quadratically constrained quadratic program (QCQP),

$$\min_{\boldsymbol{a}} \quad \boldsymbol{a}^H \boldsymbol{S} \boldsymbol{a} \tag{34a}$$

$$\text{s.t.} \quad |a_i|^2 \geq \frac{\sigma^2}{\tilde{\epsilon} + 1 - q_i}, \quad i = 1, \cdots, N. \tag{34b}$$

Proof: Let $\tilde{\boldsymbol{a}}$ be an optimal solution of (30), and the associated optimal objective is $\tilde{\epsilon} \geq 0$. Note that $\tilde{\epsilon}$ is the maximum noise power for all i.

The solution $\tilde{\boldsymbol{a}}$ satisfies (34b), where the largest noise power among all i is equal to $\tilde{\epsilon}$. Hence, $\tilde{\boldsymbol{a}}$ is a feasible solution of (34). According to Lemma 1, the power consumption of the relay is $\tilde{\boldsymbol{a}}^H \boldsymbol{S} \tilde{\boldsymbol{a}} = p$. Let $\hat{\boldsymbol{a}}$ be an optimal solution of (34). The power consumption of the relay can not be larger than p, otherwise $\hat{\boldsymbol{a}}$ is even worse than $\tilde{\boldsymbol{a}}$ for (34). If the power consumption of the relay $\hat{\boldsymbol{a}}^H \boldsymbol{S} \hat{\boldsymbol{a}}$ is strictly smaller than p, then $\hat{\boldsymbol{a}}$ is a feasible solution of (30), which is at least as good as $\tilde{\boldsymbol{a}}$ since the maximum noise power is not larger than $\tilde{\epsilon}$. That is, $\hat{\boldsymbol{a}}$ is an optimal solution of (30). However, the

[1]Essentially, it is a minmax problem with respect to (w.r.t.) the post-processing noise power. In this paper, we call it the maxmin problem w.r.t. the received SNR to keep it consistent with the following equal-SNR problem proposed in Section IV.

power consumption $\hat{a}^H S \hat{a} < p$ contradicts Lemma 1. Thus, $\hat{a}^H S \hat{a} = p$. Furthermore, there is at least one constraint in (34) in which equality holds for $\hat{a}$, otherwise $\hat{a}$ is a better solution for (30). We have proved the optimal solution of (30) is also the optimal solution of (34), vice versa. Therefore, the two problems are equivalent. ∎

Proposition 2: The maxmin problem (30) is NP-hard in the size of N when $\gamma \neq 0$.

Proof: Problem (34) is equivalent to problem (2) in [17], which has been proved to be NP-hard in general. Thus, problem (30) is NP-hard as well. We do not repeat the steps in [17]. ∎

We can use the semidefinite relaxation (SDR) technique in [17] to find a throughput upper bound and a suboptimal solution for (34). Let $X = aa^H$, then (34) can be rewritten as

$$\min_{X \in \mathbb{C}^{N \times N}, \ell_i} \quad \text{Tr}[SX] \tag{35a}$$

$$\text{s.t.} \quad \text{Tr}[E_i \odot X] - \ell_i = \frac{\sigma^2}{\tilde{\epsilon} + 1 - q_i}, \tag{35b}$$

$$\ell_i \geq 0, \; i = 1, \cdots, N, \tag{35c}$$

$$X \succeq 0, \tag{35d}$$

$$\text{rank}(X) = 1, \tag{35e}$$

where $\odot$ denotes element-by-element multiplication, i.e., the Hadamard product; $X \succeq 0$ means the matrix X is symmetric positive semidefinite; E_i is an $N \times N$ matrix, in which element (i,i) is 1 and all the other elements are 0; $\ell_i, \; i = 1, \cdots, N$ are "slack" variables. If we drop the rank-1 constraint (35e), problem (35) is in the standard form of a semidefinite programming problem (SDP).

Overview of Our Approach to Maxmin Optimization Problem: We now overview the approach of our numerical investigation of the maxmin optimization problem. Although we focus on zero-forcing relaying here, we use the same approach for the study of maxmin optimization for network-coded relaying as well after its corresponding formulation is set up in Part B.

We recognize that it is difficult to solve the maxmin problem (30). Therefore, we attempt to find its suboptimal solution. The logical steps and the rationale for our approach to finding a suboptimal solution are summarized below:

1) We have already proved that the maxmin problem is equivalent to the QCQP (34). Unfortunately, solving for the optimal solution of the QCQP is still difficult.

2) Fortunately, the SDR technique proposed in [17] can be used to find a good suboptimal solution of the QCQP. The technique consists of the following three steps:

   a) Rewrite the QCQP as an optimization problem (35) with a constraint of rank 1.

   b) By dropping the constraint of rank 1, the problem (35) is turned into a SDP. The optimal solution of the SDP can be found by some toolboxes, such as SeDuMi and SDPT3. Importantly, the optimal objective of the SDP is an upper bound of that of the QCQP, since the QCQP has one extra constraint, i.e., the constraint of rank 1.

   c) We then use the randomization technique in [17] to approximate the optimal solution of the QCQP based on the optimal solution of the SDP. With the randomization technique, we get a result, which satisfies the constraints of the QCQP, and this result is at least a suboptimal solution of the QCQP.

In [17], the authors showed that the suboptimal solution obtained as above can achieve an objective close to the global optimum. As will be shown, our simulation results also validate the near-optimality statement in [17].

Numerical Method (One-dimensional Search): As per the discussion in the above overview, we use the suboptimal solution of (35) to approximate the solution of (30). The optimal solution of (30) is the value of $\tilde{\epsilon}$ in (35), for which the objective of (35), i.e., the minimum power consumption is p. We solve for $\tilde{\epsilon}$ by one-dimension search from $\max_j q_j - 1$, i.e., when the power consumption of the relay (35a) approaches infinity. In each step of the search, $\tilde{\epsilon}$ is increased by a small amount $\tilde{\delta}$. Given an $\tilde{\epsilon}$, we can use SDP solvers to find the minimum power consumption of the relay. When the step size $\tilde{\delta}$ is small enough, the first value of $\tilde{\epsilon}$, for which the minimum power consumption is p, is the solution (30).

B. Network-coded Relaying

The network-coded problem is formulated as

$$\min_{A,B} \quad \epsilon \tag{36a}$$

$$\text{s.t.} \quad R_{ii} - 1 + \frac{\sigma^2}{|a_i|^2} \leq \epsilon, \; i = 1, \cdots, N, \tag{36b}$$

$$\Omega(A, B) \leq p, \tag{36c}$$

$$\epsilon \geq 0, \tag{36d}$$

where R and $\Omega(A, B)$ are defined in (24) and (26), respectively. In contrast to the previous problem of zero-forcing relaying, the constraint of the relay power consumption (36c) is quartic, making it more difficult than (30). We propose an iterative algorithm, in which A and B defined in (18) are optimized iteratively.

