Space-Division Approach for Multi-pair MIMO Two Way Relaying: A Principal-Angle Perspective

Haiyang Xin, Xiaojun Yuan, and Soung-Chang Liew
Dept. of Information Engineering, The Chinese University of Hong Kong, Hong Kong

***Abstract*—This work investigates the maximum sum-rate of multi-pair MIMO two-way relay channels (TWRCs), in which a relay is responsible for forwarding information between multiple pairs of users. In this system, each pair of users forms a TWRC, and a user exchanges information only with its counterpart in the same TWRC. We focus on the multi-access channel (MAC) phase of the two-pair TWRCs. We first put forth a new interpretation of the space-division (SD) method based on the classical concept of principal angles in linear algebra. We argue that the signal spaces of the two users in a TWRC can be divided into the physical-layer network coding (PNC) signal subspace and the complete decoding (CD) signal subspace according to the principal angles between the two signal spaces of different user pairs. Based on the principal-angle framework, we then propose an extended SD method to mitigate the interference among the PNC signals and CD signals. We further derive the optimal decoding strategy and optimal precoder design principles that maximize the sum-rate of the MAC phase. The associated optimization problem, however, is non-convex. We therefore propose a suboptimal solution for precoder design and analyze the asymptotic rate gap benchmarked against the cut-set bound. Significant performance improvements have been observed for the proposed hybrid PNC-CD system compared with pure complete decoding and pure PNC decoding.**

I. Introduction

Physical-layer network coding (PNC) [1]-[3] is an emerging relaying technique for efficient communications over wireless networks. In PNC, the relay(s) may decode a linear function of the incoming messages rather than the explicit messages themselves. Doing so can potentially boost the network throughput significantly.

The efficient PNC design for a multiuser two-way relay channel (TWRC) system, in which multiple users exchange information in a pairwise manner via a single relay, has recently attracted much research interest [5]-[11]. Particularly, the authors in [5], [6] studied the case that the relay is equipped with a single antenna. Later, several other approaches studied the case that multiple antennas are equipped at the relay [7]-[11]. Various design criteria have been proposed to exploit the benefit of spatial multiplexing provided by the multiple relay antennas.

The existing work on multiuser TWRCs has two limitations. First, most of these approaches focus on analogue network coding (ANC) [4]. While ANC has the merit of implementation simplicity, it suffers from noise amplification (as denoising is absent at the relay) and power loss at the relay (as the user signals are linearly superimposed in relay transmission). It is desirable to use the more sophisticated PNC set-up to avoid these disadvantages. Second, most existing approaches only consider the configuration of single antenna at each user. Multiple antennas at the user ends can create extra spatial degrees of freedom (DoF), leading to the possibility of multifold increase in system throughput. However it poses significant challenges on the joint design of user transceivers and relay operations.

In this paper, we study the efficient transceiver strategy for multiuser MIMO TWRCs in which both the relay and the users are equipped with multiple antennas. Following the space-division (SD) approach for one-pair MIMO TWRC [13], we divide the signal space of each user into two subspaces, one for PNC decoding (in which network-coded messages are decoded) and the other for complete-decoding (in which user messages are individually decoded). We formulate a sum-rate maximization problem for the precoder design of the users and the relay in the high signal to noise (SNR) regime, and then simplify it to the optimization of PNC precoders only. Since the problem is still non-convex, we propose a suboptimal (but good) solution for it. We analyze the asymptotic rate gap between the proposed scheme and the cut-set bound. Closed-form expressions are derived based on the concept of principal angles [16]. Numerical results demonstrate the superiority of the proposed scheme compared with the pure PNC or CD method.

II. Preliminary

We start with a one-pair MIMO TWRC, in which two users exchange information via a relay. Our target is to give a new interpretation of the space-division approach [13] from the perspective of principal angles and principal vectors.

A. One-Pair MIMO TWRC

In the TWRC, the data exchange includes two transmission phases, namely, the multi-access (MAC) phase and the broadcast (BC) phase. It is known that the throughput bottleneck is the MAC phase, which is our focus in what follows. In the MAC phase, the received signal at the relay is given by

$$\mathbf{y} = \mathbf{H}_1\mathbf{F}_1\mathbf{x}_1 + \mathbf{H}_2\mathbf{F}_2\mathbf{x}_2 + \mathbf{n} \qquad (1)$$

where $\mathbf{H}_i \in \mathbb{R}^{m\times n}$ is the channel matrix between the relay and the user-i, $\mathbf{x}_i \in \mathbb{R}^{n\times 1}$ is a signal vector of user-i with $E[\mathbf{x}_i\mathbf{x}_i^T] = \mathbf{I}$, $\mathbf{F}_i \in \mathbb{R}^{n\times n}$ is the corresponding precoding matrix satisfying the power constraint of $\text{tr}\{\mathbf{F}_i^T\mathbf{F}_i\} \le P_i$, $i \in \{1, 2\}$, and $\mathbf{n} \in \mathbb{R}^{m\times 1}$ is the white Gaussian noise vector with the elements independently drawn from $\mathcal{N}(0, 1)$. We assume that the entries of $\mathbf{H}_i$, $i \in \{1, 2\}$, are independently drawn. Thus,

978-1-4799-1353-4/13/$31.00 ©2013 IEEE

$\mathbf{H}_i$ is of full column or row rank, whichever is smaller, with probability one.

Upon receiving $\mathbf{y}$, the key issue for the relay design is to determine what to decode. Let $\mathcal{C}(\mathbf{H}_i)$ be the column space of $\mathbf{H}_i$. Intuitively, we expect that when $\mathcal{C}(\mathbf{H}_1)$ and $\mathcal{C}(\mathbf{H}_2)$ are nearly orthogonal, meaning that the signals of the two users nearly do not interfere with each other, then decoding both signals explicitly is preferable. Otherwise, if $\mathcal{C}(\mathbf{H}_1)$ and $\mathcal{C}(\mathbf{H}_2)$ are overlapping to a large extent, a better choice is to project them onto a common subspace and then decode a function of the two signals, following the basic idea of PNC [1]. "Principal angles" characterize the amount of overlapping between the two column spaces and allows the above intuition to be made concrete.

B. Principal Angle

Definition 1 (*Principal angle* [16]) The principal angles of $\mathcal{C}(\mathbf{H}_1)$ and $\mathcal{C}(\mathbf{H}_2)$, denoted by $\theta_1 \le \theta_2, .., \le \theta_n \in [0, \pi/2]$, with $n \le m$, are defined recursively by

$$\cos(\theta_j) = \max_{\mathbf{u}\in\mathcal{C}(\mathbf{H}_1), \mathbf{v}\in\mathcal{C}(\mathbf{H}_2)} \mathbf{u}^T\mathbf{v} = \mathbf{u}_j^T\mathbf{v}_j \tag{2}$$

subject to: $\|\mathbf{u}\|_2 = \|\mathbf{v}\|_2 = 1$,$\mathbf{u}^T\mathbf{u}_k = 0$, $\mathbf{v}^T\mathbf{v}_k = 0$, $k = 1, ..., j-1$, where $\mathbf{u}_j$ and $\mathbf{v}_j$ are defined as the principal-vector pair of $\mathcal{C}(\mathbf{H}_1)$ and $\mathcal{C}(\mathbf{H}_2)$, associated with θ_j. A simple algorithm to calculate the principal angles is given as follows [16]:

Algorithm 1 (*Calculating principal angles and vectors*)

Step 1: Compute an orthonormal basis of $\mathcal{C}(\mathbf{H}_i)$ using QR decomposition:

$$\mathbf{H}_i = \mathbf{Q}_i\mathbf{R}_i \tag{3}$$

where $\mathbf{Q}_i \in \mathbb{R}^{m\times n}$, satisfying $\mathbf{Q}_i^T\mathbf{Q}_i = \mathbf{I}$, gives the required orthonormal basis, and $\mathbf{R}_i \in \mathbb{R}^{n\times n}$ is an upper-triangular matrix, $i \in \{1, 2\}$.

