Achievable Sum-Rate of MU-MIMO Cellular Two-Way Relay Channels: Lattice Code-Aided Linear Precoding

Hyun Jong Yang, *Member*, *IEEE*, Youngchol Choi, *Member*, *IEEE*, Namyoon Lee, *Student Member*, *IEEE*, and Arogyaswami Paulraj, *Fellow*, *IEEE*

***Abstract*—We derive a new sum-rate lower bound of the multiuser multi-input multi-output (MU-MIMO) cellular two-way relay channel (cTWRC) which is composed of a base station (BS) and a relay station (RS), both with multiple antennas, and non-cooperative mobile stations (MSs), each with a single antenna. In the first phase, we show that network coding based on decode-and-forward relaying can be generalized to arbitrary input cardinality through proposed lattice code-aided linear precoding, despite the fact that precoding is permitted only at the BS due to non-cooperation among the MSs. In addition, a new sum-rate lower bound for the second phase is derived by showing that the two spatial decoding orders at the BS and MSs for one-sided zero-forcing dirty-paper-coding must be identical. From the fundamental gain of network coding, our sum-rate lower bound achieves the full multiplexing gain regardless of the number of antennas at the BS or RS, and strictly exceeds the previous lower bound which is based on traditional multiuser decoding in the first phase. Furthermore, it is shown that our lower bound asymptotically achieves the sum-rate upper bound in the presence of signal-to-noise ratio (SNR) asymmetry in high SNR regime, and sufficient conditions for this SNR asymmetry are drawn.**

***Index Terms*—Multiuser (MU)-MIMO cellular two-way relay channel (cTWRC), decode-and-forward (DF), nested lattice codes, linear precoding, achievable sum-rate**

I. Introduction

A MOBILE cellular communication network is one of the rapidly growing and most demanding telecommunication systems. Not only the coverage but also the capacity of the cellular communication network can be effectively improved by wireless relays [1], and the capacity can be further increased by multi-input multi-output (MIMO) technologies [2]–[4]. Two-way relaying (TWR) [5]–[7] is regarded as a promising relay protocol to enhance the capacity of relay networks significantly with effective resource utilization. In this paper, we address the capacity of the multiuser (MU)-MIMO cellular two-way relay channel (cTWRC) which is composed of a base station (BS) and relay station (RS) with multiple antennas and non-cooperative mobile stations (MSs), each with a single antenna. This channel has a great potential to increase the capacity of cellular networks, because the gain of the TWR and MIMO techniques can be simultaneously obtained [8]–[11].

In communication theory, it is a fundamental problem to characterize the sum-rate upper and lower bounds of a given channel model through capacity analysis [1], [12]. The sum-rate upper bound is the theoretical upper limit on the sum-rate, and the sum-rate lower bound refers to the sum-rate that can be achieved by a specific encoding and decoding methodology. The cut-set bound, which is given from Shannon's cut-set theorem [12], is the tightest sum-rate upper bound of the TWRC to date. The cut-set bound of the TWRC is twice the maximum sum-rate of the one-way relaying scheme [1]. Therefore, the main challenge is to find an achievable scheme with a sum-rate that is as close to the cut-set bound as possible.

The TWR protocol, which was inspired by the concept of the two-way channel [13] and the principle of binary XOR *network coding* [14], doubles the resource efficiency compared to the one-way relaying protocol which requires four phases (time slots, frequency bands, or orthogonal code resources) to complete a message exchange between the communication nodes through the relay node [5]–[7]. That is, TWR takes only two phases, denoted by Phase 1 and 2. The development of a symbol or signal with joint information, which is sufficient for each communication node to retrieve the designated information using the known side information, is referred to as *network coding* [6], [15], [16]. Specifically, in the three-node TWRC [5]–[7], the binary XOR network-coded symbol $x_{\mathrm{NC}} = x_1 \oplus x_2$ is retrieved in Phase 1 from the received signal at the relay node, $y_{\mathrm{R}} = x_1 + x_2 + z_{\mathrm{R}}$, where $x_i \in \{0, 1\}$ is the binary symbol transmitted from communication node i, $i = 1, 2$, and z_{R} is the additive noise. In Phase 2, the relay node broadcasts the retrieved x_{NC} to the two communication nodes, and communication node i can obtain the designated symbol from $x_{3-i} = x_{\mathrm{NC}} \oplus x_i$ using the known information of x_i.

To obtain higher rate than in the binary case, XOR network coding needs to be extended. A lattice code [17], [18] is the only known channel code that generalizes XOR network coding to arbitrary input cardinality. It was shown that the cut-set bounds of the three-node single-input single-output

Manuscript received 1 August 2011; revised 1 May 2012. This work was supported in part by the Ministry of Land, Transport and Maritime Affairs of Korea under Grant D10813810H380000110. This work was also supported in part by the NSF Programmable Open Mobile Internet (POMI) 2020 project [Award number: 0832820].

