ISIT'2013 1569717915

Asymptotic Sum-Capacity of MIMO Two-Way Relay Channels within $\frac{1}{2}\log\frac{5}{4}$ Bit per User-Antenna

Xiaojun Yuan* and Tao Yang†
*Institute of Network Coding, Dept. of Information Engineering, The Chinese University of Hong Kong, Hong Kong
†Wireless Networking and Technology Lab, CSIRO ICT Centre, Australia

***Abstract*—We propose a novel space-division based network-coding scheme for multiple-input multiple-output (MIMO) two-way relay channels (TWRCs), in which two multi-antenna users exchange information via a multi-antenna relay. In the proposed scheme, the overall signal space at the relay is divided into two subspaces. In one subspace, the spatial streams of the two users have nearly orthogonal directions, and are completely decoded at the relay. In the other subspace, the signal directions of the two users are nearly parallel, and linear functions of the spatial streams are computed at the relay, following the principle of physical-layer network coding (PNC). Based on the recovered messages and message-functions, the relay generates and forwards network-coded messages to the two users. We show that, at high signal-to-noise ratio (SNR), the proposed scheme with optimized precoding achieves the asymptotic sum-rate capacity of MIMO TWRCs within $\frac{1}{2}\log(5/4) \approx 0.161$ bit per user-antenna, for any antenna configuration and any channel realization.**

I. Introduction

Physical-layer network coding (PNC) [1] is an emerging technique for wireless relay networks. The simplest model for PNC is a two-way relay channel (TWRC), in which two users A and B exchange information via an intermediate relay. Compared with conventional schemes, PNC allows the relay to deliver linear functions of the users' messages, without completely decoding the two users' messages. This can potentially double the network throughput. It has been shown that the PNC scheme can achieve the capacity of a Gaussian TWRC within 1/2 bit per user [2], and its gap to the capacity vanishes at high signal-to-noise ratio (SNR).

Multiple-input multiple-output (MIMO) technique can significantly boost the spectral efficiency of the communication systems. Recently, efficient communications over MIMO TWRCs have attracted much research interest, where the two users and the relay are all equipped with multiple antennas. Most existing work on MIMO TWRCs focuses on classical relaying strategies borrowed from one-way relay channels, such as amplify-and-forward (AF) [3][4], compress-and-forward [5][6], and decode-and-forward (DF) [7][8]. These strategies generally perform well away from the channel capacity due to noise amplification and multiplexing loss. Recently, several relaying schemes have been proposed to support PNC in MIMO TWRCs [9]-[14]. The basic idea in [9] and [10] is to jointly decompose the channel matrices of the two users to create multiple scalar channels, over which multiple PNC streams are transmitted. Let n_A, n_B, and n_R denote the numbers of antennas of user A, user B, and the relay, respectively. For configurations with $n_A, n_B \geq n_R$, a generalized singular-value-decomposition (GSVD) scheme was shown to achieve the asymptotic capacity of MIMO TWRCs at high SNR [9]. For configurations with $n_A, n_B < n_R$, all existing schemes may perform quite far from the capacity. Such configurations, however, are of most practical importance. For example, due to the constrained physical sizes of the user terminals, it is usually convenient to implement more antennas at the relay station than at the user ends, as suggested in the standards of next generation networks [15][16].

In this paper, we propose a new space-division based PNC scheme for MIMO TWRCs. In the proposed scheme, the overall signal space of the two users is divided into two subspaces. In one subspace, the channel directions of one user are orthogonal (or close to orthogonal) to those of the other user. In this subspace, the spatial streams of the two users are completely decoded. In the other subspace, the channel directions of the two users are parallel or close to parallel. In this subspace, linear functions of the corresponding spatial streams are computed, without completely decoding the individual spatial streams. We develop the optimal precoding strategy that maximizes the high-SNR achievable sum-rate for the proposed SD scheme. We then analytically show that, as SNR tends to infinity, the proposed scheme achieves the sum-capacity of the MIMO TWRC within $\min\{n_A, n_B\}\log(5/4)$ bits, or alternatively, $\frac{1}{2}\log(5/4) \approx 0.161$ bit/user-antenna, for any antenna setup and any channel realization. This gap is much smaller than the gap for the existing schemes, e.g., in [9] and [23].

II. System Model

We consider a discrete memoryless MIMO TWRC in which users A and B exchange information via a relay. User m is equipped with n_m antennas, $m \in \{A, B\}$, and the relay with n_R antennas. We assume that there is no direct link between the two users. The channel is assumed to be flat-fading and quasi-static, i.e., the channel coefficients remain unchanged during each round of information exchange.

The system operates in a half-duplex mode. Two equal-duration time slots are employed for each round of information exchange. In the first time-slot (referred to as *uplink phase*), the two users transmit to the relay simultaneously and the relay remains silent. The transmit signal matrix at user m is denoted by $\mathbf{X}_m \in \mathbb{C}^{n_m \times T}$, $m \in \{A, B\}$, where T is the number of channel uses in one time-slot. Each column of $\mathbf{X}_m$ denotes the signal vector transmitted by the n_m antennas in one channel use. The average power at each user is constrained as $\frac{1}{T}E\left[\|\mathbf{X}_m\|_F^2\right] \leq P_m$, $m \in \{A, B\}$. The received signal at the relay is denoted by $\mathbf{Y}_R \in \mathbb{C}^{n_R \times T}$ with

$$\mathbf{Y}_R = \mathbf{H}_A \mathbf{X}_A + \mathbf{H}_B \mathbf{X}_B + \mathbf{Z}_R, \tag{1}$$

where $\mathbf{H}_m \in \mathbb{C}^{n_R \times n_m}$ represents the channel from user m to the relay, and $\mathbf{Z}_R \in \mathbb{C}^{n_R \times T}$ denotes the additive white Gaussian noise (AWGN) at the relay. We assume that the channel matrices are always of full column rank, and are globally known by both users as well as by the relay. Upon receiving $\mathbf{Y}_R$, the relay generates a signal matrix $\mathbf{X}_R \in \mathbb{C}^{n_R \times T}$.