1) Optimize A for given B: Given B, problem (36) can be formulated as (30) by redefining

$$s_{ij} \triangleq [R + BP^T + PB^H + BB^H]_{ji}[(H^T H^*)^{-1}]_{ij}, \tag{37}$$

$$q_i \triangleq R_{ii}, \tag{38}$$

Thus, the network-coded problem with fixed B can be solved by the SDR technique, which is the same as solving (34).

2) Optimize B for given A: For ease of notation, let

$$M \triangleq A^H (H^*)^{-1} H^{-T} A, \tag{39}$$

$$N \triangleq I + \gamma^2 H^{-1} H^{-H}. \tag{40}$$

The relay power consumption (26) is rewritten as

$$\Omega(A, B) = \text{Tr}[(P + B)N(P + B)^H M] \tag{41}$$

$$= \text{Tr}[BNB^H M + BNP^T M + MPNB^H + PNP^T M]. \tag{42}$$

Constraint (36b) can be rewritten as

$$\left[\boldsymbol{I} + \gamma^2 (\boldsymbol{P} + \boldsymbol{B}) \boldsymbol{H}^{-1} \boldsymbol{H}^{-H} (\boldsymbol{P} + \boldsymbol{B})^H \right]_{ii} \leq \epsilon - \frac{\sigma^2}{|a_i|^2}. \quad (43)$$

Both the objective and the constraint w.r.t. $\boldsymbol{B}$ are inhomogeneous quadratic. To homogenize this problem, we let $\tilde{\boldsymbol{b}} = [\boldsymbol{b}^T, t]^T$, where $\boldsymbol{b} = \text{diag}\{\boldsymbol{B}\}$. Then the relay power consumption (42) can be written as

$$\begin{aligned}
& \begin{bmatrix} \boldsymbol{b}^H & t^* \end{bmatrix} \begin{bmatrix} \boldsymbol{S}_t & \boldsymbol{f} \\ \boldsymbol{f}^H & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{b} \\ t \end{bmatrix} + \text{Tr}[\boldsymbol{P} \boldsymbol{N} \boldsymbol{P}^T \boldsymbol{M}] \\
& \triangleq \tilde{\boldsymbol{b}}^H \tilde{\boldsymbol{S}} \tilde{\boldsymbol{b}} + \text{Tr}[\boldsymbol{P} \boldsymbol{N} \boldsymbol{P}^T \boldsymbol{M}],
\end{aligned} \quad (44)$$

where $\boldsymbol{S}_t = \boldsymbol{N}^T \odot \boldsymbol{M}$, $\boldsymbol{f} = \text{diag}\{\boldsymbol{M} \boldsymbol{P} \boldsymbol{N}\}$, $t = 1$. Let $\tilde{\boldsymbol{X}} = \tilde{\boldsymbol{b}} \tilde{\boldsymbol{b}}^H$, then this problem can be homogenized as

$$\min_{\tilde{\boldsymbol{X}} \in \mathbb{C}^{N \times N}, \ell_i} \text{Tr}[\tilde{\boldsymbol{S}} \tilde{\boldsymbol{X}}] \quad (45\text{a})$$

$$\text{s.t.} \quad \text{Tr}[\tilde{\boldsymbol{E}}_i \odot \tilde{\boldsymbol{X}}] + \ell_i \quad (45\text{b})$$

$$= \frac{1}{\gamma^2} \left(\tilde{\epsilon} - \frac{\sigma^2}{|a_i|^2} \right) - 1 - \left[\boldsymbol{P} (\boldsymbol{H}^H \boldsymbol{H})^{-1} \boldsymbol{P}^H \right]_{ii},$$

$$\ell_i \geq 0, \ i = 1, \cdots, N, \quad (45\text{c})$$

$$\text{Tr}[\tilde{\boldsymbol{E}}_{N+1} \odot \tilde{\boldsymbol{X}}] = 1, \quad (45\text{d})$$

$$\boldsymbol{X} \succeq \boldsymbol{0}, \quad (45\text{e})$$

$$\text{rank}(\boldsymbol{X}) = 1, \quad (45\text{f})$$

where $\tilde{\boldsymbol{E}}_i$, $i = 1, \cdots, N$, can be easily designed according to (43). The last element of $\tilde{\boldsymbol{E}}_{N+1}$ is 1 and all the others are 0. Thus, constraint (45d) is actually $|t|^2 = 1$. The problem can be solved in the same way as solving (35). Note that constraint (45d) only requires the magnitude of t to be 1. Assume that the solution of (45) is $\tilde{\boldsymbol{b}}_{opt} = [\hat{\boldsymbol{b}}^T, \hat{t}]^T$, then $\tilde{\boldsymbol{b}}_{opt} e^{j\alpha}$, for all α also guarantees the optimality of (45), since adding a phase rotation does not change $\tilde{\boldsymbol{X}}$. Thus, the solution of this problem is $\hat{\boldsymbol{b}} e^{-j\angle \hat{t}}$.

To sum up the iterative algorithm, in order to design $\boldsymbol{A}$ and $\boldsymbol{B}$, we can first initialize $\boldsymbol{A}$ or $\boldsymbol{B}$ by any diagonal matrix. Then we iteratively perform the two solvers to optimize $\boldsymbol{A}$ and $\boldsymbol{B}$ until convergence.

IV. THE EQUAL-SNR PROBLEM

In this section, we require the received SNR at all stations to be identical and find its maximum of the *equal-SNR problem*. Such a formulation was first investigated in [7]. The equal-SNR requirement not only guarantees perfect fairness, but also allows more efficient computation.

A. Zero-forcing Relaying

The optimization problem is formulated as

$$\min_{\boldsymbol{a}} \quad \epsilon \quad (46\text{a})$$

$$\text{s.t.} \quad |a_i|^2 = \frac{\sigma^2}{\epsilon + 1 - q_i}, \ i = 1, \cdots, N, \quad (46\text{b})$$

$$\boldsymbol{a}^H \boldsymbol{S} \boldsymbol{a} \leq p, \quad (46\text{c})$$

$$\epsilon \geq 0. \quad (46\text{d})$$

The only difference between the equal-SNR problem and the maxmin problem (30) is that the inequality (30b) is replaced by the equality (46b). In [18], we only provided suboptimal solutions. In this paper, we will investigate optimal solutions and propose analytical and numerical suboptimal solutions. As has been proved in [18], the equal-SNR problem is feasible. Another property is that the solution of the equal-SNR problem is feasible for the maxmin problem, since constraint (30b) of the maxmin problem has a larger possible set than constraint (46b) in the equal-SNR problem. Thus, the optimal objective of (46) can not be smaller than that of (30).