H. J. Yang and A. Paulraj are with Stanford Univ., Stanford, CA (e-mail: {hjdbell, apaulraj}@stanford.edu).

Y. Choi is with Korea Institute of Ocean Science and Technology, Daejeon, Republic of Korea (e-mail: ycchoi@kiost.ac).

N. Lee is with Wireless Networking and Communications Group, The Univ. of Texas at Austin, TX (e-mail: namyoon.lee@utexas.edu).

Digital Object Identifier 10.1109/JSAC.2012.120902.

(SISO) and MIMO TWRCs can be achieved asymptotically by decode-and-forward (DF) relaying which employs network coding generalized by lattice codes [16], [19]–[23]. In the case of the MIMO TWRC, linear precoding is required at both of the two communication nodes to separate the MIMO spatial streams into multiple SISO TWRCs [21]–[23]. These results of the three-node TWRCs have been extended to the multi-pair TWRC [24]–[32].

For the MU-MIMO cTWRC, although the importance of the channel model is well understood from the case of the MU-MIMO cellular one-way relay channel [2], [4], there have been only a few studies. The amplify-and-forward (AF) TWR scheme based on zero-forcing (ZF) precoding at both the BS and RS was proposed to obtain the full multiplexing gain[1] [9], and the sum-rate of this scheme is further increased by employing dirty-paper coding (DPC) in Phase 2 [34]. However, the cut-set bound of the MU-MIMO cTWRC cannot be achieved by AF TWR on account of the error propagation at the relay node, and the achievable rates of the AF TWR schemes are decreased even more in the presence of signal-to-noise ratio (SNR) asymmetry [9].

As shown in the three-node case, it is DF relaying that achieves the cut-set bound, in which the RS eliminates the additive noise from the decoding procedure. Therefore, the main focus of this paper is DF relaying. Sum-rate lower bounds of DF relaying for the MU-MIMO cTWRC were derived in [8], [35] using the multiple-access channel (MAC) bound for Phase 1. However, these bounds are subject to multiplexing gain loss compared to the cut-set bound unless the number of RS antennas is larger than twice the number of BS antennas, because all of the codewords from the BS and MSs are individually decoded at the RS without network coding. Even if the aforementioned condition of the numbers of antennas for the full multiplexing gain is satisfied, the bounds exhibit a constant SNR gap, compared to the cut-set bound. Therefore, the cut-set bound cannot be achieved by the sum-rate lower bounds in the aforementioned studies [8], [35] in any SNR regime.

To the best of the authors' knowledge, the capacity region of the MU-MIMO cTWRC is open, and no achievable scheme that attains the cut-set bound in terms of the sum-rate is known for any SNR regime even with the global channel state information (CSI) at all nodes, i.e., error-free feedbacks of CSI among all nodes.

A. Contributions and Novelty

In this paper, we propose an achievable DF scheme for the MU-MIMO cTWRC. Our main contributions are twofold.

1) The sum-rate lower bound achieved by the proposed scheme is the best known to date. In particular, network coding is made feasible in a DF manner, even with the restriction of non-cooperative MSs. From the fundamental gain of network coding, the full multiplexing gain is obtained regardless of the number of antennas at the BS or RS.
2) Our sum-rate lower bound achieves the cut-set bound in high SNR regime if the SNR asymmetry is sufficiently large. We derive sufficient conditions for the SNR asymmetry to achieve the cut-set bound asymptotically. In practice, these asymmetric SNRs can be observed, for example, when the RS is in the line-of-sight of the BS and the MSs are located at the cell edge.

[1]The multiplexing gain denotes the total number of spatial streams exchanged between the BS and the MSs, defined by $\lim_{\mathrm{SNR}\to\infty} R_{\mathrm{sum}}/\log \mathrm{SNR}$ [33], where R_{sum} denotes the achievable sum-rate.

The key ingredient of Phase 1 in the proposed scheme is network coding generalized by linear precoding and lattice codes, which is the only known methodology that achieves the cut-set bound of any TWRC. In the three-node MIMO TWRC case, where there exists only one MS with multiple antennas, it is known that the two effective channel matrices in Phase 1, one from the BS to the RS and the other from the MS to the RS, should be identical to deploy generalized network coding [21]–[23]. To accomplish this requirement, orthogonal linear precoding is performed at both the BS and MS in the three-node case. Furthermore, DPC combined with random dithering is also necessary at both the BS and MS to reduce each spatial stream at the RS to a SISO additive white Gaussian noise (AWGN) TWRC [19]. In particular, random dithering is required to obtain the maximum achievable rate. However, unlike the three-node case, linear precoding for the MU-MIMO cTWRC is not obvious nor lossless in terms of the achievable rate, because cooperation among the MSs is not allowed. Moreover, DPC with random dithering cannot be performed at the MSs for the same reason.