In the second time-slot (referred to as *downlink phase*), $\mathbf{X}_R$ is broadcast to the two users. Upon signal reception, user *A* (or *B*) recovers the user *B*'s (or *A*'s) message with the help of perfect knowledge of $\mathbf{X}_A$ (or $\mathbf{X}_B$).

III. Proposed Space-Division Approach

A. Space-Division Approach for MIMO TWRC

The proposed space-division approach is motivated by the observation that, in a MIMO TWRC, one user's potential signal directions (from which the user signal impinging upon the relay antenna array) may be either (close to) parallel or (close to) orthogonal to the potential signal directions of the other user. It is known that, when the signals of the two users are close to parallel, PNC can approach the capacity of TWRC [9]; on the other hand, when the signals of the two users are close to orthogonal, inter-user interference is significantly mitigated, and so complete decoding at the relay yields a higher rate over PNC.

With the above motivation, we construct the transmit signal of user m as

$$\mathbf{X}_m = \mathbf{F}_m^{CD} \mathbf{C}_m^{CD} + \mathbf{F}_m^{PNC} \mathbf{C}_m^{PNC}, m \in \{A, B\} \tag{2}$$

where $\mathbf{C}_m^{CD} \in \mathbb{C}^{(n_m - k) \times T}$ and $\mathbf{C}_m^{PNC} \in \mathbb{C}^{k \times T}$ are codeword matrices (with the elements having zero mean and unit variance) for the complete-decoding (CD) and PNC signal streams, respectively; $\mathbf{F}_m^{CD} \in \mathbb{C}^{n_m \times (n_m - k)}$ and $\mathbf{F}_m^{PNC} \in \mathbb{C}^{n_m \times k}$ are respectively the precoding matrices for CD and PNC signal streams. Here, k represents the number of spacial streams allocated for PNC. With a proper choice of $\{k, \mathbf{F}_m^{PNC}, \mathbf{F}_m^{CD}\}$, user m may choose any subspace of the column space of $\mathbf{H}_m$, denoted by $\mathcal{C}(\mathbf{H}_m)$, for transmitting the PNC signal $\mathbf{C}_m^{PNC}$ or the CD signal $\mathbf{C}_m^{CD}$; hence the name space division (SD).

At the relay, the CD signals are first decoded by treating the PNC signals as interference. After that, the CD signals are canceled from the received signal, and then the relay performs network-decoding for the PNC signals.

B. Achievable Rates

We now study the achievable rates of the proposed scheme. From [9], the downlink achievable rates are the same as the cut-set outer bound. Thus, we focus on the achievable rates of the uplink.

From (1) and (2), the signals $\mathbf{C}_A^{CD}$ and $\mathbf{C}_B^{CD}$ are completely decoded by treating $\mathbf{C}_A^{PNC}$ and $\mathbf{C}_B^{PNC}$ as interference. The corresponding signal model is a standard Gaussian vector multiple-access channel [17], with the achievable sum rate given by (3).

After decoding and cancelation of $\mathbf{C}_A^{CD}$ and $\mathbf{C}_B^{CD}$, the received signal at the relay can be simplified as

$$\widetilde{\mathbf{Y}}_R = \mathbf{H}_A \mathbf{F}_A^{PNC} \mathbf{C}_A^{PNC} + \mathbf{H}_B \mathbf{F}_B^{PNC} \mathbf{C}_B^{PNC} + \mathbf{Z}_R. \tag{4}$$

Without loss of generality, let

$$\mathbf{F}_m^{PNC} = \widetilde{\mathbf{F}}_m^{PNC} \mathbf{G}_m^{PNC} \tag{5}$$

where $\widetilde{\mathbf{F}}_m^{PNC} \in \mathbb{C}^{n_m \times k}$ gives an orthogonal basis of $\mathcal{C}(\mathbf{F}_m^{PNC})$, and $\mathbf{G}_A^{PNC} \in \mathbb{C}^{k \times k}$, $m \in \{A, B\}$.

The PNC channel spaces of the two users, denoted by $\mathcal{C}(\mathbf{H}_m \widetilde{\mathbf{F}}_m^{PNC})$, $m \in \{A, B\}$, are in general different to each other. Following the reduced-dimension approach in [23], we project the two equivalent channel matrices into a same subspace spanned by $\mathbf{P} \in \mathbb{C}^{n_R \times k}$. Without loss of generality, we assume the columns of $\mathbf{P}$ form an orthonomal basis, i.e., $\mathbf{P}^\dagger \mathbf{P} = \mathbf{I}$. Then, after projection, the resulting signal model is written as

$$\mathbf{P}^\dagger \widetilde{\mathbf{Y}}_R = \widetilde{\mathbf{H}}_A^{PNC} \mathbf{G}_A^{PNC} \mathbf{C}_A^{PNC} + \widetilde{\mathbf{H}}_B^{PNC} \mathbf{G}_B^{PNC} \mathbf{C}_B^{PNC} + \mathbf{P}^\dagger \mathbf{Z}_R \tag{6}$$

where † represents conjugate transpose, and

$$\widetilde{\mathbf{H}}_m^{PNC} = \mathbf{P}^\dagger \mathbf{H}_m \widetilde{\mathbf{F}}_m^{PNC}, \ m \in \{A, B\}. \tag{7}$$