Step 2: Compute the compact SVD of $\mathbf{Q}_1^T\mathbf{Q}_2$:

$$\mathbf{Q}_1^T\mathbf{Q}_2 = \mathbf{S}\boldsymbol{\Sigma}\mathbf{T}^T \tag{4}$$

where $\mathbf{S} \in \mathbb{R}^{n\times n}$ and $\mathbf{T} \in \mathbb{R}^{n\times n}$ are orthogonal matrices, and $\boldsymbol{\Sigma} = \text{diag}\{\sigma_1, .., \sigma_n\} = \text{diag}\{\cos\theta_1, .., \cos\theta_n\}$ determines the principal angles. The principal vectors are then given by $\mathbf{U} = [\mathbf{u}_1, .., \mathbf{u}_n] = \mathbf{Q}_1\mathbf{S} \in \mathbb{R}^{m\times n}$ and $\mathbf{V} = [\mathbf{v}_1, .., \mathbf{v}_n] = \mathbf{Q}_2\mathbf{T} \in \mathbb{R}^{m\times n}$.

Remark 1: By definition, we have $\mathbf{U}^T\mathbf{V} = \mathbf{S}^T\mathbf{Q}_1^T\mathbf{Q}_2\mathbf{T} = \boldsymbol{\Sigma}$, which implies that $\mathbf{u}_j \perp \mathbf{v}_k, \forall j \ne k$. Together with $\mathbf{u}_j \perp \mathbf{u}_k$ and $\mathbf{v}_j \perp \mathbf{v}_k$, $\forall j \ne k$, we see that span$\{\mathbf{u}_1, \mathbf{v}_1\} \perp ... \perp$ span$\{\mathbf{u}_n, \mathbf{v}_n\}$, i.e., the channel spaces can be decomposed into n orthogonal dimension-2 subspaces, each spanned by one principal vector pair. In each subspace, the angle between $\mathbf{u}_j$ and $\mathbf{v}_j$ is given by θ_j, i.e., $\cos\theta_j = \mathbf{u}_j^T\mathbf{v}_j$.

Remark 2: As both $\mathbf{Q}_1$ and $\mathbf{Q}_2$ have orthonormal columns, the singular values of $\mathbf{Q}_1^T\mathbf{Q}_2$ are confined in $[0, 1]$. Particularly, $\sigma_j = 0$ means that the corresponding $\mathbf{u}_j$ and $\mathbf{v}_j$ are orthogonal; $\sigma_j = 1$ means that $\mathbf{u}_j$ and $\mathbf{v}_j$ are parallel. Note that in the case of $m \le n$, $\mathcal{C}(\mathbf{H}_1) = \mathcal{C}(\mathbf{H}_2) = \mathbb{R}^m$. Then $\sigma_1 = ... = \sigma_m = 1$ and $\{\mathbf{u}_1 = \mathbf{v}_1, ..., \mathbf{u}_m = \mathbf{v}_m\}$ can be taken as any orthonormal basis of $\mathbb{R}^m$.

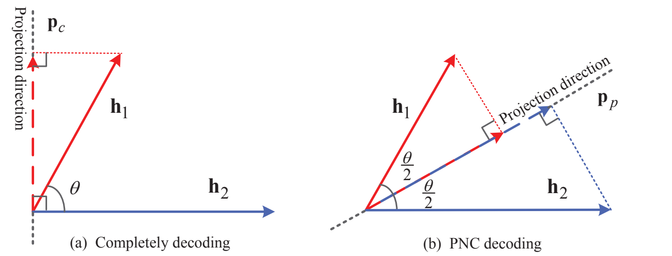

Fig. 1. Geometrical illustration of two decoding strategies

C. Complete Decoding vs. PNC Decoding

We now compare two decoding strategies at the relay, namely, complete decoding (CD) and PNC decoding. As a toy example, we assume that each user has one antenna. Then (1) reduces to

$$\mathbf{y} = \mathbf{h}_1 f_1 x_1 + \mathbf{h}_2 f_2 x_2 + \mathbf{n} \tag{5}$$

where $\mathbf{h}_1$ and $\mathbf{h}_2$ are channel vectors, x_i is a scalar signal of user-i with $E[|x_i|^2] = 1$, and f_i is the power control factor satisfying $|f_i|^2 \le P_i$, $i \in \{1, 2\}$. The relay performs either complete decoding or PNC decoding as shown in Fig. 1.

For complete decoding, the relay first decodes user-1's signal x_1 by treating x_2 as interference; then x_1 is canceled from $\mathbf{y}$, and x_2 is decoded. In the high SNR regime, decoding x_1 in the first step is equivalent to projecting $\mathbf{h}_1 x_1$ onto the unit vector $\mathbf{p}_c \in$ span$\{\mathbf{h}_1, \mathbf{h}_2\}$ that is orthogonal to $\mathbf{h}_2$, and then decoding x_1 based on the image, as illustrated in Fig. 1(a). Let θ be the angle between $\mathbf{h}_1$ and $\mathbf{h}_2$. The asymptotic achievable sum-rate is given by $\frac{1}{2}\log_2(P_1\|\mathbf{h}_1\|^2\sin^2\theta) + \frac{1}{2}\log_2(P_2\|\mathbf{h}_2\|^2)$.

For PNC decoding, the relay first projects both signal vectors onto a common direction specified by a unit vector $\mathbf{p}_p$, yielding $\mathbf{p}_p^T\mathbf{y} = \mathbf{p}_p^T\mathbf{h}_1 x_1 + \mathbf{p}_p^T\mathbf{h}_2 x_2 + \mathbf{p}_p^T\mathbf{n}$, and then performs nested lattice decoding based on the resulting scalar channel [14]. It was shown in [13] that the optimal projection direction is the angular bisector of $\mathbf{h}_1$ and $\mathbf{h}_2$, as illustrated in Fig. 1(b), and the corresponding high-SNR achievable sum-rate is $\frac{1}{2}\log_2(P_1\|\mathbf{h}_1\|^2\cos^2\frac{\theta}{2}) + \frac{1}{2}\log_2(P_2\|\mathbf{h}_2\|^2\cos^2\frac{\theta}{2})$.

In the above achievable sum-rates, $\frac{1}{2}\log_2(P_1\|\mathbf{h}_1\|^2) + \frac{1}{2}\log_2(P_2\|\mathbf{h}_2\|^2)$ is actually the sum-rate of the cut-set bound of (5). Then, it is clear that the rate gap between complete decoding and the cut-set bound is $-\frac{1}{2}\log_2\sin^2\theta$, as incurred by the power loss due to projection; the rate gap between PNC decoding and the cut-set bound is $-\log_2\cos^2\frac{\theta}{2}$, as incurred by the power loss of projection.

We next extend our discussions to the MIMO set-up in (1). Based on the concept of principal angles described in Subsection B, we propose a heuristic transceiver strategy as follows. The spatial streams of the two users are partitioned into n pairs, each pair-j consisting of one signal $x_{i,j}$ from user-i. For each pair, we choose the beamforming vectors $\mathbf{f}_{i,j}$ (i.e. the j-th column of $\mathbf{F}_i$), such that the received signal vector $\mathbf{H}_i\mathbf{f}_{i,j}x_{i,j}$, $i \in \{1, 2\}$ are aligned to the directions of the principal vector pair $\{\mathbf{u}_j, \mathbf{v}_j\}$. This is possible since $\mathbf{u}_j \in \mathcal{C}(\mathbf{H}_1)$ and $\mathbf{v}_j \in \mathcal{C}(\mathbf{H}_2)$. Then we expect that the rate gap can be characterized by the corresponding principal angles with $-\frac{1}{2}\log_2\sin^2\theta_j$ for complete decoding or $-\log_2\cos^2\frac{\theta_j}{2}$

for PNC decoding. The above approach gives good intuitions on connecting the rate gap with the principal angles of $\mathcal{C}(\mathbf{H}_1)$ and $\mathcal{C}(\mathbf{H}_2)$. However, this approach is not power-efficient due to its zero-forcing nature, i.e., $\text{span}\{\mathbf{H}_1\mathbf{f}_{1,j}, \mathbf{H}_2\mathbf{f}_{2,j}\} = \text{span}\{\mathbf{u}_j, \mathbf{v}_j\} \perp \text{span}\{\mathbf{u}_k, \mathbf{v}_k\} = \text{span}\{\mathbf{H}_1\mathbf{f}_{1,k}, \mathbf{H}_2\mathbf{f}_{2,k}\}$, $\forall j \neq k$, following from Remark 1. To reduce the power loss of zero-forcing, it is better not to align the signals exactly to the principal-vector directions. Then, the interference appears in the dimension-2 subspace spanned by each principal vector pair, and hence successive interference cancellation is necessary. The details can be found in the space-division approach proposed in [13]. Furthermore, Theorem 1 of [13] implies that the following sum-rates are achievable.