To deploy network coding without precoding at the MSs, a lattice code-aided linear precoding scheme at the BS in Phase 1 is proposed. We propose to allow discrete-valued intersymbol interference (ISI) as a means of increasing the achievable rate, which results from the difference of the two effective channel matrices and not from random dithering as in the three-node case. The precoding matrix is not merely designed in the ZF-sense such that the two effective channel matrices are identical to each other, but is optimized such that the achievable sum-rate is maximized through this ISI. In spite of the discrete nature of the ISI, which is never addressed in the literature on the lattice theory, we show that the rate of $\log(\mathrm{SINR})$ is achievable for each spatial stream as in the SISO AWGN TWRC [19] with random dithering through the proposed lattice encoding and decoding schemes, where SINR is the signal-to-interference-and-noise ratio. From this result, the optimization of the precoding matrix can be reduced to a signomial optimization problem that guarantees a global optimal solution.

In Phase 2, similarly to [8], [24], [35], the cut-set bound is achieved asymptotically by one-sided ZF-DPC in the high SNR regime, which is the best known scheme. In [8], [24], [35], the two spatial decoding orders at the BS and MSs are independently optimized to maximize the achievable rate of Phase 2. However, we demonstrate that the decoding order at the BS must be identical to the decoding order at the MSs. Therefore, the rate lower bounds of Phase 2 derived in [8], [24], [35] cannot be achieved in general, and we derive a new rate lower bound for Phase 2.

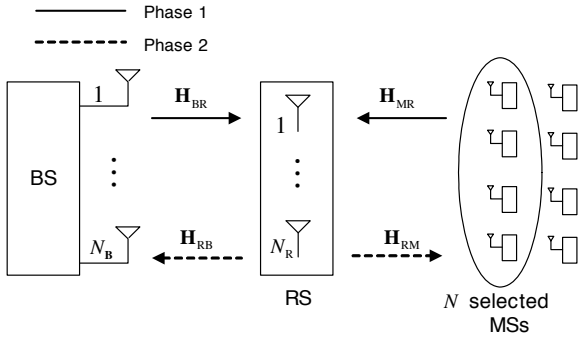


Fig. 1. Multiuser-MIMO cellular two-way relay channel

B. Organization and Notations

The remainder of this paper is organized as follows: Section II defines the MU-MIMO cTWRC and provides known sum-rate bounds. Section III presents the proposed scheme and its theoretical analysis, and Section IV evaluates the proposed scheme through numerical simulations. Section V concludes the paper.

Notations: $\mathbf{A}^{\mathrm{T}}$ and $\mathbf{A}^{\mathrm{H}}$ denote the transpose and the conjugate transpose of a matrix $\mathbf{A}$, respectively. $\Re(x)$ and $\Im(x)$ denote the real and imaginary components, respectively, of x. $\mathcal{CN}(\mu, \sigma^2)$ denotes the complex Gaussian distribution with mean μ and variance σ^2. $\mathbb{R}^{n\times m}$ and $\mathbb{C}^{n\times m}$ denote the $(n \times m)$-dimensional real and complex spaces, respectively. The diagonal matrix with diagonal elements $(a_1, a_2, \ldots, a_m)$ is denoted by $\mathrm{diag}(a_1, a_2, \ldots, a_m)$. $\mathbf{I}_m$ denotes the $m \times m$ identity matrix. $\| \cdot \|_{\mathrm{F}}$ denotes the Frobenius norm. $\log(\cdot)$ denotes the log-2 operation and $\ln(\cdot)$ denotes the natural logarithm. $[x]^+$ is defined as $[x]^+ \triangleq \max\{x, 0\}$ for any x. For the given permutation order $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_m)$, the permutation matrix is denoted by $\mathbf{\Pi}_{\boldsymbol{\beta}} = \left[\mathbf{e}_{\beta_1}{}^{\mathrm{T}}, \ldots, \mathbf{e}_{\beta_m}{}^{\mathrm{T}}\right]^{\mathrm{T}}$, where $\mathbf{e}_i$ is a row vector of length m with 1 in the ith position and 0 in every other position. For a given lattice Λ, $[\cdot]_\Lambda$ denotes the modulo-Λ operation [17] (See (58) in Appendix A-A).