Note that $\widetilde{\mathbf{H}}_m^{PNC}$ is a k-by-k square matrix. The efficient PNC design for k-by-k MIMO TWRC has been discussed in [9]-[13]. Applying the generalized singular-value decomposition (GSVD) [21] to $\widetilde{\mathbf{H}}_m^{PNC}$, we obtain

$$\widetilde{\mathbf{H}}_m^{PNC} = \mathbf{B} \mathbf{\Sigma}_m \mathbf{T}_m^\dagger, m \in \{A, B\} \tag{8}$$

where $\mathbf{B} \in \mathbb{C}^{k \times k}$ is a nonsingular matrix, $\mathbf{T}_m \in \mathbb{C}^{k \times k}$ is an orthogonal matrix, and $\mathbf{\Sigma}_m \in \mathbb{C}^{k \times k}$ is a diagonal matrix with the ith diagonal element denoted by $\Sigma_{m;i}$. We further take the QR decomposition to the matrix $\mathbf{B}$, yielding

$$\widetilde{\mathbf{H}}_m^{PNC} = \mathbf{Q} \mathbf{R} \mathbf{\Sigma}_m \mathbf{T}_m^\dagger, m \in \{A, B\} \tag{9}$$

where $\mathbf{Q} \in \mathbb{C}^{k \times k}$ is unitary, and $\mathbf{R} \in \mathbb{C}^{k \times k}$ is an upper triangular matrix. Then, $\mathbf{G}_m^{PNC}$ is designed as

$$\mathbf{G}_m^{PNC} = \mathbf{T}_m \mathbf{\Psi}_m^{1/2}, m \in \{A, B\} \tag{10}$$

where $\mathbf{\Psi}_m^{1/2} = \mathrm{diag}\{\sqrt{\psi_{m;1}}, \sqrt{\psi_{m;2}}, \cdots, \sqrt{\psi_{m;k}}\}$ is a diagonal matrix with $\psi_{m;i} \geq 0, i = 1, 2, ..., k$.

Let R_m^{PNC} be the total rate of the PNC spatial streams of user m, $m \in \{A, B\}$. From Theorem 1 in [9], the achievable rate-pair of the PNC spacial streams is given by

$$R_m^{PNC} = \sum_{i=1}^{k} \frac{1}{2} \left[\log \left(\frac{I(i) \Sigma_{m;i}^2 \psi_{m;i}}{\Sigma_{A;i}^2 \psi_{A;i} + \Sigma_{B;i}^2 \psi_{B;i}} + \frac{r_{i,i}^2 \Sigma_{m;i}^2 \psi_{m;i}}{N_0} \right) \right]^+ \tag{11}$$

$m \in \{A, B\}$, where $I(i)$ is the indicator function with $I(i) = 1$ for $i = 1$ and $I(i) = 0$ for $i \neq 1$.

C. Formulation of Sum-Rate Maximization

Now we consider the optimization of the precoding matrices in (2) to maximize the achievable sum rate (denoted by R). This problem can be formulated as follows.

$$\text{maximize} \quad R = R^{CD} + \sum_{m \in \{A, B\}} R_m^{PNC} \tag{12a}$$

$$\text{subject to} \quad \mathrm{tr}\{\mathbf{F}_m^{CD} (\mathbf{F}_m^{CD})^\dagger\} + \sum_{i=1}^{k} \psi_{m;i} \leq P_m \tag{12b}$$

$$\mathbf{P}^\dagger \mathbf{P} = \mathbf{I}, (\widetilde{\mathbf{F}}_m^{PNC})^\dagger \widetilde{\mathbf{F}}_m^{PNC} = \mathbf{I}, m \in \{A, B\}. \tag{12c}$$

$$R^{CD} = \frac{1}{2}\log \frac{\left| N_0\mathbf{I} + \sum_{m\in\{A,B\}} \mathbf{H}_m(\mathbf{F}_m^{CD}(\mathbf{F}_m^{CD})^\dagger + \mathbf{F}_m^{PNC}(\mathbf{F}_m^{PNC})^\dagger)\mathbf{H}_m^\dagger \right|}{\left| N_0\mathbf{I} + \sum_{m\in\{A,B\}} \mathbf{H}_m\mathbf{F}_m^{PNC}(\mathbf{F}_m^{PNC})^\dagger\mathbf{H}_m^\dagger \right|} \tag{3}$$

In the above, R^{CD} is given by (3), R_m^{PNC} is given by (11), and (12b) represents the power constraints. This problem involves the joint optimization of $k, \mathbf{F}_A^{CD}, \mathbf{F}_B^{CD}, \widetilde{\mathbf{F}}_A^{PNC}, \widetilde{\mathbf{F}}_B^{PNC}$, $\mathbf{P}$, $\{\psi_{A;i}\}_{i=1}^k$, and $\{\psi_{B;i}\}_{i=1}^k$, which is in general difficult to solve. We will next determine the optimal solution to Problem (12) in the high-SNR regime.

IV. Asymptotic Sum-Rate Analysis

In this section, we analyze the uplink achievable sum-rate as the SNRs, i.e., $\frac{P_A}{N_0}$ and $\frac{P_B}{N_0}$, tend to infinity. We start with a brief description of principle angles and vectors of two subspaces.

The orthogonality between two subspaces are characterized by the principle angles [21]. By definition, the first principle angle of two subspaces is the smallest angle between a pair of vectors, one from each subspace, that are mostly close to parallel to each other; and these two vectors are called the first pair of principle vectors. Then, the ith principle angle is recursively defined as the smallest angle between two *most parallel* vectors, one from each subspace, that are orthogonal to the first $i-1$ pairs of principle vectors. The principle angles and principle vectors can be calculated as follows.

Let the QR decomposition of $\mathbf{H}_m$ be

$$\mathbf{H}_m = \mathbf{Q}_m\mathbf{R}_m, m \in \{A,B\} \tag{13}$$

where $\mathbf{Q}_m \in \mathbb{C}^{n_R\times n_m}$ satisfies $\mathbf{Q}_m\mathbf{Q}_m^\dagger = \mathbf{I}$, and $\mathbf{R}_m \in \mathbb{C}^{n_m\times n_m}$. Then, the singular value decomposition (SVD) of $\mathbf{Q}_A\mathbf{Q}_B^\dagger$ is generally written as

$$\mathbf{Q}_A^\dagger\mathbf{Q}_B = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\dagger \tag{14}$$

where $\mathbf{U} \in \mathbb{C}^{n_A\times n_A}$ and $\mathbf{V} \in \mathbb{C}^{n_B\times n_B}$ are both unitary matrices, and $\mathbf{\Sigma} = \mathrm{diag}\{\sigma_1, \ldots, \sigma_{\min\{n_A,n_B\}}\}$ with the diagonal elements arranged in the *descending* order. It is known that the ith smallest principle angle of $\mathcal{C}(\mathbf{H}_A)$ and $\mathcal{C}(\mathbf{H}_B)$ is given by $\theta_i = \arccos\sigma_i$ [21]. The corresponding principle vectors are given by $\mathbf{Q}_A\mathbf{u}_i$ and $\mathbf{Q}_B\mathbf{v}_i$, where $\mathbf{u}_i$ and $\mathbf{v}_i$ are respectively the ith column of $\mathbf{U}$ and $\mathbf{V}$.