Proposition 1: As $P_1, P_2 \to \infty$, the asymptotically optimal $\mathbf{F}_i$ to maximize the achievable sum-rate satisfies $\mathbf{F}_i^T\mathbf{F}_i = \frac{P_i}{n}\mathbf{I}$, $i \in \{1, 2\}$. Moreover, the rate gap of complete decoding and PNC decoding with respect to the cut-set bound are given by

$$R_{cs} - R_{cd} \longrightarrow -\sum_{j=1}^{n} \frac{1}{2}\log_2 \sin^2\theta_j \tag{6}$$

$$R_{cs} - R_{pnc} \longrightarrow -\sum_{j=1}^{n} \log_2 \cos^2\frac{\theta_j}{2} \tag{7}$$

where $R_{cs} = \sum_{i=1}^{2}\frac{1}{2}\log_2\left|\mathbf{I} + \frac{P_i}{n}\mathbf{H}_i\mathbf{H}_i^T\right|$ is the sum-rate of the cut-set bound of the MAC phase in the high SNR regime.

Sketch of proof: Let λ_j be the j-th largest eigenvalue of $\mathbf{Q}_1\mathbf{Q}_1^T + \mathbf{Q}_2\mathbf{Q}_2^T$. From [15], λ_j is related to θ_j by $\lambda_j = 1 + \cos\theta_j$. Then, Proposition 1 follows immediately from Theorem 1 of [13] by setting the number of PNC streams to be 0 and n, respectively.

III. System Model

We now consider the multi-pair MIMO TWRCs. We focus on the case of two user-pairs, which captures the key issues involved in the generalization from one pair to multiple pairs. The system model is similar to the one-pair case. The only difference is that here two pairs of users, namely, pair-A and pair-B, simultaneously exchange information via the relay in a pair-wise manner. Each round of information exchange still consists of the MAC and BC phases. As in the one-pair case, we focus on the MAC phase, with the channel model given by

$$\mathbf{y} = \sum_{l\in\{A,B\}}\sum_{i\in\{1,2\}} \mathbf{H}_{li}\mathbf{F}_{li}\mathbf{x}_{li} + \mathbf{n} \tag{8}$$

where $\mathbf{H}_{li} \in \mathbb{R}^{m\times n}$ is the channel matrix for user-i in pair-l; $\mathbf{x}_{li} \in \mathbb{R}^{n\times 1}$ and $\mathbf{F}_{li} \in \mathbb{R}^{n\times n}$ are the signal vector and corresponding precoder, with $E[\mathbf{x}_{li}\mathbf{x}_{li}^T] = \mathbf{I}$ and $\text{tr}\{\mathbf{F}_{li}^T\mathbf{F}_{li}\} \leq P_{li}$, $l \in \{A, B\}$, $i \in \{1, 2\}$; and $\mathbf{n}$ is defined in (1). Upon receiving $\mathbf{y}$, the relay performs decoding operations. The details will be elaborated in the next section.

The cut-set bound of the above scheme can be obtained by a straightforward extension of the results in [6] for multi-pair SISO TWRCs. Let R_{li} be the information rate of user-i in pair-l. Then the rate inequalities related to the MAC phase are listed below:

$$R_{li} \leq \frac{1}{2}\log_2\left|\mathbf{I} + \mathbf{H}_{li}\bar{\mathbf{Q}}_{li}\mathbf{H}_{li}^T\right|, l \in \{A, B\}, i \in \{1, 2\} \tag{9}$$

$$R_{Ai} + R_{Bj} \leq \frac{1}{2}\log_2\left|\mathbf{I} + \mathbf{H}_{Ai}\bar{\mathbf{Q}}_{Ai}\mathbf{H}_{Ai}^T + \mathbf{H}_{Bj}\bar{\mathbf{Q}}_{Bj}\mathbf{H}_{Bj}^T\right|, \quad i \in \{1, 2\}, j \in \{1, 2\} \tag{10}$$

where $\bar{\mathbf{Q}}_{li} = \mathbf{F}_{li}\mathbf{F}_{li}^T$ is the channel input covariance of user-i in pair-l. It can be shown that, at high SNR, the optimal $\bar{\mathbf{Q}}_{li}$ is given by $\bar{\mathbf{Q}}_{li} = P_{li}/n\mathbf{I}$. This cut-set bound will be used as a benchmark to evaluate our proposed transceiver strategy.

IV. Proposed Space-Division Approach

A. Space-Division Signaling

In this section, we propose a space-division approach for the multi-pair MIMO TWRCs in (8). In the proposed scheme, the signal streams of each user are divided into two groups: one for PNC and the other for CD. As in the one-pair case, we require that the PNC signals for each user-pair be close to parallel (as measured by the principal angles) so that they can be efficiently projected onto common directions for PNC decoding; and the CD signals in each user-pair be close to orthogonal. The new challenge is that, when we identify the PNC and CD directions for each pair, the interference from the other pair should be taken into account. In what follows, we develop a solution to this problem.

To achieve full spatial multiplexing, each user transmits n spatial streams. Without loss of generality, we assume that within each user-pair l, $l \in \{A, B\}$, there are k_l pairs of PNC signals. The remaining $n - k_l$ streams are CD signals for each user in user-pair l. Let $\mathbf{x}_{li}^p \in \mathbb{R}^{k_l\times 1}$ and $\mathbf{x}_{li}^c \in \mathbb{R}^{(n-k_l)\times 1}$ be the PNC and CD streams, $\mathbf{F}_{li}^p \in \mathbb{R}^{n\times k_l}$ and $\mathbf{F}_{li}^c \in \mathbb{R}^{n\times(n-k_l)}$ be the corresponding precoding matrices, respectively. Then

$$\mathbf{x}_{li} = \left[(\mathbf{x}_{li}^p)^T, (\mathbf{x}_{li}^c)^T\right]^T \text{ and } \mathbf{F}_{li} = [\mathbf{F}_{li}^p, \mathbf{F}_{li}^c]. \tag{11}$$

Note that $k_l = n$, $l \in \{A, B\}$, corresponds to the pure PNC method; $k_l = 0$, $l \in \{A, B\}$, corresponds to the pure CD method.

Remark 3: We assume that the number of antennas at the relay (i.e., m) is sufficiently large, so that the degree of freedom (DoF) of the scheme is limited by the number of antennas at each user (i.e., n). DoF analysis shows that, to ensure the validity of the above assumption, we need to set $m \geq 3n$ and $n \geq k_l \geq \max(4n - m, 0)$. Details are omitted here due to space limitation and will be provided in a later report.

B. Decoding Order at the Relay

The decoding at the relay consists of four steps. In each step, either all the CD signals or all the PNC signals of one user-pair are decoded. Then, there are in total $4! = 24$ different choices of decoding orders. For example, $\{A^c, A^p, B^c, B^p\}$ means that the CD signals of pair-A, the PNC signals of pair-A, the CD signals of pair-B, and the PNC signals of pair-B are decoded sequentially. Note that once the CD signals are decoded, they are immediately cancelled from the received signals without causing interference to those undecoded signals in the remaining decoding process; but the decoded PNC signals can not be cancelled, since only the linear functions of them are decoded. Clearly, the achievable sum-rate of the scheme depends on the decoding order at the relay. The optimal decoding order is given bellow,

Proposition 2: For the space-division approach, the optimal decoding order for sum-rate maximization satisfies: the first two steps for complete decoding; the remaining two steps for PNC decoding. That it, the four possible optimal decoding orders are

$$\begin{aligned}&\{A^c, B^c, A^p, B^p\}, \quad \{A^c, B^c, B^p, A^p\},\\ &\{B^c, A^c, A^p, B^p\}, \quad \{B^c, A^c, B^p, A^p\}. \end{aligned} \tag{12}$$

Proof: To prove Proposition 2, it suffices to establish the following two points: (i) for any decoding order, if two consecutive steps decode signals from the same user-pair, then the order of compete decoding followed by PNC decoding always achieves a higher sum-rate; (ii) if two consecutive steps are both for complete decoding (or both for PNC decoding), swapping these steps does not affect the achievable sum-rate.