Proposition 3: If no additional noise is introduced at the switch, then the optimal solution of (30) is such that each station has exactly the same post-processing noise power. That is, in this case, the two optimization problems (30) and (46) are equivalent.

Proof: As proved in Proposition 2, the optimization problem (34) is equivalent to (30). When $\gamma = 0$ in (34), $q_i = 1$ for all $i = 1, \cdots, N$. The objective can be rewritten as $\sum_{i=1}^{N} s_{ii} |a_i|^2$. The optimal solution of $\boldsymbol{a}$ is obvious, and it satisfies $|a_i|^2 = \frac{\sigma^2}{\tilde{\epsilon} - 1}$. That is, for the optimal solution of (30), each station has equal post-processing $\tilde{\epsilon}$. ∎

Thus, we can use the solution of the equal-SNR problem to approximate that of the maxmin problem in the high SNR regime. The gap will be evaluated numerically in Section V.

Optimal Solution in the Case of Two Stations

In order to minimize the post-processing noise power ϵ, we should try to maximize $|a_1|$ and $|a_2|$. Given any ϵ, the amplitudes $|a_1|$ and $|a_2|$ can be calculated by the equal noise power constraint. Then we should find their optimal phases to minimize the relay power consumption.

If $N = 2$, the power constraint of the relay in (46c) can be expanded as

$$\boldsymbol{a}^H \boldsymbol{S} \boldsymbol{a} = s_{11} |a_1|^2 + s_{22} |a_2|^2 + s_{21} a_1 a_2^* + s_{12} a_1^* a_2, \quad (47)$$

where $s_{11} \geq 0$, $s_{22} \geq 0$, and $s_{12} = s_{21}^*$. By the definition of $\boldsymbol{S}$ in (12), $s_{12} = s_{21} \geq 0$. Formula (47) can be written as

$$\boldsymbol{a}^H \boldsymbol{S} \boldsymbol{a} = \left| \sqrt{s_{11}} a_1 + \frac{s_{12}}{\sqrt{s_{11}}} a_2 \right|^2 + \left(s_{22} - \frac{s_{12}^2}{s_{11}} \right) |a_2|^2. \quad (48)$$

Since in (48) only $\left| \sqrt{s_{11}} a_1 + \frac{s_{12}}{\sqrt{s_{11}}} a_2 \right|$ is related to the phases of variables a_1 and a_2, the global minimum is achieved by real-valued a_1 and a_2 with opposite signs. Without loss of optimality, assume $a_1 \geq 0$ and $a_2 \leq 0$. Then the power consumption constraint can be simplified as

$$\boldsymbol{a}^T \boldsymbol{S} \boldsymbol{a} = s_{11} a_1^2 + 2 s_{12} a_1 a_2 + s_{22} a_2^2. \quad (49)$$

According to the equal noise power constraint (46b), we have

$$a_2 = - \frac{\sigma}{\sqrt{q_1 - q_2 + \frac{\sigma^2}{a_1^2}}}. \quad (50)$$

Plugging (50) into the power constraint, we have a biquartic equation (51),

$$\begin{aligned}
0 = & s_{11}^2 q_\delta^2 a_1^8 + 2 q_\delta [s_{11}^2 \sigma^2 + s_{11} s_{22} \sigma^2 - s_{11} p q_\delta \\
& - 2 s_{12}^2 \sigma^2] a_1^6 + [\sigma^4 (s_{11} + s_{22})^2 + p^2 q_\delta^2 \\
& - 2 p \sigma^2 q_\delta (2 s_{11} + s_{22}) - 4 s_{12}^2 \sigma^4] a_1^4 \\
& + [2 p^2 \sigma^2 q_\delta - 2 p \sigma^4 (s_{11} + s_{22})] a_1^2 + p^2 \sigma^4,
\end{aligned} \quad (51)$$

where $q_\delta = q_1 - q_2$. Since the equal-SNR problem is feasible, there exists solutions for (51). In order to maximize a_1 and a_2, the largest real root of (51) is the optimal solution of a_1, which admits an analytical solution [19]. Consequently, a_2 can be calculated by (50).

Alternatively, after we deduce that a_1 and a_2 have opposite signs, the one-dimensional search in Section III.A can be used to solve the problem as well. Thus, we have solved the equal-SNR problem of (46) in the case of $N = 2$.

It is not difficult to see that letting the two user signals take opposite signs (phases) minimizes the relay power consumption. The optimal solution of the maxmin problem when $N = 2$ also has the opposite-phase property. As we shall see, in the case of more than two users with users forming pairs, assigning each pair of users opposite phases is an effective scheme. Thus, in general the opposite-phase setting is effective for pairwise transmission in any problem which needs to minimize the relay power consumption, including the preceding maxmin problem.

Suboptimal Solution in the Case of $N > 2$

If there are more than two stations, we propose a suboptimal solution to (46). The vector $\boldsymbol{a}$ consists of arbitrary complex numbers. Given ϵ, one can obtain $|a_j|$ from (46b). Let the phases $\theta_j = \angle a_j$, $j = 1, \cdots, N$, be fixed. We solve (46) to obtain the minimum noise variance $\epsilon(\theta_1, \cdots, \theta_N)$, which satisfies $\boldsymbol{a}^H \boldsymbol{S} \boldsymbol{a} = p$.

We now consider the general case of complex-valued $\boldsymbol{a}$. With the numerical method of one-dimensional search, the optimization of a_j's amplitudes and phases can be decoupled. There exists an ϵ such that $\boldsymbol{a}^H \boldsymbol{S} \boldsymbol{a} = p$. Denote such an ϵ by $\epsilon(\theta_1, \cdots, \theta_N)$ since in each step we regard the amplitudes as constant values. It then suffices to solve for

$$\epsilon^* = \arg \min_{\theta_1, \cdots, \theta_N} \epsilon(\theta_1, \cdots, \theta_N). \tag{52}$$

In the following we provide two suboptimal algorithms for (46) via (52).