We are now ready to present the main result of this paper. Define

$$R^{UB} = \frac{1}{2}\sum_{m\in\{A,B\}} \log\left|\frac{P_m}{N_0 n_m}\mathbf{H}_m^\dagger\mathbf{H}_m\right| \tag{15}$$

which is known as the uplink cut-set upper bound of the MIMO TWRC at high SNR [9]. Also define

$$\Delta^{SD} \triangleq -\log\prod_{i=1}^{k}\frac{\sigma_i+1}{2} - \log\prod_{i=k+1}^{\min(n_A,n_B)}\sqrt{1-\sigma_i^2}. \tag{16}$$

Denote by $\mathbf{A}_{(i,j)}$ the submatrix of $\mathbf{A}$ consisting of the ith-to-jth columns of $\mathbf{A}$, and by $\mathbf{A}_{\overline{(i,j)}}$ the submatrix of $\mathbf{A}$ obtained by excluding the ith-to-jth columns. Then

Theorem 1: As $N_0 \to 0$, the minimum gap between the achievable sum-rate of the proposed SD scheme and the sum-capacity upper bound R^{UB} is given by

$$R^{UB} - R \xrightarrow{N_0\longrightarrow 0} \Delta^{SD}, \tag{17}$$

which is achieved when

$$\mathbf{P}: \quad \mathbf{p}_i = \frac{\mathbf{Q}_A\mathbf{u}_i + \mathbf{Q}_B\mathbf{v}_i}{\|\mathbf{Q}_A\mathbf{u}_i + \mathbf{Q}_B\mathbf{v}_i\|}, \text{ for } i=1,\ldots,k \tag{18a}$$

$$\widetilde{\mathbf{F}}_A^{PNC} = \mathbf{R}_A^{-1}\mathbf{U}_{(1,k)}(\mathbf{U}_{(1,k)}^\dagger(\mathbf{R}_A\mathbf{R}_A^\dagger)^{-1}\mathbf{U}_{(1,k)})^{-\frac{1}{2}} \tag{18b}$$

$$\widetilde{\mathbf{F}}_B^{PNC} = \mathbf{R}_B^{-1}\mathbf{V}_{(1,k)}(\mathbf{V}_{(1,k)}^\dagger(\mathbf{R}_B\mathbf{R}_B^\dagger)^{-1}\mathbf{V}_{(1,k)})^{-\frac{1}{2}} \tag{18c}$$

$$\psi_{m,i} = \frac{P_m}{n_m}, \text{ for } i=1,\ldots,k \tag{18d}$$

$$\mathbf{F}_m^{CD} = \sqrt{\frac{P_m}{n_m}}\widetilde{\mathbf{F}}_m^{PNC\perp}, \; m \in \{A,B\} \tag{18e}$$

where $\mathbf{p}_i$ is the ith column of $\mathbf{P}$, and $\widetilde{\mathbf{F}}_m^{PNC\perp} \in \mathbb{C}^{n_m\times(n_m-k)}$ satisfies $(\widetilde{\mathbf{F}}_m^{PNC\perp})^\dagger\widetilde{\mathbf{F}}_m^{PNC\perp} = \mathbf{I}$ and $(\widetilde{\mathbf{F}}_m^{PNC\perp})^\dagger\widetilde{\mathbf{F}}_m^{PNC} = \mathbf{0}$.

Remark 1: Theorem 1 can be interpreted as follows. (18b) (or (18c)) implies $\mathcal{C}(\mathbf{H}_A\widetilde{\mathbf{F}}_A^{PNC}) = \mathcal{C}(\mathbf{Q}_A[\mathbf{u}_1,\ldots,\mathbf{u}_k])$ (or $\mathcal{C}(\mathbf{H}_B\widetilde{\mathbf{F}}_B^{PNC}) = \mathcal{C}(\mathbf{Q}_B[\mathbf{v}_1,\ldots,\mathbf{v}_k])$), that is, the PNC signal of user A (or B) falls into the subspace spanned by the principle vectors of $\mathbf{H}_A$ (or $\mathbf{H}_B$) associated with the k smallest principle angles. (18a) implies that the column space of the projection matrix $\mathbf{P}$ is spanned by the angular bisectors of the principle vector pairs $\{\mathbf{u}_i, \mathbf{v}_i\}$, for $i=1,\ldots,k$. Moreover, (18d) reveals that equal power allocation is asymptotically optimal in the high SNR regime, which is similar to the case of point-to-point MIMO communications.

Remark 2: The first term in (16), namely, $-\log\prod_{i=1}^{k}\frac{\sigma_i+1}{2}$, is the rate loss incurred by the projection operation of PNC; and the remaining term is the rate loss incurred by the complete-decoding operation.

Remark 3: For the case of $n_A = n_B = n_R$, we have $\sigma_i = 1$ for $i = 1, ..., n_R$. Then, from (16), we choose $k = n_R$ spacial streams for PNC (and none for CD), which leads to $\Delta^{SD} = 0$, meaning that our proposed scheme is asymptotically optimal for $n_A = n_B = n_R$. This agrees with the fact that our proposed scheme reduces to the GSVD-PNC scheme which is indeed asymptotically optimal at high SNR [9].