II. System Model and Known Sum-Rate Bounds

A. System Model

We consider the MU-MIMO cTWRC as depicted in Fig. 1. The BS and RS are assumed to have N_{B} and N_{R} antennas, respectively, and each MS is assumed to have a single antenna. The number of selected MSs is $N \triangleq \min\{N_{\mathrm{B}}, N_{\mathrm{R}}\}$. The discrete memoryless channels with quasi-static and frequency-flat fading coefficients are considered [2], [4], [8]–[11], [15], [16], [19]–[32], [34]–[36]. The channel matrices from the BS to RS and from the RS to BS are denoted by $\mathbf{H}_{\mathrm{BR}} \in \mathbb{C}^{N_{\mathrm{R}}\times N_{\mathrm{B}}}$ and $\mathbf{H}_{\mathrm{RB}} \in \mathbb{C}^{N_{\mathrm{B}}\times N_{\mathrm{R}}}$, respectively. The channel vectors from the kth selected MS to RS and from the RS to kth selected MS are denoted by $\mathbf{h}_{\mathrm{MR},k} \in \mathbb{C}^{N_{\mathrm{R}}\times 1}$ and $\mathbf{h}_{\mathrm{RM},k} \in \mathbb{C}^{1\times N_{\mathrm{R}}}$, respectively. The horizontally stacked channel matrices between the RS and selected MSs are defined by $\mathbf{H}_{\mathrm{MR}} \triangleq [\mathbf{h}_{\mathrm{MR},1}, \ldots, \mathbf{h}_{\mathrm{MR},N}]$ and $\mathbf{H}_{\mathrm{RM}} \triangleq \left[\mathbf{h}_{\mathrm{RM},1}{}^{\mathrm{T}}, \ldots, \mathbf{h}_{\mathrm{RM},N}{}^{\mathrm{T}}\right]^{\mathrm{T}}$. It is assumed that each channel coefficient is generated from a continuous probability distribution so that each channel matrix for any instance has a full rank [37]. The two phases are assumed to be isolated, sidestepping the self-interference problem as assumed in [1], [2], [4]–[11], [15], [16], [19]–[32], [34]–[36].

In Phase 1, the integer message at the BS to be transmitted to the kth MS is denoted by $W_{\mathrm{B},k} \in \mathcal{W}_{\mathrm{B},k} \triangleq \{1, 2, \ldots, 2^{TR_{\mathrm{B},k}}\}$ with a transmission rate of $R_{\mathrm{B},k}$. The corresponding codeword and the transmit signal, denoted by $\mathbf{x}_{\mathrm{B},k} \in \mathbb{C}^{1\times T}$ and $\mathbf{S}_{\mathrm{B}} \in \mathbb{C}^{N_{\mathrm{B}}\times T}$, respectively, are given with the encoding function $f_{\mathrm{B},k}$ and the MIMO preprocessing function u_{B} as

$$\mathbf{x}_{\mathrm{B},k} = f_{\mathrm{B},k}\left(W_{\mathrm{B},k}\right), k = 1, \ldots, N, \tag{1}$$

$$\mathbf{S}_{\mathrm{B}} = u_{\mathrm{B}}\left(\mathbf{x}_{\mathrm{B},1}, \ldots, \mathbf{x}_{\mathrm{B},N}\right). \tag{2}$$

Similarly, the integer message at the kth MS to be transmitted to the BS is denoted by $W_{\mathrm{M},k} \in \mathcal{W}_{\mathrm{M},k} \triangleq \{1, 2, \ldots 2^{TR_{\mathrm{M},k}}\}$ with a transmission rate of $R_{\mathrm{M},k}$. The corresponding codeword and the transmit signal, denoted by $\mathbf{x}_{\mathrm{M},k} \in \mathbb{C}^{1\times T}$ and $\mathbf{s}_{\mathrm{M},k} \in \mathbb{C}^{1\times T}$, respectively, are given with the encoding function $f_{\mathrm{M},k}$ and the preprocessing function $u_{\mathrm{M},k}$ as

$$\mathbf{x}_{\mathrm{M},k} = f_{\mathrm{M},k}\left(W_{\mathrm{M},k}\right), \tag{3}$$

$$\mathbf{s}_{\mathrm{M},k} = u_{\mathrm{M},k}\left(\mathbf{x}_{\mathrm{M},k}\right), \quad k = 1, \ldots, N. \tag{4}$$

The maximum transmit powers per dimension at the BS and the kth MS are denoted by P_{B} and $P_{\mathrm{M},k}$, respectively, i.e., $\frac{1}{T}E\left\{\|\mathbf{S}_{\mathrm{B}}\|_{\mathrm{F}}^2\right\} \le P_{\mathrm{B}}$, $\frac{1}{T}E\left\{\|\mathbf{s}_{\mathrm{M},k}\|_{\mathrm{F}}^2\right\} \le P_{\mathrm{M},k}$. The codeword matrices are defined as $\mathbf{X}_\chi \triangleq \left[\mathbf{x}_{\chi,1}{}^{\mathrm{T}}, \ldots, \mathbf{x}_{\chi,N}{}^{\mathrm{T}}\right]^{\mathrm{T}}$, $\chi \in \{\mathrm{B}, \mathrm{M}\}$, for later notational simplicity, and the horizontally stacked matrix of $\mathbf{s}_{\mathrm{M},k}$ is defined by $\mathbf{S}_{\mathrm{M}} \triangleq \left[\mathbf{s}_{\mathrm{M},1}{}^{\mathrm{T}}, \ldots, \mathbf{s}_{\mathrm{M},N}{}^{\mathrm{T}}\right]^{\mathrm{T}}$. The BS and all selected MSs simultaneously transmit to the RS, and the received signal at the RS can be written by

$$\mathbf{Y}_{\mathrm{R}} = \mathbf{H}_{\mathrm{BR}}\mathbf{S}_{\mathrm{B}} + \mathbf{H}_{\mathrm{MR}}\mathbf{S}_{\mathrm{M}} + \mathbf{Z}_{\mathrm{R}}, \tag{5}$$

where $\mathbf{Z}_{\mathrm{R}}$ is the additive noise at the RS, each element of which is independent and identically distributed (i.i.d.) according to $\mathcal{CN}(0, 1)$.