Point (i) is true since complete decoding followed by PNC decoding allows interference cancellation of CD signals prior to PNC decoding, which implies a higher PNC rate (while the CD rate remains the same no matter whether CD or PNC decoding is performed first). Point (ii) is true due to the following reasons: for the CD case, all the CD signals form a MAC channel, and it is well-known that the MAC sum-rate does not dependent on the decoding order; for the PNC case, the decoding of the PNC signals in one user-pair always treats the PNC signals from the other pair as interference (i.e., no interference cancellation is applied), and therefore the achievable sum-rate does not change, regardless of the PNC signals of which user-pair are decoded first. This completes the proof of Proposition 2.

C. Achievable Rate

We now present the achievable rate of the above scheme. Suppose that at a certain step t ($1 \le t \le 4$), the signals $\{\mathbf{x}_{l1}^s, \mathbf{x}_{l2}^s\}$ of pair-l are to be decoded. Note that $s = c$ means that the CD signals are decoded at Step t; $s = p$ means that the PNC signals are decoded at Step t. The relay pre-cancels the CD signals decoded in the steps prior to Step t (if exists), yielding

$$\mathbf{y}_l^s = \mathbf{H}_{l1}\mathbf{F}_{l1}^s\mathbf{x}_{l1}^s + \mathbf{H}_{l2}\mathbf{F}_{l2}^s\mathbf{x}_{l2}^s + \mathbf{e}_l^s + \mathbf{n}. \tag{13}$$

In the above, $\mathbf{e}_l^s$ is the interference from all the PNC signals and also from the CD signals of the other pair (if not decoded yet). For example, suppose that the decoding order is $\{B^c, A^c, A^p, B^p\}$. Then

$$\mathbf{e}_B^c = \sum_{i=1}^{2}\mathbf{H}_{Ai}\mathbf{F}_{Ai}^c\mathbf{x}_{Ai}^c + \sum_{l\in\{A,B\}}\sum_{i=1}^{2}\mathbf{H}_{li}\mathbf{F}_{li}^p\mathbf{x}_{li}^p \tag{14}$$

$$\mathbf{e}_A^c = \sum_{l\in\{A,B\}}\sum_{i=1}^{2}\mathbf{H}_{li}\mathbf{F}_{li}^p\mathbf{x}_{li}^p \tag{15}$$

$$\mathbf{e}_A^p = \sum_{i=1}^{2}\mathbf{H}_{Bi}\mathbf{F}_{Bi}^p\mathbf{x}_{Bi}^p \tag{16}$$

$$\mathbf{e}_B^p = \sum_{i=1}^{2}\mathbf{H}_{Ai}\mathbf{F}_{Ai}^p\mathbf{x}_{Ai}^p. \tag{17}$$

Here we describe an interference whitening approach to give an asymptotic sum-rate expression based on Proposition 1. Let $\mathbf{W}_l^s = E\left[\mathbf{e}_l^s(\mathbf{e}_l^s)^T\right] + \mathbf{I}$ be the covariance matrix of the interference-plus-noise term, i.e., $\mathbf{e}_l^s + \mathbf{n}$. Then we left-multiply $\mathbf{y}_l^s$ by $(\mathbf{W}_l^s)^{-\frac{1}{2}}$, so that the effective noise $\mathbf{n}_l^s = (\mathbf{W}_l^s)^{-\frac{1}{2}}(\mathbf{e}_l^s + \mathbf{n})$ is an AWGN vector with $E\left[\mathbf{n}_l^s(\mathbf{n}_l^s)^T\right] = \mathbf{I}$. The effective channel is given by

$$(\mathbf{W}_l^s)^{-\frac{1}{2}}\mathbf{y}_l^s = \tilde{\mathbf{H}}_{l1}^s\mathbf{F}_{l1}^s\mathbf{x}_{l1}^s + \tilde{\mathbf{H}}_{l2}^s\mathbf{F}_{l2}^s\mathbf{x}_{l2}^s + \mathbf{n}_l^s, \tag{18}$$

where $\tilde{\mathbf{H}}_{li}^s = (\mathbf{W}_l^s)^{-\frac{1}{2}}\mathbf{H}_{li}$, $i \in \{1, 2\}$. From Proposition 1, the rate achieved in the high SNR regime is

$$R_l^c \approx \sum_{i=1}^{2}\frac{1}{2}\log_2\left|\mathbf{I} + \tilde{\mathbf{H}}_{li}^c\mathbf{F}_{li}^c(\tilde{\mathbf{H}}_{li}^c\mathbf{F}_{li}^c)^T\right| + \sum_{j=1}^{n-k_l}\frac{1}{2}\log_2\sin^2\theta_{l,j}^c \tag{19}$$

$$R_l^p \approx \sum_{i=1}^{2}\frac{1}{2}\log_2\left|\mathbf{I} + \tilde{\mathbf{H}}_{li}^p\mathbf{F}_{li}^p(\tilde{\mathbf{H}}_{li}^p\mathbf{F}_{li}^p)^T\right| + \sum_{j=1}^{k_l}\log_2\cos^2\frac{\theta_{l,j}^p}{2} \tag{20}$$

where $\theta_{l,j}^s$ is the j-th principal angle between $\mathcal{C}(\tilde{\mathbf{H}}_{l1}^s\mathbf{F}_{l1}^s)$ and $\mathcal{C}(\tilde{\mathbf{H}}_{l2}^s\mathbf{F}_{l2}^s)$, and R_l^c (or R_l^p) is the sum-rate of user-pair l's CD (or PNC) signals.

V. Asymptotic Precoder Design

A. Problem Formulation

We now consider the precoder design to maximize the achievable sum-rate of the proposed SD approach. Without loss of generality, let

$$\mathbf{F}_{li}^s = \tilde{\mathbf{F}}_{li}^s\mathbf{\Gamma}_{li}^s, \ s \in \{p, c\}, l \in \{A, B\}, i \in \{1, 2\} \tag{21}$$

where $\tilde{\mathbf{F}}_{li}^p \in \mathbb{R}^{n\times k_l}$ and $\tilde{\mathbf{F}}_{li}^c \in \mathbb{R}^{n\times(n-k_l)}$, satisfying $(\tilde{\mathbf{F}}_{li}^s)^T\tilde{\mathbf{F}}_{li}^s = \mathbf{I}$, specify the beamforming directions of the precoders, and $\mathbf{\Gamma}_{li}^p \in \mathbb{R}^{k_l\times k_l}$ and $\mathbf{\Gamma}_{li}^c \in \mathbb{R}^{(n-k_l)\times(n-k_l)}$ are for power allocation. Then the optimization problem for maximizing the sum-rate of the MAC phase is formulated as follows:

$$\begin{aligned}\max_{\tilde{\mathbf{F}}_{li}^s, \mathbf{\Gamma}_{li}^s} \quad & \sum_{l\in\{A,B\}}\sum_{s\in\{p,c\}} R_l^s\\ \text{s.t.} \quad & (\tilde{\mathbf{F}}_{li}^s)^T\tilde{\mathbf{F}}_{li}^s = \mathbf{I},\\ & \text{tr}\{(\mathbf{\Gamma}_{li}^p)^T\mathbf{\Gamma}_{li}^p\} + \text{tr}\{(\mathbf{\Gamma}_{li}^c)^T\mathbf{\Gamma}_{li}^c\} \le P_{li},\\ & s \in \{p, c\}, \ l \in \{A, B\}, \ i \in \{1, 2\}. \end{aligned} \tag{22}$$

B. Asymptotic Precoder Design

The problem in (22) is in general difficult to solve. For ease of analysis, we look at the precoder design in the high SNR regime.