Non-PNC Random-phase Algorithm: In (52), we note that ϵ is a complicated nonlinear function of θ_j. A time-consuming exhaustive search can be used to find the solution to (52). To reduce the computation time, we find the best set of phases over a randomly generated candidates in lieu of an exhaustive search. We call this the *random-phase algorithm*. We divide the interval of $[0, 2\pi)$ equally into M bins with the values of $0, \frac{2\pi}{M}, \cdots, \frac{2(M-1)\pi}{M}$ respectively, and we randomly pick among them to set the the value of θ_j for each and every $j = 1, \cdots, N$. After that, we compute the corresponding $\epsilon(\theta_1, \cdots, \theta_N)$ by solving (46b) and (46c). Given an ϵ, substituting it into (46b) yields $|a_j|$ for all j. We perform L trials of these random phase assignments to obtain L phase vectors of $(\theta_1, \cdots, \theta_N)$. Calculating the relay power consumption by (46c), we choose the phase vector using the least power among the L candidates as our approximated phase solution. By one-dimensional search, the estimated ϵ^* can be achieved when the least power consumption of the relay is p. Accordingly, we can calculate $|a_j|$ for all j with the estimated ϵ^*, then $\boldsymbol{a}$ with the approximated phase solution. Hence, $\boldsymbol{G}$ can be calculated by its definition (5). This best-out-of-L-trials feasible solution is

in general larger than the actual optimal ϵ^*. In Section V, we will show that large gains can be achieved with only small M and L. Moreover, increasing M and L further yields very little improvement, suggesting that the estimated ϵ^* with small M and L is close to the result achieved by an extensive search.

Non-PNC Opposite-phase Algorithm: Recall that in the case of $N = 2$ the solution has a_1 and a_2 with opposite signs, or in general the minimum power consumption is achieved with a_1 and a_2 being complex numbers with opposite phases, i.e., $\theta_1 = \theta_2 + \pi$. For large even number of N, consider the situation in which the transmissions are pairwise. The stations form pairs, and two stations in a pair exchange data with each other only. Assume that station pair ℓ consists of stations $\pi(\ell)$ and $\kappa(\ell)$. Define

$$\begin{bmatrix} h_{\ell 1} h_{\ell 2} \\ h_{\ell 2}^* h_{\ell 3} \end{bmatrix} \triangleq \begin{bmatrix} [(\boldsymbol{H}^H \boldsymbol{H})^{-1}]_{\pi(\ell), \pi(\ell)} & [(\boldsymbol{H}^H \boldsymbol{H})^{-1}]_{\pi(\ell), \kappa(\ell)} \\ [(\boldsymbol{H}^H \boldsymbol{H})^{-1}]_{\kappa(\ell), \pi(\ell)} & [(\boldsymbol{H}^H \boldsymbol{H})^{-1}]_{\kappa(\ell), \kappa(\ell)} \end{bmatrix} \tag{53}$$

where $h_{\ell 1} \geq 0$, $h_{\ell 3} \geq 0$, $h_{\ell 2} \in \mathbb{C}$. The post-processing noise power can be factorized in terms of $b_{\ell 1}$ and $b_{\ell 2}$ as follows:

$$\epsilon = \gamma^2 h_{\ell 3} + \frac{\sigma^2}{|a_{\ell 1}|^2} \tag{54}$$

$$= \gamma^2 h_{\ell 1} + \frac{\sigma^2}{|a_{\ell 2}|^2}. \tag{55}$$

The relay power consumption can be written as

$$\Omega_1(\boldsymbol{a}) = \gamma^2 \mathcal{O}_1(\boldsymbol{a}) + \sum_{\ell=1}^{N/2} \left\{ \left| a_{\ell 1} + \gamma^2 |h_{\ell 2}|^2 a_{\ell 2} \right|^2 + \left(h_{\ell 1} + \gamma^2 h_{\ell 1} h_{\ell 3} - 1 \right) |a_{\ell 1}|^2 + \left(h_{\ell 3} + \gamma^2 h_{\ell 1} h_{\ell 3} - \gamma^4 |h_{\ell 2}|^4 \right) |a_{\ell 2}|^2 \right\}, \tag{56}$$

where the summation consists of the inner-pair quadratic items of $\boldsymbol{a}$, and $\mathcal{O}_1(\boldsymbol{a})$ denotes the sum of the quadratic items of $\boldsymbol{a}$ across pairs, e.g., $a_i a_j$ where i, j from different pairs. For fixed amplitudes of $a_{\ell 1}$ and $a_{\ell 2}$, it is obvious that the phase vector $(\theta_1, \cdots, \theta_N)$ that minimizes (56) is the optimal phase vector of (46). We could let two stations of a pair to have opposite phases to lower the inner-pair part of (56). Simulation results indicate that as long as the relative phase is π within a pair, the throughput performance remains essentially the same regardless of the phase differences between different pairs. Thus, to simplify the problem we use real numbers for the elements of $\boldsymbol{a}$. The amplitudes are calculated in the same way as that of the non-PNC random-phase algorithm.

Remark 1: The opposite-phase solution approaches the optimal solution of the equal-SNR problem as SNR increases, since $\gamma^2 \mathcal{O}_1(\boldsymbol{a})$ becomes negligible. Thus, the opposite-phase solution approaches the optimal solution of the maxmin problem as well according to Proposition 3.

B. Network-Coded Relaying

The optimization problem is formulated as

$$\min_{\boldsymbol{A},\boldsymbol{B}} \quad \epsilon \tag{57a}$$

$$\text{s.t.} \quad \boldsymbol{R}_{ii} - 1 + \frac{\sigma^2}{|a_i|^2} = \epsilon, \; i = 1, \cdots, N, \tag{57b}$$

$$\Omega(\boldsymbol{A},\boldsymbol{B}) \le p, \tag{57c}$$

$$\epsilon \ge 0. \tag{57d}$$

The only difference between the equal-SNR problem and the maxmin problem (36) is that the inequality (36b) is replaced by the equality (57b). A proposition for the network-coded problem can be proved in the same way as Proposition 3. We then use the solution of the equal-SNR problem (57) to approximate that of the maxmin problem (36) at high SNR. The gap will be also evaluated in Section V.

PNC Identical-b Random-phase Algorithm simplifies the matter by introducing an extra constraint: $\boldsymbol{B} = b\boldsymbol{I}$, where b is a real scalar. We set a range for searching b. For each trial of b, we use essentially the same method as the non-PNC random-phase algorithm proposed in Section IV.A to find the phases θ_j of a_j and φ_j of b_j, and find the corresponding estimated ϵ^* as well. We then find b that yields the least post-processing noise power. The beamformer $\boldsymbol{G}$ can be calculated from $\boldsymbol{P}$, $|a_j|$, b, θ_j and φ_j.