In Phase 2, from the received signal $\mathbf{Y}_{\mathrm{R}}$, the RS generates the codeword matrix $\mathbf{X}_{\mathrm{R}} \in \mathbb{C}^{N\times T}$ and the transmit signal matrix $\mathbf{S}_{\mathrm{R}} \in \mathbb{C}^{N_{\mathrm{R}}\times T}$ from

$$\mathbf{X}_{\mathrm{R}} = f_{\mathrm{R}}\left(\mathbf{Y}_{\mathrm{R}}\right), \quad \mathbf{S}_{\mathrm{R}} = u_{\mathrm{R}}\left(\mathbf{X}_{\mathrm{R}}\right), \tag{6}$$

where f_{R} is an encoding function and u_{R} is a MIMO preprocessing function. The maximum transmit power per dimension at the RS is denoted by P_{R}, i.e., $\frac{1}{T}E\left\{\|\mathbf{S}_{\mathrm{R}}\|_{\mathrm{F}}^2\right\} \le P_{\mathrm{R}}$. The RS broadcasts $\mathbf{S}_{\mathrm{R}}$ to the BS and MSs, and the received signals at the BS and the selected MSs are written by

$$\mathbf{Y}_{\mathrm{B}} = \mathbf{H}_{\mathrm{RB}}\mathbf{S}_{\mathrm{R}} + \mathbf{Z}_{\mathrm{B}}, \tag{7}$$

and

$$\mathbf{Y}_{\mathrm{M}} = \left[\mathbf{y}_{\mathrm{M},1}{}^{\mathrm{T}}, \ldots, \mathbf{y}_{\mathrm{M},N}{}^{\mathrm{T}}\right]^{\mathrm{T}} = \mathbf{H}_{\mathrm{RM}}\mathbf{S}_{\mathrm{R}} + \mathbf{Z}_{\mathrm{M}}, \tag{8}$$

respectively, where $\mathbf{y}_{\mathrm{M},k} \in \mathbb{C}^{1\times T}$ is the received signal at the kth MS, and $\mathbf{Z}_{\mathrm{B}}$ and $\mathbf{Z}_{\mathrm{M}}$ are the AGWN at the BS and the selected MSs, respectively. Each element of $\mathbf{Z}_{\mathrm{B}}$ and $\mathbf{Z}_{\mathrm{M}}$ is i.i.d. according to $\mathcal{CN}(0, 1)$. With the side information of the known message, the BS estimates the designated message $(W_{\mathrm{M},1}, \ldots, W_{\mathrm{M},N})$ from the received signal $\mathbf{Y}_{\mathrm{B}}$, and the kth

MS estimates $W_{\mathrm{B},k}$ from the received signal $\mathbf{y}_{\mathrm{M},k}$. Therefore, the error probability is defined as

$$P_e \triangleq \Pr\Big\{ d_{\mathrm{B}}\left(\mathbf{Y}_{\mathrm{B}}, W_{\mathrm{B},1}, \ldots, W_{\mathrm{B},N}\right) \neq \left(W_{\mathrm{M},1}, \ldots, W_{\mathrm{M},N}\right) \text{ or } \cup_{k=1}^{N} d_{\mathrm{M},k}(\mathbf{y}_{\mathrm{M},k}, W_{\mathrm{M},k}) \neq W_{\mathrm{B},k} \Big\}, \quad (9)$$

where d_{B} and $d_{\mathrm{M},k}$ are the decoding functions at the BS and the kth MS, respectively.

A rate tuple $(R_{\mathrm{B},1}, \ldots, R_{\mathrm{B},N}, R_{\mathrm{M},1}, \ldots, R_{\mathrm{M},N})$ is said to be *achievable* if there exist encoding functions $f_{\mathrm{B},k}$ and $f_{\mathrm{M},k}$ and preprocessing functions u_{B} and $u_{\mathrm{M},k}$ as well as decoding functions d_{B} and $d_{\mathrm{M},k}$ for all $k \in \{1, \ldots, N\}$ such that P_e tends to zero as $T \to \infty$. The sum-rate lower bound is defined by the maximum sum of achievable rates.