Proposition 3: As $P_{li} \to \infty$, $l \in \{A, B\}$ and $i \in \{1, 2\}$, the asymptotically optimal solution to problem (22) satisfies $(\tilde{\mathbf{F}}_{li}^c)^T\tilde{\mathbf{F}}_{li}^p = \mathbf{0}$ and $(\mathbf{\Gamma}_{li}^p)^T\mathbf{\Gamma}_{li}^p = P_{li}/n\mathbf{I}$ and $(\mathbf{\Gamma}_{li}^c)^T\mathbf{\Gamma}_{li}^c = P_{li}/n\mathbf{I}$.

Proof: To prove Proposition 3, it suffices to consider the optimality of the precoders for one user in one user-pair by fixing those for the other users. Without losing generality, we choose user-1 of pair-A to be of concern. Then, to prove Proposition 3, we need to show that, for any given $\tilde{\mathbf{F}}^p_{A1}$ (and the precoders for the other users), the optimal $\{\tilde{\mathbf{F}}^c_{A1}, \mathbf{\Gamma}^c_{A1}, \mathbf{\Gamma}^p_{A1}\}$ satisfies $(\tilde{\mathbf{F}}^c_{A1})^T\tilde{\mathbf{F}}^p_{A1} = \mathbf{0}$ and $(\mathbf{\Gamma}^p_{A1})^T\mathbf{\Gamma}^p_{A1} = P_{A1}/n\mathbf{I}$ and $(\mathbf{\Gamma}^c_{A1})^T\mathbf{\Gamma}^c_{A1} = P_{A1}/n\mathbf{I}$. From Proposition 2, there are four optimal decoding orders achieve the same sum-rate. Thus, without less of generality, we focus on the decoding order $\{B^c, A^c, A^p, B^p\}$ in studying the achievable sum-rate.

We first show that R^c_B (i.e., the achievable rate in Step 1) and R^p_B (i.e., the achievable rate in Step 4) are independent of $\{\tilde{\mathbf{F}}^c_{A1}, \mathbf{\Gamma}^c_{A1}, \mathbf{\Gamma}^p_{A1}\}$. To see this, we recall that R^c_B is the achievable sum-rate of the CD signals of pair-B by treating $\mathbf{e}^c_B$ in (14) as the interference. At high SNR, R^c_B only depends on the interference subspace occupied by $\mathbf{e}^c_B$ (that may be related to the optimization variables). This interference subspace can be represented as $\mathcal{C}(\mathbf{H}_{B1}\tilde{\mathbf{F}}^p_{B1}) \cup \mathcal{C}(\mathbf{H}_{B2}\tilde{\mathbf{F}}^p_{B2}) \cup \mathcal{C}(\mathbf{H}_{A1}) \cup \mathcal{C}(\mathbf{H}_{A2})$, by noting that, to achieve full spatial multiplexing, $\mathcal{C}(\mathbf{H}_{Ai}\tilde{\mathbf{F}}^c_{Ai}) \cup \mathcal{C}(\mathbf{H}_{Ai}\tilde{\mathbf{F}}^p_{Ai}) = \mathcal{C}(\mathbf{H}_{Ai})$, $i \in \{1,2\}$. Thus the interference subspace of $\mathbf{e}^c_B$ is independent of $\{\tilde{\mathbf{F}}^c_{A1}, \mathbf{\Gamma}^c_{A1}, \mathbf{\Gamma}^p_{A1}\}$. We next consider R^p_B. Similarly, at high SNR, R^p_B only depends on the interference subspace of $\mathbf{e}^p_B$ in (17), which is also independent of $\{\tilde{\mathbf{F}}^c_{A1}, \mathbf{\Gamma}^c_{A1}, \mathbf{\Gamma}^p_{A1}\}$. Therefore, there is no need to consider R^c_B and R^p_B in optimizing $\{\tilde{\mathbf{F}}^c_{A1}, \mathbf{\Gamma}^c_{A1}, \mathbf{\Gamma}^p_{A1}\}$.

We next consider R^c_A in Step 2 and R^p_A in step 3. We divide Step 2 into two substeps: first decode the CD signals of user-2 of pair-A; then decode the CD signals of user-1 of pair-A. (Note that reversing the order leads to the same sum-rate.) Let R^c_{A2} be the achievable rate for the first substep and R^c_{A1} be the achievable rate for the second substep. Then we have $R^c_A = R^c_{A2} + R^c_{A1}$. At high SNR, R^c_{A2} only depends on the interference subspace in decoding the CD signals of user-2 of pair-A. This interference subspace is given by $\mathcal{C}(\mathbf{H}_{A1}) \cup \mathcal{C}(\mathbf{H}_{A2}\tilde{\mathbf{F}}^p_{A2}) \cup \mathcal{C}(\mathbf{H}_{B1}\tilde{\mathbf{F}}^p_{B1}) \cup \mathcal{C}(\mathbf{H}_{B2}\tilde{\mathbf{F}}^p_{B2})$, which is independent of $\{\tilde{\mathbf{F}}^c_{A1}, \mathbf{\Gamma}^c_{A1}, \mathbf{\Gamma}^p_{A1}\}$. Then we only need to analyze $R^c_{A1} + R^p_A$, with the effective channel given as

$$\bar{\mathbf{y}} = \mathbf{H}_{A1}\mathbf{F}^c_{A1}\mathbf{x}^c_{A1} + \sum_{i=1}^{2}\mathbf{H}_{Ai}\mathbf{F}^p_{Ai}\mathbf{x}^p_{Ai} + \mathbf{e}^p_A + \mathbf{n} \tag{23}$$

where $\bar{\mathbf{y}}$ is the remaining signal vector after cancelling the CD signals of pair-B and user-2 of pair-A from $\mathbf{y}$ in (8), and $\mathbf{e}^p_A$ is given in (16).

From (23), we first whiten the interference-plus-noise term of $\mathbf{e}^p_A + \mathbf{n}$, to get the effective channel $\tilde{\mathbf{H}}^p_{Ai} = (\mathbf{W}^p_A)^{-\frac{1}{2}}\mathbf{H}_{Ai}$, $i \in \{1,2\}$. We decode user-1 of pair-A by treating other signals as interference, with the achievable rate given by

$$R^c_{A1} = \frac{1}{2}\log_2\left|\mathbf{I} + \tilde{\mathbf{H}}^p_{A1}\mathbf{F}^c_{A1}(\tilde{\mathbf{H}}^p_{A1}\mathbf{F}^c_{A1})^T + \sum_{i=1}^{2}\tilde{\mathbf{H}}^p_{Ai}\mathbf{F}^p_{Ai}(\tilde{\mathbf{H}}^p_{Ai}\mathbf{F}^p_{Ai})^T\right| - \frac{1}{2}\log_2\left|\mathbf{I} + \sum_{i=1}^{2}\tilde{\mathbf{H}}^p_{Ai}\mathbf{F}^p_{Ai}(\tilde{\mathbf{H}}^p_{Ai}\mathbf{F}^p_{Ai})^T\right|. \tag{24}$$

Then, the CD signal of user-1 of pair-A is cancelled, and we decode the PNC signals of pair-A. At high SNR, the achievable rate is given by (20), i.e.

$$R^p_A = \sum_{i=1}^{2}\frac{1}{2}\log_2\left|\mathbf{I} + \tilde{\mathbf{H}}^p_{Ai}\mathbf{F}^p_{Ai}(\tilde{\mathbf{H}}^p_{Ai}\mathbf{F}^p_{Ai})^T\right| + \sum_{j=1}^{k_A}\log_2\cos^2\frac{\theta^p_{A,j}}{2}. \tag{25}$$

In the above, $\theta^p_{A,j}$ is the j-th principal angle of $\mathcal{C}(\tilde{\mathbf{H}}^p_{A1}\tilde{\mathbf{F}}^p_{A1})$ and $\mathcal{C}(\tilde{\mathbf{H}}^p_{A2}\tilde{\mathbf{F}}^p_{A2})$, and is independent of $\{\tilde{\mathbf{F}}^c_{A1}, \mathbf{\Gamma}^c_{A1}, \mathbf{\Gamma}^p_{A1}\}$. Also, it can be shown that (cf(19))

$$\frac{1}{2}\log_2\left|\mathbf{I} + \sum_{i=1}^{2}\tilde{\mathbf{H}}^p_{Ai}\mathbf{F}^p_{Ai}(\tilde{\mathbf{H}}^p_{Ai}\mathbf{F}^p_{Ai})^T\right| \approx \sum_{i=1}^{2}\frac{1}{2}\log_2\left|\mathbf{I} + \tilde{\mathbf{H}}^p_{Ai}\mathbf{F}^p_{Ai}(\tilde{\mathbf{H}}^p_{Ai}\mathbf{F}^p_{Ai})^T\right| + \sum_{j=1}^{k_A}\frac{1}{2}\log_2\sin^2\theta^p_{A,j}. \tag{26}$$