PNC Phase-aligned Algorithm: Consider pairwise transmission, i.e., the case of N being an even number. Define

$$\boldsymbol{B}_\ell \triangleq \begin{bmatrix} b_{\ell 1} & 0 \\ 0 & b_{\ell 2} \end{bmatrix}, \tag{58}$$

where $b_{\ell 1}$ and $b_{\ell 2}$ are the diagonal elements $\pi(\ell)$ and $\kappa(\ell)$ of $\boldsymbol{B}_\ell$, respectively. The post-processing noise power can be factorized in terms of $b_{\ell 1}$ and $b_{\ell 2}$ as follows:

$$\epsilon = \gamma^2 h_{\ell 1} \left| b_{\ell 1} + \frac{h_{\ell 2}^*}{h_{\ell 1}} \right|^2 - \frac{\gamma^2 |h_{\ell 2}|^2}{h_{\ell 1}} + \gamma^2 h_{\ell 3} + \frac{\sigma^2}{|a_{\ell 1}|^2} \tag{59}$$

$$= \gamma^2 h_{\ell 3} \left| b_{\ell 2} + \frac{h_{\ell 2}}{h_{\ell 3}} \right|^2 - \frac{\gamma^2 |h_{\ell 2}|^2}{h_{\ell 3}} + \gamma^2 h_{\ell 1} + \frac{\sigma^2}{|a_{\ell 2}|^2}. \tag{60}$$

This formula shows the potential advantage of introducing $b_{\ell 1}$ and $b_{\ell 2}$, i.e., the application of physical-layer network coding. If we set $b_{\ell 1}$ and $b_{\ell 2}$ as

$$b_{\ell 1} = -\frac{h_{\ell 2}^*}{h_{\ell 1}}, \quad b_{\ell 2} = -\frac{h_{\ell 2}}{h_{\ell 3}}, \tag{61}$$

the post-processing noise power can be minimized in terms of $b_{\ell 1}$ and $b_{\ell 2}$. The relay power consumption can be written as

$$\Omega_2(\boldsymbol{A}) = \gamma^2 \mathcal{O}_2(\boldsymbol{A}) + \sum_{\ell=1}^{N/2} \left\{ |a_{\ell 1} + \lambda_\ell a_{\ell 2}|^2 \right.$$
$$+ \left(h_{\ell 1} + \frac{|h_{\ell 2}|^2}{h_{\ell 1}} - \gamma^2 |h_{\ell 2}|^2 + \gamma^2 h_{\ell 1} h_{\ell 3} - 1 \right) |a_{\ell 1}|^2 \tag{62}$$
$$+ \left. \left(h_{\ell 3} + \frac{|h_{\ell 2}|^2}{h_{\ell 3}} - \gamma^2 |h_{\ell 2}|^2 + \gamma^2 h_{\ell 1} h_{\ell 3} - \lambda_\ell^2 \right) |a_{\ell 2}|^2 \right\},$$

where

$$\lambda_\ell = |h_{\ell 2}|^2 \left[\gamma^2 \left(\frac{|h_{\ell 2}|^2}{h_{\ell 1} h_{\ell 3}} - 1 \right) - \frac{1}{h_{\ell 1}} - \frac{1}{h_{\ell 3}} \right].$$

Note that $\mathcal{O}_2(\boldsymbol{A})$ denotes the sum of the quadratic items of $\boldsymbol{A}$ across pairs. Since $\boldsymbol{H}^{-1}\boldsymbol{H}^{-H}$ is positive semidefinite, we can prove $|h_{\ell 2}|^2 \le h_{\ell 1} h_{\ell 3}$, then $\lambda_\ell < 0$. Thus, for fixed amplitudes of $a_{\ell 1}$ and $a_{\ell 2}$, the inner-pair part of (62) can be lowered when the phases of $a_{\ell 1}$ and $a_{\ell 2}$ are aligned, e.g., $a_{\ell 1}, \; a_{\ell 2} \ge 0$. Simulation results indicate that as long as the phases of $\boldsymbol{A}$'s diagonal elements are aligned within pairs, the throughput performance is not sensitive to the phase differences among different pairs.

Remark 2: Note that the phase-aligned solution is the optimal solution of the maxmin problem (36) when the noise level at the relay is zero. The result is obvious after setting γ to 0 in (36).

In summary, Remarks 1 and 2 show that when the noise at the relay is small, the opposite-phase solution approaches optimal for non-PNC case and the phase-aligned solution approaches optimal for the PNC case. Our simulation results validate this conclusion.

V. NUMERICAL RESULTS

In this section, we evaluate the throughputs of various designs of MIMO switches. We assume the maximum transmit power of the relay and every station are the same (thus $p = 1$), and the noise level at the relay and the stations is the same. Our simulation indicates that the system throughputs are roughly the same with different symmetric permutations. The same result can be concluded for asymmetric permutations. Thus, we use one permutation for each of them ($\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ are given by the matrices described in (19) and (20) with the diagonal elements set to 0) in simulations.

Observation 1: For equal-SNR zero-forcing (i.e., non-PNC) relaying, the optimal setting for the case of two stations ($N = 2$) has the property that the two elements of $\boldsymbol{a}$ have opposite signs. In general, the non-PNC opposite-phase algorithm is an effective scheme for pairwise transmission with larger even number of stations.

A similar framework as ours is investigated in [20], which focuses on optimizing the sum rate of all stations. Therein, a suboptimal beamforming scheme is proposed, which also uses zero-forcing detection and zero-forcing precoding, and simply uses a positive scalar weight to control the relay power consumption instead of our diagonal $\boldsymbol{A}$. We regard this scheme as a benchmark and call it "the basic scheme." All schemes proposed in this paper have an advantage over the basic scheme in that they guarantees fairness.

Compared with the basic scheme in Fig. 4, the optimal setting of $N = 2$, i.e., the non-PNC opposite-phase algorithm proposed in Section IV.A achieves more than 0.6 dB gain in the low SNR regime. The gain becomes smaller as the SNR increases, e.g., around 0.25 dB gain at the SNR of 15 dB. We explain why the gain diminishes for high SNR as follows. When the relay noise power is zero, $|a_1| = |a_2|$. In this case, s_{12} becomes 0 in (48). Then the throughput performance does not depend on the phase difference of a_1 and a_2. The opposite-sign setting is equivalent to the identical-gain setting, i.e., the basic scheme. Thus, the gain over the basic scheme becomes trivial in the high-SNR regime.
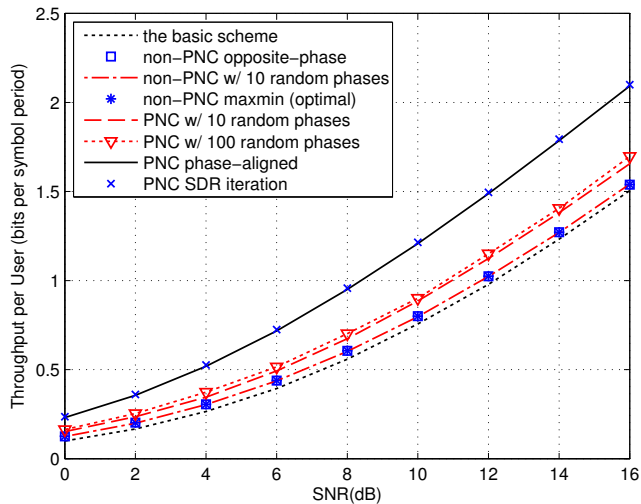
Fig. 4.   Throughput comparison of different relaying schemes in the case of two stations.
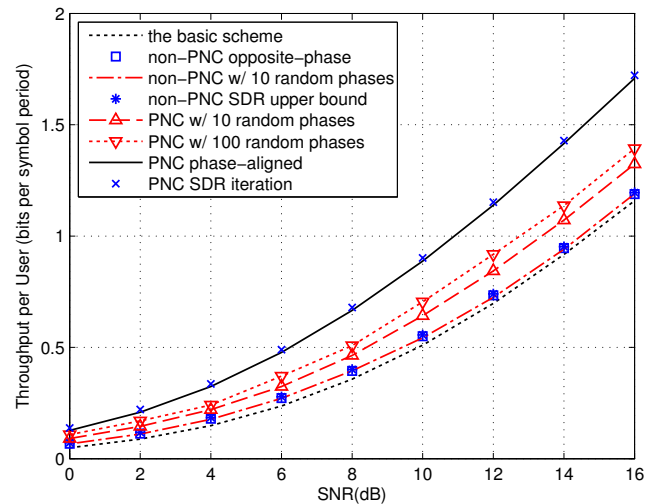