B. Known Sum-Rate Bounds

From Shannon's cut-set theorem [12], the sum-rate upper bound, which is known as the *cut-set bound* [19], [21], is given by the maximum sum of the rates that satisfy

$$\sum_{k=1}^{N} R_{\chi,i} \leq \min\Big\{ \log\det\left(\mathbf{I}_{N_{\mathrm{R}}} + \mathbf{H}_{\chi\mathrm{R}}\mathbf{\Omega}_{\chi}{\mathbf{H}_{\chi\mathrm{R}}}^{\mathrm{H}}\right), \log\det\left(\mathbf{I}_{N_{\tilde{\chi}}} + \mathbf{H}_{\mathrm{R}\tilde{\chi}}\mathbf{\Omega}_{\mathrm{R}}{\mathbf{H}_{\mathrm{R}\tilde{\chi}}}^{\mathrm{H}}\right) \Big\}, \quad (10)$$

where $\chi \in \{\mathrm{B}, \mathrm{M}\}$, $\tilde{\chi} = \{\mathrm{B}, \mathrm{M}\} - \chi$, and $\mathbf{\Omega}_{\chi} \triangleq \frac{1}{T} E\left\{\mathbf{S}_{\chi}{\mathbf{S}_{\chi}}^{\mathrm{H}}\right\}$ and $\mathbf{\Omega}_{\mathrm{R}} \triangleq \frac{1}{T} E\left\{\mathbf{S}_{\mathrm{R}}{\mathbf{S}_{\mathrm{R}}}^{\mathrm{H}}\right\}$ are the signal covariance matrices. The power constraints are satisfied from $\mathrm{tr}\left(\mathbf{\Omega}_{\chi}\right) \leq P_{\chi}$ and $\mathrm{tr}\left(\mathbf{\Omega}_{\mathrm{R}}\right) \leq P_{\mathrm{R}}$, where $\mathrm{tr}(\cdot)$ is the trace operation. The optimal signal covariance matrices can be found in [36].

If we denote $\mathcal{N} = \{1, \ldots, N\}$, the sum-rate lower bounds of [8], [24], [35] are bounded by the maximum sum of the rates that satisfy[2]

$$t_{\mathrm{B}} \sum_{k=1}^{N} R_{\mathrm{B},k} + \sum_{m \in \mathcal{M}} R_{\mathrm{M},m} \leq \log\det\left(\mathbf{I}_{N_{\mathrm{R}}} + t_{\mathrm{B}}\mathbf{H}_{\mathrm{BR}}{\mathbf{H}_{\mathrm{BR}}}^{\mathrm{H}} + \sum_{m \in \mathcal{M}} \mathbf{h}_{\mathrm{MR},m}{\mathbf{h}_{\mathrm{MR},m}}^{\mathrm{H}}\right), \quad (11)$$

for all $t_{\mathrm{B}} \in \{0, 1\}, \mathcal{M} \subseteq \mathcal{N}$. The maximum multiplexing gain for the bound in (11) is given by $\min\{2N, N_{\mathrm{R}}\}$, while the maximum multiplexing gain is $2N$ for the cut-set bound in (10) [33].

III. Proposed TWR Scheme for the MU-MIMO Cellular TWRC

A. Phase 1: ZF-based Precoding

For a comprehensive understanding of the proposed precoding scheme, we first consider immediate ZF-based precoding at the BS to make the two effective channel matrices at the RS identical to each other. Because $\mathbf{x}_{\chi,k}$, $k = 1, \ldots, N$, are generated independently from one another without DPC, they should be decoded independently using successive interference cancelation (SIC). We consider the QR-SIC at the RS, which provides optimal performance in the high SNR regime [21], [22], [38] (See Remark 2). The QR decomposition of $\mathbf{H}_{\mathrm{MR}}$ can be represented by

$$\mathbf{H}_{\mathrm{MR}} = \mathbf{\Theta}\tilde{\mathbf{R}}, \quad (12)$$

where $\mathbf{\Theta} \in \mathbb{C}^{N \times N_{\mathrm{R}}}$ consists of orthonormal columns, and $\tilde{\mathbf{R}} \in \mathbb{C}^{N \times N}$ is upper-triangular. The transmit signals are given by

$$\mathbf{S}_{\mathrm{B}} = \left(\mathbf{\Theta}^{\mathrm{H}}\mathbf{H}_{\mathrm{BR}}\right)^{\dagger}\tilde{\mathbf{R}}\mathbf{X}_{\mathrm{B}}, \quad \mathbf{s}_{\mathrm{M},k} = \mathbf{x}_{\mathrm{M},k}, \quad (13)$$

where $\left(\mathbf{\Theta}^{\mathrm{H}}\mathbf{H}_{\mathrm{BR}}\right)^{\dagger}\tilde{\mathbf{R}}$ is the precoding matrix, and $\mathbf{A}^{\dagger} \triangleq \mathbf{A}^{\mathrm{H}}\left(\mathbf{A}\mathbf{A}^{\mathrm{H}}\right)^{-1}$. Here, $\mathbf{\Theta}^{\mathrm{H}}$ is multiplied for this scheme to work for $N_{\mathrm{B}} < N_{\mathrm{R}}$ as well. Inserting (13) to $\mathbf{Y}_{\mathrm{R}}$ in (5) and multiplying $\mathbf{\Theta}^{\mathrm{H}}$ to $\mathbf{Y}_{\mathrm{R}}$, we get