Corollary 1: The optimal k to minimize the rate gap Δ^{SD} in (16) satisfies $1 \geq \sigma_1 \geq ... \geq \sigma_k \geq \frac{3}{5} > \sigma_{k+1} \geq ... \geq \sigma_{\min(n_A,n_B)} \geq 0$. With the optimal choice of k, the asymptotic rate gap Δ^{SD} is at most $\min(n_A,n_B)\log(5/4)$ bits, or $\frac{1}{2}\log(5/4)$ bits per user-antenna, which occurs when $\sigma_i = \frac{3}{5}$, for $i = 1,\ldots,\min(n_A,n_B)$.

Proof: The proof follows directly from (16) and (17), together with the fact that the minimum of $\max\{\frac{x+1}{2}, \sqrt{1-x^2}\}$ in $x \in [0,1]$ is $4/5$, which is achieved at $x = 3/5$. ■

V. Conclusions and Discussions

In this paper, we proposed a space-division approach for efficient communications over MIMO TWRCs. The proposed approach exploits the benefits provided by both PNC and complete-decoding strategies, but avoids their disadvantages. The asymptotically optimal design of the proposed scheme was established. We showed that the asymptotic gap to the sum-capacity is at most $\frac{1}{2}\log(5/4) \approx 0.161$ bits per user-antenna.

Our further analysis revealed that the average asymptotic gap over fading channels is much smaller than the above worst-case gap. We also studied the optimization of the proposed SD scheme in the finite SNR regime. Numerical results demonstrate that the SD scheme can approach the capacity cut-set out bound very closely throughout the entire practical SNR range. These results are not presented in this paper due to space limitation. Interested readers may refer to [24] for details.

Appendix A
Proof of Theorem 1

Proof: We first consider the PNC achievable rate. As $N_0 \to 0$, the PNC achievable rate of user m is

$$R_m^{PNC} = \sum_{i=1}^{k} \frac{1}{2}\log\left(\frac{r_{i,i}^2 \Sigma_{m;i}^2 \psi_{m;i}}{N_0}\right) \tag{19a}$$

$$= \frac{1}{2}\log\left|\frac{1}{N_0}\mathbf{R}\mathbf{\Sigma}_m^2\mathbf{R}^\dagger\right| + \frac{1}{2}\log|\Psi_m| \tag{19b}$$

$$= \frac{1}{2}\log\left|\frac{1}{N_0}\widetilde{\mathbf{H}}_m^{PNC}(\widetilde{\mathbf{H}}_m^{PNC})^\dagger\right| + \frac{1}{2}\log|\Psi_m| \tag{19c}$$

where (19a) follows from (11), (19b) from the fact that $\mathbf{R}$ is square and upper-triangular, and (19c) from (9). From (7) and (13), we obtain

$$\widetilde{\mathbf{H}}_m^{PNC} = \mathbf{P}^\dagger\mathbf{Q}_m\mathbf{R}_m\widetilde{\mathbf{F}}_m^{PNC}. \tag{20}$$

Let $\mathbf{V}_m^{PNC} \in \mathbb{C}^{n_m \times k}$ be an orthonomal matrix with the columns span $\mathcal{C}(\mathbf{R}_m\widetilde{\mathbf{F}}_m^{PNC})$. By definition, $(\mathbf{V}_m^{PNC})^\dagger\mathbf{V}_m^{PNC} = \mathbf{I}$. Then, as $N_0 \to 0$, R_m^{PNC} in (19) becomes

$$R_m^{PNC} = \Delta_{m;1} + \Delta_{m,2} + \frac{1}{2}\log|\Psi_m| \tag{21}$$

where

$$\Delta_{m,1} = \frac{1}{2}\log\left|\mathbf{P}^\dagger\mathbf{Q}_m\mathbf{V}_m^{PNC}(\mathbf{Q}_m\mathbf{V}_m^{PNC})^\dagger\mathbf{P}\right| \tag{22a}$$

$$\Delta_{m,2} = \frac{1}{2}\log\left|\frac{1}{N_0}(\widetilde{\mathbf{F}}_m^{PNC})^\dagger\mathbf{R}_m^\dagger\mathbf{R}_m\widetilde{\mathbf{F}}_m^{PNC}\right|. \tag{22b}$$

Now we consider the achievable rates of the CD signals. Similarly to (5), we can express

$$\mathbf{F}_m^{CD} = \widetilde{\mathbf{F}}_m^{CD}\mathbf{G}_m^{CD} \tag{23}$$

where $\widetilde{\mathbf{F}}_m^{CD} \in \mathbb{C}^{n_m\times(n_m-k)}$ gives an orthonormal basis of $\mathcal{C}(\mathbf{F}_m^{CD})$, and $\mathbf{G}_m^{CD} \in \mathbb{C}^{(n_m-k)\times(n_m-k)}$, $m \in \{A,B\}$. Let $\mathbf{V}_m^{CD} \in \mathbb{C}^{n_m\times(n_m-k)}$ be an orthonormal basis of the null space of $\mathcal{C}(\mathbf{R}_m\widetilde{\mathbf{F}}_m^{PNC})$. By definition, $(\mathbf{V}_m^{CD})^\dagger\mathbf{V}_m^{CD} = \mathbf{I}$ and $(\mathbf{V}_m^{CD})^\dagger\mathbf{V}_m^{PNC} = \mathbf{0}$. Denote

$$\mathbf{B}_m^{CD} = (\mathbf{V}_m^{CD})^\dagger\mathbf{R}_m\widetilde{\mathbf{F}}_m^{CD} \tag{24}$$

where $\mathbf{B}_m^{CD} \in \mathbb{C}^{(n_m-k)\times(n_m-k)}$. Then, it can be shown that, as $N_0 \to 0$, the CD achievable sum rate in (3) becomes

$$R^{CD} = \sum_{m\in\{A,B\}}\left(\log\left|\mathbf{G}_m^{CD}(\mathbf{G}_m^{CD})^\dagger\right| + \Delta_{m;2}\right) + \Delta_3 + \Delta_4 \tag{25}$$