Combining (24)-(26), we see that maximizing $R^c_{A1} + R^p_A$ over $\{\tilde{\mathbf{F}}^c_{A1}, \mathbf{\Gamma}^c_{A1}, \mathbf{\Gamma}^p_{A1}\}$ is equivalent to maximizing $\left|\mathbf{I} + \tilde{\mathbf{H}}^p_{A1}\mathbf{F}_{A1}(\tilde{\mathbf{H}}^p_{A1}\mathbf{F}_{A1})^T + \tilde{\mathbf{H}}^p_{A2}\mathbf{F}^p_{A2}(\tilde{\mathbf{H}}^p_{A2}\mathbf{F}^p_{A2})^T\right|$, by noting $\mathbf{F}_{A1}\mathbf{F}^T_{A1} = \mathbf{F}^c_{A1}(\mathbf{F}^c_{A1})^T + \mathbf{F}^p_{A1}(\mathbf{F}^p_{A1})^T$. As $\tilde{\mathbf{H}}^p_{A2}\mathbf{F}^p_{A2}$ is fixed, we only need to consider the following optimization problem:

$$\begin{aligned}
\max_{\tilde{\mathbf{F}}^c_{A1}, \mathbf{\Gamma}^c_{A1}, \mathbf{\Gamma}^p_{A1}} \quad & \left|\mathbf{I} + \bar{\mathbf{H}}_{A1}\mathbf{F}_{A1}\mathbf{F}^T_{A1}\bar{\mathbf{H}}^T_{A1}\right| \\
\text{s.t.} \quad & (\tilde{\mathbf{F}}^c_{A1})^T\tilde{\mathbf{F}}^c_{A1} = \mathbf{I}, \\
& \text{tr}\{(\mathbf{\Gamma}^p_{A1})^T\mathbf{\Gamma}^p_{A1}\} + \text{tr}\{(\mathbf{\Gamma}^c_{A1})^T\mathbf{\Gamma}^c_{A1}\} \le P_{A1}, \\
& l \in \{A, B\},\ i \in \{1,2\}.
\end{aligned} \tag{27}$$

where $\bar{\mathbf{H}}_{A1} = (\mathbf{I} + \tilde{\mathbf{H}}^p_{A2}\mathbf{F}^p_{A2}(\tilde{\mathbf{H}}^p_{A2}\mathbf{F}^p_{A2})^T)^{-\frac{1}{2}}\tilde{\mathbf{H}}^p_{A1}$ is the effective channel after whitening.

Note that at high SNR

$$\begin{aligned}
& \left|\mathbf{I} + \bar{\mathbf{H}}_{A1}\mathbf{F}_{A1}\mathbf{F}^T_{A1}\bar{\mathbf{H}}^T_{A1}\right| \\
& \approx \left|\tilde{\mathbf{H}}^T_{A1}\tilde{\mathbf{H}}_{A1}\mathbf{F}_{A1}\mathbf{F}^T_{A1}\right| \\
& = \left|\tilde{\mathbf{H}}^T_{A1}\tilde{\mathbf{H}}_{A1}\right|\left|\mathbf{F}_{A1}\mathbf{F}^T_{A1}\right|.
\end{aligned} \tag{28}$$

Our problem is simplified to maximize $\left|\mathbf{F}_{A1}\mathbf{F}^T_{A1}\right| = \left|\tilde{\mathbf{F}}^c_{A1}\Gamma^c_{A1}(\Gamma^c_{A1})^T(\tilde{\mathbf{F}}^c_{A1})^T + \tilde{\mathbf{F}}^p_{A1}\Gamma^p_{A1}(\Gamma^p_{A1})^T(\tilde{\mathbf{F}}^p_{A1})^T\right|$. The optimum is achieved when $(\mathbf{F}_{A1})^T\mathbf{F}_{A1} = P_{A1}/n\mathbf{I}$, or equivalently $(\tilde{\mathbf{F}}^c_{A1})^T\tilde{\mathbf{F}}^p_{A1} = \mathbf{0}$, $(\mathbf{\Gamma}^c_{A1})^T\mathbf{\Gamma}^c_{A1} = P_{A1}/n\mathbf{I}$ and $(\mathbf{\Gamma}^p_{A1})^T\mathbf{\Gamma}^p_{A1} = P_{A1}/n\mathbf{I}$. This completes the proof of Proposition 3.

C. Approximate Design of PNC Precoders

With Proposition 3, the precoder optimization problem in (22) reduces to

$$\begin{aligned}
\max_{\tilde{\mathbf{F}}^p_{li}} \quad & \sum_{l \in \{A,B\}}\sum_{s \in \{p,c\}} R^s_l \\
\text{s.t.} \quad & (\tilde{\mathbf{F}}^p_{li})^T\tilde{\mathbf{F}}^p_{li} = \mathbf{I},\ l \in \{A, B\},\ i \in \{1,2\}.
\end{aligned} \tag{29}$$

Problem (29) is still difficult to solve. Here we propose an approximate solution. First, instead of maximizing the sum-rate in (29), we target at maximizing the sum-rate of PNC decoding, i.e., $R^p_A + R^p_B$. Then we only need to study the signal model involved in PNC decoding:

$$\mathbf{y}^p_l = \sum_{i=1}^{2}\sqrt{\frac{P_{Ai}}{n}}\mathbf{H}_{Ai}\tilde{\mathbf{F}}^p_{Ai}\mathbf{x}^p_{Ai} + \sum_{i=1}^{2}\sqrt{\frac{P_{Bi}}{n}}\mathbf{H}_{Bi}\tilde{\mathbf{F}}^p_{Bi}\mathbf{x}^p_{Bi} + \mathbf{n}. \tag{30}$$

In the above, $\Gamma_{li}^p = \sqrt{P_{li}/n}\mathbf{I}$, $\forall l, i$ are assumed, since we are only interested in the precoder design at high SNR. Now consider the PNC decoding of pair-l, $l \in \{A, B\}$ by treating the other pair as interference. Recall that $\mathbf{W}_l^p$ is the covariance matrix of the interference-plus-noise term. Then we have

$$(\mathbf{W}_l^p)^{-\frac{1}{2}}\mathbf{y}_l^p = \sqrt{\frac{P_{l1}}{n}}\tilde{\mathbf{H}}_{l1}^p\mathbf{F}_{l1}^p\mathbf{x}_{l1}^p + \sqrt{\frac{P_{l2}}{n}}\tilde{\mathbf{H}}_{l2}^p\mathbf{F}_{l2}^p\mathbf{x}_{l2}^p + \mathbf{n}_l^p. \quad (31)$$

Clearly, $\mathcal{C}(\tilde{\mathbf{H}}_{li}^p\tilde{\mathbf{F}}_{li}^p) \subset \mathcal{C}(\tilde{\mathbf{H}}_{li}^p)$, $\forall\, l, i$. As inspired by the SD approach for the one-pair case, our purpose for the design of $\tilde{\mathbf{F}}_{li}^p$ is to ensure that $\mathcal{C}(\tilde{\mathbf{H}}_{l1}^p\tilde{\mathbf{F}}_{l1}^p)$ and $\mathcal{C}(\tilde{\mathbf{H}}_{l2}^p\tilde{\mathbf{F}}_{l2}^p)$ are as parallel to each other as possible. Therefore, we choose $\tilde{\mathbf{F}}_{li}^p$ such that the PNC signal space $\mathcal{C}(\tilde{\mathbf{H}}_{l1}^p\tilde{\mathbf{F}}_{l1}^p)$ is aligned to the subspace spanned by the k_l principal vector pairs with corresponding the first k_l principal angles of $\mathcal{C}(\tilde{\mathbf{H}}_{l1}^p)$, i.e.

$$\mathcal{C}(\tilde{\mathbf{H}}_{li}^p\tilde{\mathbf{F}}_{li}^p) = \mathcal{C}(\tilde{\mathbf{U}}_l^{k_l}) \text{ and } \mathcal{C}(\tilde{\mathbf{H}}_{li}^p\tilde{\mathbf{F}}_{li}^p) = \mathcal{C}(\tilde{\mathbf{V}}_l^{k_l}) \quad (32)$$

where $\tilde{\mathbf{U}}_l^{k_l} = [\tilde{\mathbf{u}}_{l,1}, ..., \tilde{\mathbf{u}}_{l,k_l}]$ and $\tilde{\mathbf{V}}_l^{k_l} = [\tilde{\mathbf{v}}_{l,1}, ..., \tilde{\mathbf{v}}_{l,k_l}]$, with $\tilde{\mathbf{u}}_{l,j}$ and $\tilde{\mathbf{v}}_{l,j}$ being the j-th principal vector of $\tilde{\mathbf{H}}_{l1}^p$ and $\tilde{\mathbf{H}}_{l2}^p$. We have the following proposition.