$$\tilde{\mathbf{Y}}_{\mathrm{R}} \triangleq \mathbf{\Theta}^{\mathrm{H}}\mathbf{Y}_{\mathrm{R}} = \tilde{\mathbf{R}}\left(\mathbf{X}_{\mathrm{B}} + \mathbf{X}_{\mathrm{M}}\right) + \tilde{\mathbf{Z}}_{\mathrm{R}}, \quad (14)$$

where $\tilde{\mathbf{Z}}_{\mathrm{R}} \triangleq \mathbf{\Theta}^{\mathrm{H}}\mathbf{Z}_{\mathrm{R}}$. For this channel of (14), it can be easily shown that network coding can be performed in a DF manner using the lattice codes used in [22] and the QR-SIC.

However, this ZF-based precoding unnecessarily penalizes the power allocated to $\mathbf{X}_{\mathrm{B}}$ to satisfy the power constraint, especially if $\mathbf{H}_{\mathrm{BR}}$ is ill-conditioned. In the subsequent section, we propose a general approach for linear precoding and network coding to maximize the achievable sum-rate, in which ZF-based precoding is included as a special case.

B. Phase 1: Sum-Rate Maximizing Linear Precoding

The fundamental challenge for network coding based on DF relaying is to make the network-coded symbol decodable by channel codes. In the binary case, any binary linear code can be used for $\mathbf{x}_{\mathrm{B},k}$ and $\mathbf{x}_{\mathrm{M},k}$, because $\mathbf{x}_{\mathrm{B},k} \oplus \mathbf{x}_{\mathrm{M},k}$ is decodable due to the linearity of the channel codes [36]. For arbitrary cardinalities of $\mathbf{x}_{\mathrm{B},k}$ and $\mathbf{x}_{\mathrm{M},k}$, i.e., arbitrary $R_{\mathrm{B},k}$ and $R_{\mathrm{M},k}$, network coding is generalized from $\mathbf{x}_{\mathrm{B},k} \oplus \mathbf{x}_{\mathrm{M},k}$ to $\left[\mathbf{x}_{\mathrm{B},k} + \mathbf{x}_{\mathrm{M},k}\right]_{\Lambda}$ by lattice codes in the three-node case [19], [21], [23] with the aid of random dithering, where Λ is the lattice used in the construction of the lattice codes. However, we generalize network coding to $\mathbf{x}_{\mathrm{B},k} + \mathbf{x}_{\mathrm{M},k}$, because the random dithering cannot be used in the cTWRC.

1) Encoding at the BS and MSs: The encoding procedure to generate $\mathbf{X}_{\chi} = \left[{\mathbf{x}_{\chi,1}}^{\mathrm{T}}, \ldots, {\mathbf{x}_{\chi,N}}^{\mathrm{T}}\right]^{\mathrm{T}}$ is as follows:

1) To generate complex signals using two independent codewords, the integer message set $\mathcal{W}_{\chi,k}$, $\chi \in \{\mathrm{B}, \mathrm{M}\}$, is divided for every $k \in \{1, \ldots, N\}$ as
$$\mathcal{W}_{\chi,k} = \mathcal{W}_{\chi,k}^{(\mathrm{R})} \times \mathcal{W}_{\chi,k}^{(\mathrm{I})}, \quad (15)$$
where $\mathcal{W}_{\chi,k}^{(\mathrm{R})} = \mathcal{W}_{\chi,k}^{(\mathrm{I})} \triangleq \left\{1, \ldots, 2^{TR_{\chi,k}{}'}\right\}$ and $R_{\chi,k}{}' \triangleq R_{\chi,k}/2$.
2) We construct three lattices $\Lambda_{\mathrm{B},k}, \Lambda_{\mathrm{M},k}, \Lambda_{C,k} \subset \mathbb{R}^{1\times T}$ to develop the lattice codes $\mathcal{C}_{\chi,k}$ which is defined by the elements of $\Lambda_{C,k}$, that are nested in a T-dimensional shaping region $\mathcal{R}\left(\Lambda_{\chi,k}\right)$, where $\mathcal{R}\left(\Lambda\right)$ denotes the Voronoi region of the lattice Λ (See Appendix A-A for lattices preliminaries). The messages $W_{\chi,k}^{(\mathrm{R})} \in \mathcal{W}_{\chi,k}^{(\mathrm{R})}$ and $W_{\chi,k}^{(\mathrm{I})} \in \mathcal{W}_{\chi,k}^{(\mathrm{I})}$ are one-to-one mapped to $\mathbf{x}_{\chi,k}^{(\mathrm{R})} \in \mathcal{C}_{\chi,k}$