Fig. 5.   Throughput comparison of different relaying schemes for pairwise switching pattern in the case of four stations.

Fig. 5 presents the throughput in the case where $N = 4$ stations form two pairs for pairwise transmission. The throughputs are roughly the same when we vary the phase differences of the pairs while keeping the phase difference within each pair to π. (Experimentation with the phases is not shown in Fig. 5 to avoid cluttering.) With this result, we could set the elements of $\boldsymbol{a}$ such that one element in each pair is positive and the other element is negative. The results are similar to that of $N = 2$. The non-PNC opposite-phase algorithm achieves 0.8 dB gain in the low SNR regime and 0.25 dB gain over the basic scheme in the high SNR regime.

Observation 2: For equal-SNR relaying, physical-layer network coding can be applied as a relaying method to improve throughput performance significantly. For pairwise transmission, the PNC phase-aligned algorithm achieves significant gains over any other non-PNC scheme of zero-forcing relaying.

Simulation results on the PNC phase-aligned scheme are also presented in Fig. 4 and Fig. 5. When we apply physical-layer network coding in our MIMO switching, significant gains can be achieved over other schemes. The proposed PNC phase-aligned scheme outperforms all the other non-PNC schemes. Note in particular that compared with the basic scheme, it does not involve complicated calculations during the one-dimensional search of $\boldsymbol{A}$ and the setting of $\boldsymbol{B}$. However, the PNC phase-aligned scheme can not be applied to non-pairwise transmissions.

Observation 3: For equal-SNR zero-forcing relaying with the non-PNC random-phase algorithm, the simulation results indicate that large gains can be achieved with a small number of phase bins and trials. For network-coded relaying, the PNC identical-b random-phase algorithm can be applied for both pairwise and non-pairwise transmissions to achieve the network coding gain. It is worth mentioning that network coding helps not only for the traditional pairwise switching pattern but also for the non-pairwise pattern.

In Figs. 4 and 5, when $M = 8$ and $L = 10$, the non-PNC

random-phase scheme proposed in Section IV.A can achieve good enough throughput performance. With the PNC random-phase algorithm proposed in Section IV.B, the throughput performance is even better than that of the best non-PNC scheme. For $M = 8$, it achieves around 1.2 dB gain with $L = 10$ and 1.4 dB with $L = 100$ compared to the basic scheme in the case of two stations; it can achieve around 1.5 dB and 2 dB gains with 10 and 100 trials, respectively, over the basic scheme in the case of four stations. However, it needs larger M, L and more random trials to perform as good as the PNC phase-aligned scheme.

Consider the non-pairwise transmission in Fig. 6. Note that the non-PNC opposite-phase scheme and the PNC phase-aligned scheme can not be applied to non-pairwise transmissions. The non-PNC random-phase scheme and the PNC random-phase scheme can still outperform the basic scheme. For $M = 8$, when $L = 10$ the non-PNC random-phase scheme could achieve around 0.15 dB gain. The PNC random-phase scheme could achieve around 1.1 dB and 1.3 dB gains over the basic scheme with 10 and 100 trials, respectively.

Observation 4: In general, for zero-forcing (i.e., non-PNC) relaying, the throughput of the maxmin problem is larger than that of the equal-SNR problem. However, the throughput gap between the equal-SNR problem and the maxmin problem is small over a wide range of SNR.

We first evaluate the throughput performances with two stations. An interesting question is how large is the throughput gap between (30) and (46) corresponding to their optimal noise powers. In Fig. 4, the throughput gap between the two curves is very small. Note that the optimal solutions of the maxmin problem are found by exhaustive search. In Table I, the gap is evaluated over a wide range of SNR. The results indicate that the gap is less than 0.1% for the SNR regime from 0 to 30 dB.

In the case of four stations, the exhaustive search for solving the maxmin problem becomes computationally expensive. Therefore, we use the upper bound of the maxmin problem

TABLE I
WORST-STATION THROUGHPUTS OF THE MAXMIN PROBLEM (EXHAUSTIVE SEARCH & SDR) AND THROUGHPUTS OF THE EQUAL-SNR PROBLEM (NON-PNC OPPOSITE-PHASE & PNC PHASE-ALIGNED ALGORITHMS) WHEN $N = 2$; Δ DENOTES THE VARIATIONAL RATIO W.R.T. THE EQUAL-SNR SOLUTION OF EITHER NON-PNC OPPOSITE-PHASE OR PNC PHASE-ALIGNED ALGORITHMS, RESPECTIVELY.

| SNR | Non-PNC | | | PNC | |
| | Equal-SNR | Maxmin | | Equal-SNR | Maxmin |
(dB)	Opposite-phase	Optimal/$\Delta(\%)$	SDR/$\Delta(\%)$	Phase-aligned	SDR/$\Delta(\%)$
0	0.1270	0.1271 / 0.08	0.1270 / -0.00	0.2305	0.2363 / 2.52
10	0.7990	0.7994 / 0.05	0.7987 / -0.03	1.2091	1.2264 / 1.43
20	2.1249	2.1256 / 0.03	2.1202 / -0.22	2.7250	2.7549 / 1.10
30	3.7075	3.7083 / 0.02	3.6864 / -0.57	4.3762	4.4016 / 0.58