where

$$\Delta_{m;3} = \frac{1}{2}\log\left|\frac{1}{N_0}(\mathbf{V}_m^{CD})^\dagger\mathbf{R}_m\widetilde{\mathbf{F}}_m^{CD}(\widetilde{\mathbf{F}}_m^{CD})^\dagger\mathbf{R}_m^\dagger\mathbf{V}_m^{CD}\right| \tag{26a}$$

$$\Delta_4 = \frac{1}{2}\log\left|\mathbf{I} - (\mathbf{V}_A^{CD})^\dagger\mathbf{Q}_A^\dagger\mathbf{Q}_B\mathbf{Q}_B^\dagger\mathbf{Q}_A\mathbf{V}_A^{CD}\right| \tag{26b}$$

$$\Delta_5 = \frac{1}{2}\log\left|\mathbf{I} - (\mathbf{V}_B^{CD})^\dagger\mathbf{Q}_B^\dagger\widetilde{\mathbf{Q}}_A^{PNC}(\widetilde{\mathbf{Q}}_A^{PNC})^\dagger\mathbf{Q}_B\mathbf{V}_B^{CD}\right| \tag{26c}$$

$$\widetilde{\mathbf{Q}}_m^{PNC} = \mathbf{Q}_m\mathbf{V}_m^{PNC}, m \in \{A,B\}. \tag{26d}$$

We interpret the above rate terms as follows. Suppose that the CD signal of user *A* is first decoded and canceled, and then the CD signal of user *B* is decoded. Then, Δ_4 (≤ 0) represents the rate loss by projecting the CD signal of user *A* (that spans the subspace $\mathcal{C}(\mathbf{Q}_A\mathbf{V}_A^{CD})$) to the null space of user *B*'s overall signal (i.e., the null space of $\mathcal{C}(\mathbf{Q}_B)$). Δ_5 (≤ 0) represents the rate loss of decoding user *B*'s CD signal. This loss results from projecting user *B*'s CD signal to the null space of user *A*'s PNC signal (by noting that user *A*'s CD signal has been canceled).

Now we consider the maximization of the sum rate. This problem, by inspection of (21) and (25), can be decomposed into four separate problems, as listed below.

$$\begin{aligned}
\text{P1}: \ & \text{maximize} \ \ \Delta_{A;1} + \Delta_{B;1} \\
& \text{subject to} \ \ \mathbf{P}^\dagger\mathbf{P} = \mathbf{I}, (\mathbf{V}_m^{PNC})^\dagger\mathbf{V}_m^{PNC} = \mathbf{I}, m \in \{A,B\}. \\
\text{P2}: \ & \text{maximize} \ \ \Delta_{m;2} + \Delta_{m;3} \\
& \text{subject to} \ \ (\widetilde{\mathbf{F}}_m^{PNC})^\dagger\widetilde{\mathbf{F}}_m^{PNC} = \mathbf{I} \text{ and } (\widetilde{\mathbf{F}}_m^{CD})^\dagger\widetilde{\mathbf{F}}_m^{CD} = \mathbf{I}. \\
\text{P3}: \ & \text{maximize} \ \ \log|\Psi_m| + \log\left|\mathbf{G}_m^{CD}(\mathbf{G}_m^{CD})^\dagger\right| \\
& \text{subject to} \ \ \text{tr}\{\Psi_m + \mathbf{G}_m^{CD}(\mathbf{G}_m^{CD})^\dagger\} \leq P_m, m \in \{A,B\}. \\
\text{P4}: \ & \text{maximize} \ \ \Delta_4 + \Delta_5 \\
& \text{subject to} \ \ (\mathbf{V}_m^{CD})^\dagger\mathbf{V}_m^{CD} = \mathbf{I}, m \in \{A,B\}.
\end{aligned}$$

Then, to prove Theorem 1, it suffices to show that the solution in (18) simultaneously achieves the maxima of P1-P4.

We start with P1. From Theorem 1 of [23], for any given $\mathbf{V}_A^{PNC}$ and $\mathbf{V}_B^{PNC}$, and the maximum of P1 is $\prod_{i=1}^{k}\frac{\widetilde{\lambda}_i}{2}$, where $\widetilde{\lambda}_i$ is the *i*th largest eigenvalue of $\widetilde{\mathbf{Q}}_A^{PNC}(\widetilde{\mathbf{Q}}_A^{PNC})^\dagger + \widetilde{\mathbf{Q}}_B^{PNC}(\widetilde{\mathbf{Q}}_B^{PNC})^\dagger$; the optimal $\mathbf{P}$ is given by the eigenvectors corresponding to the *k* largest eigenvalues. From Lemma 2 of [22], the *i*th largest singular value of $(\widetilde{\mathbf{Q}}_A^{PNC})^\dagger\widetilde{\mathbf{Q}}_B^{PNC}$ is given by $\widetilde{\sigma}_i = \widetilde{\lambda}_i - 1$. Recall from (26d) that $(\widetilde{\mathbf{Q}}_A^{PNC})^\dagger\widetilde{\mathbf{Q}}_B^{PNC} = (\mathbf{V}_A^{PNC})^\dagger(\mathbf{Q}_A)^\dagger\mathbf{Q}_B\mathbf{V}_B^{PNC}$. From Lemma 3.3.1 of [20], we obtain $\widetilde{\sigma}_i \leq \sigma_i$, where σ_i is defined below (14). Thus

$$\Delta_{A;1} + \Delta_{B;1} \leq \prod_{i=1}^{k}\frac{\widetilde{\lambda}_i}{2} = \prod_{i=1}^{k}\frac{\widetilde{\sigma}_i + 1}{2} \leq \prod_{i=1}^{k}\frac{\sigma_i + 1}{2} \tag{28}$$

where the equalities hold when $\mathbf{V}_A^{PNC} = \mathbf{U}_{(1,k)}$ and $\mathbf{V}_B^{PNC} = \mathbf{V}_{(1,k)}$ and $\mathbf{P}$ is given by (18a). Noting $\mathcal{C}(\mathbf{V}_m^{PNC}) = \mathcal{C}(\mathbf{R}_m \widetilde{\mathbf{F}}_m^{PNC})$ and $(\widetilde{\mathbf{F}}_m^{PNC})^\dagger \widetilde{\mathbf{F}}_m^{PNC} = \mathbf{I}$, we obtain that the optimal $\widetilde{\mathbf{F}}_m^{PNC}$, $m \in \{A, B\}$ are given by (18b) and (18c).