Proposition 4: The orthonormal matrix $\tilde{\mathbf{F}}_{li}^p$, $i \in \{1, 2\}$, satisfying (32), are given by

$$\tilde{\mathbf{F}}_{l1}^p = \mathbf{A}_{l1}\mathbf{U}_l^{k_l}\left[(\mathbf{U}_l^{k_l})^T\mathbf{A}_{l1}^2\mathbf{U}_l^{k_l}\right]^{-\frac{1}{2}} \quad (33)$$

$$\tilde{\mathbf{F}}_{l2}^p = \mathbf{A}_{l2}\mathbf{V}_l^{k_l}\left[(\mathbf{V}_l^{k_l})^T\mathbf{A}_{l2}^2\mathbf{V}_l^{k_l}\right]^{-\frac{1}{2}} \quad (34)$$

where $\mathbf{A}_{li} = [(\tilde{\mathbf{H}}_{li}^p)^T\tilde{\mathbf{H}}_{li}^p]^{-1}$.

Proof: Since $\mathcal{C}(\tilde{\mathbf{H}}_{li}^p\tilde{\mathbf{F}}_{li}^p) = \mathcal{C}(\tilde{\mathbf{U}}_l^{k_l})$, we can write that $\tilde{\mathbf{H}}_{li}^p\tilde{\mathbf{F}}_{li}^p = \mathbf{U}_l^{k_l}\mathbf{G}_{l1}$, where $\mathbf{G}_{l1} \in \mathbb{R}^{k_l \times k_1}$. Then $\tilde{\mathbf{F}}_{l1}^p = [(\tilde{\mathbf{H}}_{li}^p)^T\tilde{\mathbf{H}}_{li}^p]^{-1}\mathbf{U}_l^{k_l}\mathbf{G}_{l1}$. Since $(\tilde{\mathbf{F}}_{l1}^p)^T\tilde{\mathbf{F}}_{l1}^p = \mathbf{I}$ and $\mathbf{G}_{l1}$ is invertible, we have $\mathbf{G}_{l1} = \left[(\mathbf{U}_l^{k_l})^T A_{l1}^2\mathbf{U}_l^{k_l}\right]^{-\frac{1}{2}}$. The expression of $\tilde{\mathbf{F}}_{l2}^p$ can be derived similarly. This concludes the proof of Proposition 4.

It is worth noting that we cannot directly obtain $\tilde{\mathbf{F}}_{l1}^p$ and $\tilde{\mathbf{F}}_{l2}^p$ from (33) and (34). The reason is that the PNC decoders for the two user-pairs are coupled with each other. To see this, we recall that $\mathbf{W}_A^p = \sum_{i=1}^{2}\sqrt{\frac{P_{Bi}}{n}}\tilde{\mathbf{H}}_{Bi}^p\tilde{\mathbf{F}}_{Bi}^p(\tilde{\mathbf{H}}_{Bi}^p\tilde{\mathbf{F}}_{Bi}^p)^T + \mathbf{I}$. Therefore, $\tilde{\mathbf{F}}_{Ai}^p$ depends on $\tilde{\mathbf{F}}_{Bi}^p$. Similarly, $\tilde{\mathbf{F}}_{Bi}^p$ depends on $\tilde{\mathbf{F}}_{Ai}^p$. To decouple these two precoders, we choose $\mathbf{W}_A^p = \sum_{i=1}^{2}\sqrt{\frac{P_{Bi}}{n}}\tilde{\mathbf{H}}_{Bi}^p(\tilde{\mathbf{H}}_{Bi}^p)^T + \mathbf{I}$ in determining $\tilde{\mathbf{F}}_{Ai}^p$ (using (33) and (34)), and then determine $\tilde{\mathbf{F}}_{Bi}^p$ based on $\tilde{\mathbf{F}}_{Ai}^p$. Of course, we can alternatively first design $\tilde{\mathbf{F}}_{Bi}^p$ and then design $\tilde{\mathbf{F}}_{Ai}^p$. We choose the one with a higher achievable rate in simulation. Once the PNC precoders are obtained, we can directly determine the CD precoders using Proposition 3.

D. Asymptotic Rate Analysis

As mentioned before, when $m \geq 3n$, the proposed SD approach achieves the same DoF as the cut-set bound does. We next investigate the rate gap between the proposed SD approach and the cut-set bound in the high SNR regime from the viewpoint of principal angles.

We start with introducing some notations. Let

$$C = \sum_{l\in\{A,B\}}\sum_{i\in\{1,2\}}\left\{\frac{1}{2}\log_2\left|\mathbf{I} + \frac{P_{li}}{n}\mathbf{H}_{li}\mathbf{H}_{li}^T\right|\right\}. \quad (35)$$

Let $\mathbf{H}_{li}\mathbf{F}_{li}^s = \mathbf{H}_{li}^s$, $\forall s, l, i$; $\theta_{\{li,\bar{l}\bar{i}\},j}$ be the principal angles of $\mathcal{C}(\mathbf{H}_{li})$ and $\mathcal{C}(\mathbf{H}_{\bar{l}\bar{i}})$; $\theta_{\{li^s,\bar{l}\bar{i}^{\bar{s}}\},j}$ be the principal angles of $\mathcal{C}(\mathbf{H}_{li}^s)$ and $\mathcal{C}(\mathbf{H}_{\bar{l}\bar{i}}^{\bar{s}})$; $\theta_{\{li^s,\bar{l}^{\bar{s}}\},j}$ be the principal angles of $\mathcal{C}(\mathbf{H}_{li}^s)$ and $\mathcal{C}([\mathbf{H}_{\bar{l}1}^{\bar{s}}, \mathbf{H}_{\bar{l}2}^{\bar{s}}])$; $\theta_{\{li^s,l^{\bar{s}}\bar{l}^{\bar{s}}\},j}$ be the principal angles of $\mathcal{C}(\mathbf{H}_{li}^s)$ and $\mathcal{C}([\mathbf{H}_{l1}^{\bar{s}}, \mathbf{H}_{l2}^{\bar{s}}, \mathbf{H}_{\bar{l}1}^{\bar{s}}, \mathbf{H}_{\bar{l}2}^{\bar{s}}])$, where $s \in \{p, c\}$, $l \in \{A, B\}$, $i \in \{1, 2\}$, and $\bar{s}$, $\bar{l}$, and $\bar{i}$ are respectively the complements of s, l, and i.