TABLE II
WORST-STATION THROUGHPUTS OF THE MAXMIN PROBLEM (SDR UPPER BOUND & SDR) AND THROUGHPUTS OF THE EQUAL-SNR PROBLEM (NON-PNC OPPOSITE-PHASE & PNC PHASE-ALIGNED ALGORITHMS) WHEN $N = 4$; Δ DENOTES THE VARIATIONAL RATIO W.R.T. THE EQUAL-SNR SOLUTION OF EITHER NON-PNC OPPOSITE-PHASE OR PNC PHASE-ALIGNED ALGORITHMS, RESPECTIVELY.

| SNR | Non-PNC | | | PNC | |
| | Equal-SNR | Maxmin | | Equal-SNR | Maxmin |
(dB)	Opposite-phase	SDR-Upper/$\Delta(\%)$	SDR/$\Delta(\%)$	Phase-aligned	SDR/$\Delta(\%)$
0	0.0661	0.0720 / 8.85	0.0720 / 8.85	0.1266	0.1376 / 8.69
10	0.5504	0.5562 / 1.05	0.5552 / 0.87	0.8842	0.9018 / 1.99
20	1.7314	1.7328 / 0.08	1.7302 / -0.08	2.3302	2.3435 / 0.57
30	3.2906	3.2918 / 0.04	3.2818 / -0.27	3.9611	3.9734 / 0.31

calculated by the SDR scheme for benchmarking instead. Generally, the throughput gap between the equal-SNR and the upper bound is small, and becomes even smaller in the high SNR regime in Fig. 5 and Table II. Since we use the upper bound for the maxmin solution for benchmarking, we conclude that the throughput gap between the maxmin problem and the equal-SNR problem is also small.

Therefore, the maxmin problem can be well approximated by the equal-SNR problem. We note that the trend as indicated by the simulations results in Fig. 4, Fig. 5, Table I and Table II is consistent with the analytical result of Proposition 3. In the high SNR regime, the noise becomes negligible, and the gap between equal-SNR and maxmin diminishes.

Observation 5: The throughput of the equal-SNR problem is roughly the same as the approximate throughput of the maxmin problem achieved by the SDR technique for pairwise switching pattern. However, for the non-pairwise pattern, the SDR technique achieves good throughput performance for both non-PNC and PNC relaying schemes.

With reference to Table I and Table II for pairwise switching pattern, for zero-forcing relaying our non-PNC opposite-phase algorithm is better than the SDR scheme in the high SNR regime, and for network-coded relaying the SDR scheme is better than our PNC phase-aligned algorithm. However, the gap is mostly smaller than 2% for SNR larger than 0 dB.

For the non-pairwise switching pattern, the random-phase scheme is close to the SDR upper bound for the non-PNC relaying. However, for the PNC relaying, the iterative SDR scheme outperforms the PNC identical-b random-phase algorithm since the latter needs more trials to achieve better throughput performance.
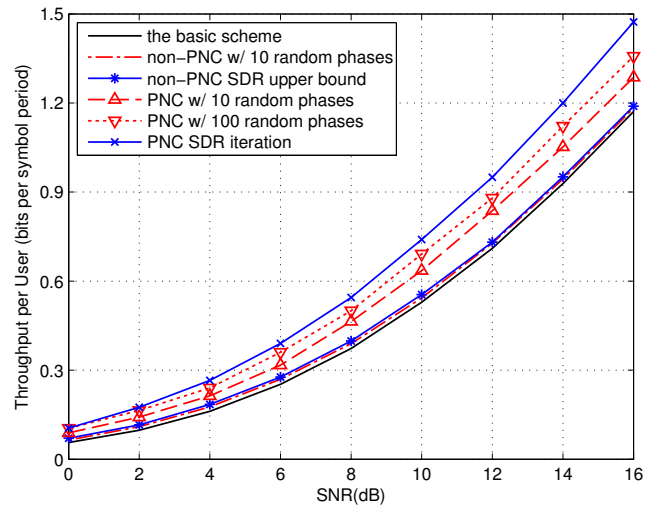

Fig. 6. Throughput comparison of different relaying schemes for non-pairwise switching pattern in the case of four stations.

The overall implications of our analytical and simulation results are as follows. The equal-SNR scheme, with the target of achieving perfect fairness among the links, is also a good approximation to the maxmin problem when the relay noise is small. Given a symmetric switch matrix that realizes pairwise transmissions, we could use the non-PNC opposite-phase algorithm or the non-PNC random-phase algorithm for zero-forcing relaying, and the PNC phase-aligned algorithm or the PNC identical-b random-phase algorithm for network-coded relaying, to identify a suitable gain vector. The PNC

phase-aligned algorithm has good throughput performance as well as fast execution time. Given an asymmetric switch matrix that realizes non-pairwise transmissions, we could use the random-phase algorithm for zero-forcing relaying and network-coded relaying to identify a suitable gain vector. The SDR scheme also achieves good throughput performance; however, its complexity is generally higher than the equal-SNR schemes.

VI. CONCLUSION

We have proposed a framework for wireless MIMO switching to facilitate communication among multiple wireless stations. With optimized precoders, network-coded relaying improves the throughput performance significantly over non-network-coded relaying.

The maxmin solution and the equal-SNR solution have their respective advantages. The former yields better throughput performance. However, the equal-SNR solution guarantees perfect fairness. With the equal-SNR solution, MIMO switching can be easily extended to multiple transmissions, by which general transmission patterns can be realized, including unicast, multicast, broadcast, or a mixture of them [18]. Moreover, the maxmin problem is NP-hard, which is solved using exhaustive search. Even approximating it with the SDR scheme using interior point methods has the complexity cost of $O(N^7)$ [17]. However, the proposed schemes for the equal-SNR problem has the complexity cost of $O(N^3)$ only, induced mainly by calculating matrix inverses. Hence, the equal-SNR setting is perhaps more amenable to practical deployment.

In future work, it will be interesting to explore switch matrices that realize more complicated patterns than unicast. It would also be interesting to study the case where the number of antennas at the relay is fewer than the number of stations.

REFERENCES

[1] D. Gunduz, A. Yener, A. Goldsmith, and H. V. Poor, "The multi-way relay channel," in *Proc. IEEE Int'l Symp. Inform. Theory*, 2009, pp. 339–343.

[2] Y. E. Sagduyu, D. Guo, and R. A. Berry, "Throughput optimal control for relay-assisted wireless broadcast with network coding," in *IEEE International Workshop on Wireless Network Coding (WiNC)*, San Francisco, CA, USA, Jun. 2008.

[3] Y. E. Sagduyu, R. A. Berry, and D. Guo, "Throughput and stability for relay-assisted wireless broadcast with network coding," *IEEE J. Sel. Area Commun.*, 2012, this issue.

[4] T. Cui, T. Ho, and J. Kliewer, "Space-time communication protocols for N-way relay networks," in *IEEE Global Telecommun. Conf.*, 2008, pp. 1–5.