Next consider P2. To solve P2, we first consider an equivalent two-user MIMO multiple-access channel with channel matrices given by $\mathbf{R}_m \widetilde{\mathbf{F}}_m^{PNC}$ and $\mathbf{R}_m \widetilde{\mathbf{F}}_m^{CD}$. With the input covariances being the identity matrix, the asymptotic sum rate (as $N_0 \to 0$) is given by

$$\begin{aligned}
&\log\left|\mathbf{I} + \frac{1}{N_0}\mathbf{R}_m\left(\widetilde{\mathbf{F}}_m^{PNC}(\widetilde{\mathbf{F}}_m^{PNC})^\dagger + \widetilde{\mathbf{F}}_m^{CD}(\widetilde{\mathbf{F}}_m^{CD})^\dagger\right)\mathbf{R}_m^\dagger\right| \\
&= \log\left|\mathbf{I} + \frac{1}{N_0}\mathbf{R}_m\widetilde{\mathbf{F}}_m^{PNC}(\widetilde{\mathbf{F}}_m^{PNC})^\dagger\mathbf{R}_m^\dagger\right| \\
&\quad + \log\frac{\left|\mathbf{I} + \frac{1}{N_0}\mathbf{R}_m\left(\widetilde{\mathbf{F}}_m^{PNC}(\widetilde{\mathbf{F}}_m^{PNC})^\dagger + \widetilde{\mathbf{F}}_m^{CD}(\widetilde{\mathbf{F}}_m^{CD})^\dagger\right)\mathbf{R}_m^\dagger\right|}{\left|\mathbf{I} + \frac{1}{N_0}\mathbf{R}_m\widetilde{\mathbf{F}}_m^{PNC}(\widetilde{\mathbf{F}}_m^{PNC})^\dagger\mathbf{R}_m^\dagger\right|} \\
&= \Delta_{m;2} + \log\left|\mathbf{I} + \frac{1}{N_0}(\mathbf{V}_m^{CD})^\dagger\mathbf{R}_m\widetilde{\mathbf{F}}_m^{CD}(\widetilde{\mathbf{F}}_m^{CD})^\dagger\mathbf{R}_m^\dagger\mathbf{V}_m^{CD}\right| \\
&= \Delta_{m;2} + \Delta_{m;3}
\end{aligned}$$

where the second last step follows from the fact that $(\mathbf{I} + \frac{1}{N_0}\mathbf{R}_m\widetilde{\mathbf{F}}_m^{PNC}(\widetilde{\mathbf{F}}_m^{PNC})^\dagger\mathbf{R}_m^\dagger)^{-1} \xrightarrow{N_0 \longrightarrow 0} \mathbf{I} - \mathbf{V}_m^{PNC}(\mathbf{V}_m^{PNC})^\dagger = \mathbf{V}_m^{CD}(\mathbf{V}_m^{CD})^\dagger$. On the other hand, we have

$$\begin{aligned}
&\log\left|\mathbf{I} + \frac{1}{N_0}\mathbf{R}_m\left(\widetilde{\mathbf{F}}_m^{PNC}(\widetilde{\mathbf{F}}_m^{PNC})^\dagger + \widetilde{\mathbf{F}}_m^{CD}(\widetilde{\mathbf{F}}_m^{CD})^\dagger\right)\mathbf{R}_m^\dagger\right| \\
&\le \log\left|\mathbf{I} + \frac{1}{N_0}\mathbf{R}_m\mathbf{R}_m^\dagger\right| = \log\left|\mathbf{I} + \frac{1}{N_0}\mathbf{H}_m\mathbf{H}_m^\dagger\right|
\end{aligned}$$

where the equality holds when $\widetilde{\mathbf{F}}_m^{PNC}(\widetilde{\mathbf{F}}_m^{PNC})^\dagger + \widetilde{\mathbf{F}}_m^{CD}(\widetilde{\mathbf{F}}_m^{CD})^\dagger = \mathbf{I}$, or equivalently, $(\widetilde{\mathbf{F}}_m^{CD})^\dagger\widetilde{\mathbf{F}}_m^{PNC} = \mathbf{0}$. Thus, the maximum of P2 is achieved at (18).

It is straightforward to see that the maximum of P3 is achieved when equal power allocation is adopted, i.e., $\Psi_m = \frac{P_m}{n_m}\mathbf{I}$ and $\mathbf{G}_m^{CD}(\mathbf{G}_m^{CD})^\dagger = \frac{P_m}{n_m}\mathbf{I}$.

Now consider P4. The eigenvalues of $\mathbf{Q}_A^\dagger\mathbf{Q}_B\mathbf{Q}_B^\dagger\mathbf{Q}_A$ are given by $\sigma_1^2, \ldots, \sigma_{\min\{n_A, n_B\}}^2, 0, \ldots, 0$. Let λ_i^{CD} be the ith smallest eigenvalue of $(\mathbf{V}_A^{CD})^\dagger\mathbf{Q}_A^\dagger\mathbf{Q}_B\mathbf{Q}_B^\dagger\mathbf{Q}_A\mathbf{V}_A^{CD}$. From the interlacing theorem of Hermitian matrices [19], λ_i^{CD} is no less than the ith smallest eigenvalue of $\mathbf{Q}_A^\dagger\mathbf{Q}_B\mathbf{Q}_B^\dagger\mathbf{Q}_A$. Thus, $\Delta_4 = \frac{1}{2}\log\prod_{i=1}^{n_A-k}(1-\lambda_i^2) \le \frac{1}{2}\log\prod_{i=k+1}^{\min(n_A,n_B)}(1-\sigma_i^2)$, where the equality holds when $\mathbf{V}_A^{CD} = \mathbf{U}_{(1,k)}$.