Proposition 5: The asymptotic rate R_{SD} of the proposed SD approach satisfies $C - R_{SD} \to \Delta_{SD}$, as $P_{li} \to \infty$, $\forall l, i$, with

$$\begin{aligned}
\Delta_{SD} =& -\sum_{l\in\{A,B\}}\sum_{i\in\{1,2\}}\sum_{j=1}^{n-k_l}\left[\frac{1}{2}\log_2\sin^2\theta_{\{li^c,A^pB^p\},j}\right] \\
&-\sum_{i\in\{1,2\}}\sum_{j=1}^{n-k_A}\left[\frac{1}{2}\log_2\sin^2\theta_{\{Ai^c,B^c\},j}\right] \\
&-\sum_{i\in\{1,2\}}\sum_{j=1}^{k_A}\left[\frac{1}{2}\log_2\sin^2\theta_{\{Ai^p,B^p\},j}\right] \\
&-\sum_{i\in\{1,2\}}\sum_{j=1}^{k_B}\left[\frac{1}{2}\log_2\sin^2\theta_{\{Bi^p,A^p\},j}\right] \\
&-\sum_{l\in\{A,B\}}\sum_{j=1}^{n-k_l}\left[\frac{1}{2}\log_2\sin^2\theta_{\{l1^c,l2^c\},j}\right] \\
&-\sum_{l\in\{A,B\}}\sum_{j=1}^{k_l}\left[\log_2\cos^2\frac{\theta_{\{l1^p,l2^p\},j}}{2}\right].
\end{aligned} \quad (36)$$

Remark 4: Note that Δ_{SD} depends on the choice of k_l (i.e. the number of PNC spatial streams for pair-l, $l \in \{A, B\}$). This implies that Δ_{SD} can be minimized by enumerating over all possible choices of k_A and k_B.

Proposition 6: The asymptotic cut-set bound satisfies $C - R_{CS} \to \Delta_{CS}$, as $P_{li} \to \infty$, $\forall l, i$, with

$$\begin{aligned}
\Delta_{CS} = \max\Bigg\{&-\sum_{i=1}^{2}\sum_{j=1}^{n}\left[\frac{1}{2}\log_2\sin^2\theta_{\{Ai,Bi\},j}\right], \\
&-\sum_{i=1}^{2}\sum_{j=1}^{n}\left[\frac{1}{2}\log_2\sin^2\theta_{\{Ai,B\{3-i\}\},j}\right]\Bigg\}.
\end{aligned} \quad (37)$$

The proofs of Propositions 5 and 6 are straightforward but tedious. We omit the details here due to space limitation. With Propositions 5 and 6, we can readily obtain the rate gap between the proposed SD method and the cut-set bound.

VI. Numerical Results

In this section, numerical results are provided to demonstrate. In Fig. 2, we compare the rate of the SD method with the cut-set bound, the pure PNC method ($k_l = n$) and the pure CD method ($k_l = 0$). We assume $P_{li} = P$, for $l \in \{A, B\}$,

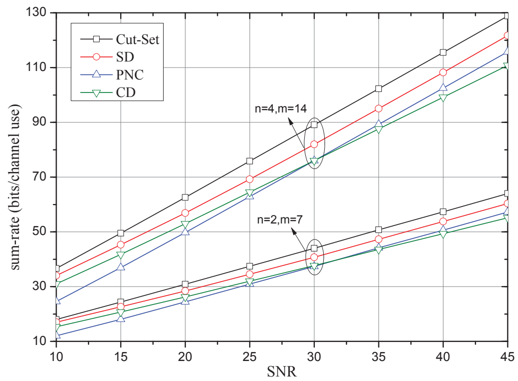

Fig. 2. Rate comparison of three different methods and the cut-set bound

$i \in \{1, 2\}$, and equal power allocation is also assumed for each spatial stream. Two cases of antenna configuration are considered: $\{m=7, n=2\}$ and $\{m=14, n=4\}$. We see that the SD method has at least 2.4 dB (for $n=2$) and 2.2 dB (for $n=4$) power gain compared with the pure PNC method and CD method, and has 2.6 dB (for $n=2$) and 2.8 dB (for $n=4$) rate gap compared with the cut-set bound. Note that the pure CD method performs better than the pure PNC method at low SNR, and the case is reversed at high SNR.

Fig. 3 illustrates the asymptotical rate gap between the SD method and the cut-set bound at high SNR. The number of user antennas varies from 1 to 4, while the number of relay antennas from 3 to 18. Note that we only consider the antenna set-up of $m \geq 3n$, in which the DoF of the system is n per user. From Fig. 3, for $m=3n, 4n, 5n$, the rate gap is about 1.2 bits, 0.33 bits, 0.21 bits per user-antenna, respectively. The results also show that, for fixed n, the rate gap vanishes as increasing the number of relay antenna. This is expected since the user channels tend to be orthogonal to each other as $m \to \infty$.

VII. Conclusion

In this paper, based on the concept of principal angles, we develop an extended space-division approach for the MAC phase of two-pair TWRCs over a single relay. We provide the optimal decoding strategy and optimal precoder design principles for the maximization of the sum-rate. Due to its non-convex nature, we propose a sub-optimal precoder design algorithm and derive the closed-form expression of the rate gap with respect to the cut-set bound in the high SNR regime. In two case studies, numerical results show that the proposed SD method outperforms the pure PNC and CD methods with more than 2dB power improvement.

Apart from the detailed results, an important contribution of this paper is the use of the classical concept of principal angles in linear algebra to delineate the PNC and CD decoding methods in optimal multi-pair TWRCs design. In particular, this tool gives us a systematic way to study the interplay between signal and interference alignments in the existence of multiple users. Going forward, there are many interesting directions for further exploration. For example, this paper only focuses on the MAC phase. The precoder design for the BC phase is needed to complete the overall design. In addition, the design in this paper is guided by the asymptotic analysis targeted for the high SNR regime. The studies for the low and medium SNR regimes remain to be conducted.

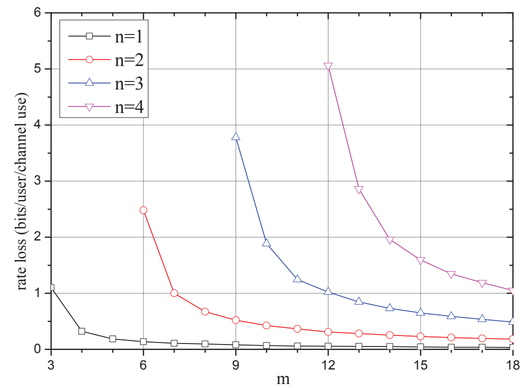

Fig. 3. Rate gap between the SD method and the cut-set bound

Acknowledgment

This work is partially supported by AoE grant E-02/08 and the General Research Funds Project No. 414911 and No. 418712, established under the University Grant Committee of the Hong Kong Special Administrative Region, China; and the China 973 Program Project No. 2012CB315904, and the China NSFC grant Project No. 61271277.

References

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

[2] B. Nazer and M. Gastpar, "The case for structured random codes in network communication theorems," *ITW*, Lake Tahoe, California, 2007.

[3] P. Popovski and H. Yomo, "Physical network coding in two-way wireless relay channels," *IEEE ICC*, Glasgow, Scotland, June 2007.

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

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

[6] A. Sezgin, A. Avestimehr, M. Khajehnejad, B. Hassibi, "Divide-and-Conquer: Approaching the Capacity of the Two-Pair Bidirectional Gaussian Relay Network," *IEEE Trans. Inform. Theory*, vol.58, Apr. 2012.

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

[8] H. Xin, Y. Peng, C. Wang, Y. Yang, W. Wang, "Coordinated Eigen Beamforming for Multi-Pair MIMO Two-Way Relay Network," *IEEE GLOBECOM*, Dec. 2011.

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

[10] C. Sun, C. Yang, Y. Li, and B. Vucetic, "Transceiver design for multi-user multi-antenna two-way relay cellular systems," *IEEE Trans. Commun.*,vol. 60, no. 10, Oct. 2012.

[11] F. Wang, S. C. Liew, and D. Guo, "Wireless mimo switching with zero forcing and network coding," *IEEE J. Select. Areas Commun.*, vol. 30, no. 8, pp. 1452-1463, Sep. 2012.

[12] 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.

[13] X. Yuan, T. Yang, and I. B. Collings, "MIMO two-way relaying: A space-division approach," submitted to *IEEE Trans. Inform. Theory*.

[14] 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.

[15] J. Miao and A. Ben-Israel, "On principle angles between subspaces in Rn, " *Lin. Algeb. and its Appl.*, vol. 171, pp. 81-98, Nov. 1992.

[16] G. H. Golub and C. F. V. Loan, *Matrix Computation, 3rd ed.* Baltimore, MD: The Johns Hopkins University Press, 1997.