[5] F. Gao, T. Cui, B. Jiang, and X. Gao, "On communication protocol and beamforming design for amplify-and-forward N-way relay networks," in *3rd IEEE Int'l Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, 2009.

[6] 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.

[7] 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 (Allerton)*, Sept. 2011, pp. 551 –558.

[8] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1204–1216, 2000.

[9] S.-Y. R. Li, R. W. Yeung, and N. Cai, "Linear network coding," *IEEE Trans. Inf. Theory*, vol. 49, no. 2, pp. 371–381, 2003.

[10] R. W. Yeung, S.-Y. R. Li, and N. Cai, *Network Coding Theory*. Now Publisher, 2006.

[11] S. Zhang, S. C. Liew, and P. P. Lam, "Hot topic: physical-layer network coding," in *Proc. of ACM Mobicom*, 2006.

[12] G. H. Golub and C. F. V. Loan, *Matrix Computations*. Johns Hopkins University Press, 1996.

[13] K. H. Hui, Y. E. Sagduyu, D. Guo, and R. A. Berry, "The maximum stable broadcast throughput for wireless line networks with network coding and topology control," in *Proc. Conf. Inform. Sci. Sys.*, Princeton, NJ, USA, Mar. 2010, pp. 1–6.

[14] 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.

[15] C. K. Ho, K. T. Gowda, and S. Sun, "Relaying for pair-wise information exchange," in *Proc. IEEE Inf. Theory Workshop*, Taormina, Sicily, Italy, Oct. 2009, pp. 421–425.

[16] R. Wang and M. Tao, "Joint source and relay precoding designs for MIMO two-way relay systems," in *Proc. IEEE Int'l Conf. on Commun. (ICC)*, Jun. 2011, pp. 1 –5.

[17] N. D. Sidiropoulos, T. N. Davidson, and Z.-Q. Luo, "Transmit beamforming for physical-layer multicasting," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2239–2251, 2006.

[18] F. Wang and S. C. Liew, "Wireless MIMO switching," Apr. 2011. [Online]. Available: http://arxiv.org/abs/1104.4035v1

[19] Wikipedia, "Quartic function — Wikipedia, the free encyclopedia," 2011. [Online]. Available: http://en.wikipedia.org/wiki/Quartic-equation

[20] A. U. T. Amah and A. Klein, "Non-regenerative multi-way relaying with linear beamforming," in *Proc. 20th IEEE Int'l Symp. on Personal, Indoor and Mobile Radio Commun.*, 2009.
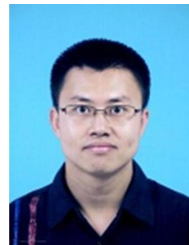
Fanggang Wang (S'10-M'11) received the B.S. degree in 2005 and the Ph.D. degree in 2010 from the School of Information and Communication Engineering at Beijing University of Posts and Telecommunications, Beijing, China. From 2008 to 2010, he worked as a Visiting Scholar in Electrical Engineering Department, Columbia University, New York City, New York, USA. Since 2010, he has been working in the State Key Lab of Rail Traffic Control and Safety, School of Electronic and Information Engineering, Beijing Jiaotong University as an Assistant Professor, and also in Institute of Network Coding, The Chinese University of Hong Kong as a Postdoctoral Fellow. In 2011, he visited Department of Electrical Engineering & Computer Science, Northwestern University, Evanston, IL, USA. His research interests are in the area of MIMO, OFDM and network coding techniques in wireless communications.

He chaired two workshops on wireless network coding (NRN 2011 & NRN 2012) and served as TPC member in several conferences.

Soung Chang Liew received his S.B., S.M., E.E., and Ph.D. degrees from the Massachusetts Institute of Technology. From 1984 to 1988, he was at the MIT Laboratory for Information and Decision Systems, where he investigated Fiber-Optic Communications Networks. From March 1988 to July 1993, he was at Bellcore (now Telcordia), New Jersey, where he engaged in Broadband Network Research. He has been a Professor at the Department of Information Engineering, the Chinese University of Hong Kong, since 1993. He is an Adjunct Professor at Peking University and Southeast University, China.

Prof. Liew's current research interests include wireless networks, Internet protocols, multimedia communications, and packet switch design. Prof. Liew's research group won the best paper awards in IEEE MASS 2004 and IEEE WLN 2004. Separately, TCP Veno, a version of TCP to improve its performance over wireless networks proposed by Prof. Liew's research group, has been incorporated into a recent release of Linux OS. In addition, Prof. Liew initiated and built the first inter-university ATM network testbed in Hong Kong in 1993. More recently, Prof. Liew's research group pioneers the concept of Physical-layer Network Coding (PNC).

Besides academic activities, Prof. Liew is also active in the industry. He co-founded two technology start-ups in Internet Software and has been serving as a consultant to many companies and industrial organizations. He is currently consultant for the Hong Kong Applied Science and Technology Research Institute (ASTRI), providing technical advice as well as helping to formulate R&D directions and strategies in the areas of Wireless Internetworking, Applications, and Services.

Prof. Liew is the holder of eight U.S. patents and a Fellow of IEEE, IET and HKIE. He currently serves as Editor for IEEE Transactions on Wireless Communications and Ad Hoc and Sensor Wireless Networks. He is the recipient of the first Vice-Chancellor Exemplary Teaching Award at the Chinese University of Hong Kong. Publications of Prof. Liew can be found in www.ie.cuhk.edu.hk/soung.
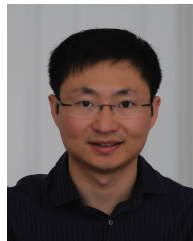
Dongning Guo (S'97-M'05-SM'11) joined the faculty of Northwestern University, Evanston, IL, in 2004, where he is currently an Associate Professor in the Department of Electrical Engineering and Computer Science. He received the B.Eng. degree from the University of Science & Technology of China, the M.Eng. degree from the National University of Singapore, and the M.A. and Ph.D. degrees from Princeton University, Princeton, NJ. He was an R&D Engineer in the Center for Wireless Communications (now the Institute for Infocom Research), Singapore, from 1998 to 1999. He has held visiting positions at Norwegian University of Science and Technology in summer 2006 and in the Institute of Network Coding at the Chinese University of Hong Kong in 2010–2011. He is an Associate Editor of the IEEE Transactions on Information Theory and an Editor of Foundations and Trends in Communications and Information Theory.

Dongning Guo received the Huber and Suhner Best Student Paper Award in the International Zurich Seminar on Broadband Communications in 2000 and is a co-recipient of the IEEE Marconi Prize Paper Award in Wireless Communications in 2010 (with Y. Zhu and M. L. Honig). He is also a recipient of the National Science Foundation Faculty Early Career Development (CAREER) Award in 2007. His research interests are in information theory, communications, and networking.