Finally consider Δ_5. Noting $\Delta_5 \le 0$, we need to show that $\Delta_5 = 0$ when $\mathbf{V}_A^{PNC} = \mathbf{U}_{(1,k)}$ and $\mathbf{V}_B^{CD} = \mathbf{V}_{\overline{(1,k)}}$. To see this, it suffices to show that $(\mathbf{V}_B^{CD})^\dagger\mathbf{Q}_B^\dagger\widetilde{\mathbf{Q}}_A^{PNC} = (\mathbf{V}_B^{CD})^\dagger\mathbf{Q}_B^\dagger\mathbf{Q}_A\mathbf{V}_A^{PNC} = \mathbf{0}$, which is straightforward by noting the orthogonality between the principle vectors, i.e., $\mathbf{Q}_A\mathbf{u}_i \perp \mathbf{Q}_B\mathbf{v}_j$, for any $i \neq j$ (cf., [23], the proof of Theorem 1).

Combining the optimal solutions to P1-P4, we see that the minimum achievable sum-rate away from the capacity is at most Δ^{SD} in (16), which is achieved at (18). ∎

ACKNOWLEDGMENT

This work was partially supported by the grants from the University Grants Committee (Project No. 418712 and AoE/E-02/08) of the Hong Kong S.A.R., China. The work of Tao Yang is supported by CSIRO OCE postdoctoral fellowships.

REFERENCES

[1] S. Zhang, S. Liew, and P. P. Lam, "Hot topic: physical-layer network coding," in Proc. 12th Annual International Conf. Mobile Computing Networking (ACM MobiCom'06), Sept. 2006, pp. 358-365.

[2] W. Nam, S. Chung, Y. H. Lee, "Capacity of the Gaussian two-way relay channel to within 1/2 bit", *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5488-5494, Nov. 2010.

[3] S. Katti, S. Gollakota, and D. Katabi, "Embracing wireless interference: Analog network coding", *ACM SIGCOMM'07*.

[4] 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. Area. Comm.*, Vol. 27, No. 5, pp. 699-712, June 2009.

[5] M. Aleksic, P. Razaghi, and W. Yu, "Capacity of a class of modulo-sum relay channels," *IEEE Trans. Inf. Theory*, vol. 55, pp. 921–930, March 2009.

[6] S. H. Lim, Y.-H. Kim, A. El Gamal, and S.-Y. Chung, "Noisy network coding," *IEEE Trans. Inf. Theory*, vol. 57, pp. 3132–3152, May 2011.

[7] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inf. Theory*, vol. 51, pp. 3037–3063, September 2005.

[8] A. El Gamal, N. Hassanpour, and J. Mammen, "Relay networks with delays," *IEEE Trans. Inf. Theory*, vol. 53, pp. 3413–3431, October 2007.

[9] H. J. Yang, J. Chun and A. Paulraj, "Asymptotic capacity of the separated MIMO two-way relay channel," *IEEE Trans. Inf. Theory*, vol. 57, no. 11, pp. 7542-7554, Nov. 2011.

[10] T. Yang, X. Yuan, P. Li, I. B. Collings and J. Yuan., "A new physical-layer network coding scheme with eigen-direction alignment precoding for MIMO two-way relaying," *IEEE Trans. Comm.*, vol. 61, no. 3, March 2013.

[11] F. Wang, X. Yuan, S. C. Liew, and D. Guo, "Wireless MIMO switching: Weighted sum mean square error and sum rate optimization," to appear in *IEEE Trans. Inf. Theory*.

[12] T. Huang, T. Yang, J. Yuan and I. Land, "Design of irregular repeat-accumulate coded physical-layer network coding for Gaussian two-way relay channels," to appear in *IEEE Trans. Commun.*.

[13] A. Khina, Y. Kochman, U. Erez, "Physical-layer MIMO relaying," in *Proc. ISIT'11*, Saint Petersburg, Russia, July/August 2011.

[14] B. Nazer and M. Gasper, "Compute-and-forward: Harnessing interference through structured codes," *IEEE Trans. Inf. Theory*, vol. 57, pp. 6463-6486, Oct. 2011.

[15] 3GPP, "Overview of 3GPP Release 10," Release 10 V0.1.1 (2011-6), June 2011. (Available at http://www.3gpp.org/Release-10)

[16] IEEE, "IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Broadband Wireless Access Systems Amendment 3: Advanced Air Interface," IEEE Std 802.16m-2011, May 2011. (Available at http://www.ieee802.org/16/published.html)

[17] T. M. Cover, and J. A. Thomas, *Elements of Information Theory*, Wiley, New York, 1991.

[18] D. Tse and P. Visanath, *Fundamentals of Wireless Communications*, Cambridge University Press, 2006.

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

[20] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge Unversity Press, 1991.

[21] G. H. Golub and C. F. v. Loan, *Matrix Computations* , 3rd ed., Baltimore, MD: The Johns Hopkins University Press, 1996.

[22] J. Miao and A. Ben-Israel, "On principle angles between subspaces in $\mathbb{R}^n$," *Lin. Algeb. and its Appl.*, vol. 171, pp. 81-98, Nov. 1992.

[23] T. Yang, X. Yuan, and I. B. Collings, "Reduced-dimension cooperative precoding for MIMO two-way relay channels," *IEEE Trans. Wireless Commun.*, vol. 11, no. 11, Nov. 2012.

[24] X. Yuan, T. Yang, and I. B. Collings, "MIMO two-way relaying: A space-division approach," submitted to *IEEE Trans. Inform. Theory*. Available at: http://arxiv.org/abs/1209.3702