[2] Note that (11) is the bound on the achievable rates for Phase 1. The achievable rates for Phase 2 are not discussed, because the scheme for Phase 2 proposed in [8], [24], [35] does not work in the MU-MIMO cTWRC.

and $\mathbf{x}_{\chi,k}^{(\mathrm{I})} \in \mathcal{C}_{\chi,k}$, respectively. For given power constants $\gamma_{\chi,k}$, the lattices $\Lambda_{\mathrm{B},k}, \Lambda_{\mathrm{M},k}$, and $\Lambda_{C,k}$ are jointly constructed such that $|\mathcal{C}_{\chi,k}| = 2^{TR_{\chi,k}}$ and

$$\frac{1}{T}\left\|\mathbf{x}_{\chi,k}^{(\mathrm{R})}\right\|^2 \le \frac{\gamma_{\chi,k}}{2}, \quad \frac{1}{T}\left\|\mathbf{x}_{\chi,k}^{(\mathrm{I})}\right\|^2 \le \frac{\gamma_{\chi,k}}{2}. \tag{16}$$

Assuming $\gamma_{\mathrm{B},k} \ge \gamma_{\mathrm{M},k}$ without loss of generality, we can construct these three lattices such that $\Lambda_{\mathrm{B},k} \subseteq \Lambda_{\mathrm{M},k} \subseteq \Lambda_{C,k}$ and thereby $\mathbf{x}_{\mathrm{B},k}^{(\mathrm{R})} + \mathbf{x}_{\mathrm{M},k}^{(\mathrm{R})}, \mathbf{x}_{\mathrm{B},k}^{(\mathrm{I})} + \mathbf{x}_{\mathrm{M},k}^{(\mathrm{I})} \in \Lambda_{C,k}$. With this group property, $\mathbf{x}_{\mathrm{B},k}^{(\mathrm{R})} + \mathbf{x}_{\mathrm{M},k}^{(\mathrm{R})}$ and $\mathbf{x}_{\mathrm{B},k}^{(\mathrm{I})} + \mathbf{x}_{\mathrm{M},k}^{(\mathrm{I})}$ are *decodable* by lattice decoding [39]. The detailed construction of $\mathcal{C}_{\chi,k}$ is provided in Appendix A-B.

3) The complex-valued codeword $\mathbf{x}_{\chi,k}$ is given from

$$\mathbf{x}_{\chi,k} = \mathbf{x}_{\chi,k}^{(\mathrm{R})} + j\mathbf{x}_{\chi,k}^{(\mathrm{I})}, \quad \chi \in \{\mathrm{B}, \mathrm{M}\}. \tag{17}$$

The precoding matrix at the BS is denoted by $\tilde{\mathbf{W}} \in \mathbb{C}^{N_{\mathrm{B}} \times N}$, which is calculated as a function of $\mathbf{H}_{\mathrm{BR}}$ and $\mathbf{H}_{\mathrm{MR}}$ at the RS. We assume that $\tilde{\mathbf{W}}$ is known by the BS through error-free feedback as assumed in [2], [8]–[11], [16], [21]–[24], [28]. The compact SVD of $\mathbf{H}_{\mathrm{BR}}$ can be denoted as

$$\mathbf{H}_{\mathrm{BR}} = \mathbf{U}_{\mathrm{BR}}\mathbf{\Sigma}_{\mathrm{BR}}{\mathbf{V}_{\mathrm{BR}}}^{\mathrm{H}}, \tag{18}$$

where $\mathbf{U}_{\mathrm{BR}} \in \mathbb{C}^{N_{\mathrm{R}} \times N}$ and $\mathbf{V}_{\mathrm{BR}} \in \mathbb{C}^{N_{\mathrm{B}} \times N}$ consist of N orthogonal columns and $\mathbf{\Sigma}_{\mathrm{BR}}$ is an $(N \times N)$-dimensional diagonal matrix. The transmit signals are obtained from

$$\mathbf{S}_{\mathrm{B}} = \tilde{\mathbf{W}}\mathbf{X}_{\mathrm{B}} = \mathbf{V}_{\mathrm{BR}}\mathbf{W}\mathbf{X}_{\mathrm{B}}, \quad \mathbf{s}_{\mathrm{M},k} = \mathbf{x}_{\mathrm{M},k}, \tag{19}$$

where $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_N] \in \mathbb{C}^{N \times N}$ is to be optimized in the sequel. Here, $\mathbf{V}_{\mathrm{BR}}$ is to simplify the optimization of the precoding matrix without loss of optimality, that is, we find $\mathbf{W}$ instead of $\tilde{\mathbf{W}}$. The power constraints are given by

$$\begin{aligned}\frac{1}{T}E\left\{\|\mathbf{V}_{\mathrm{BR}}\mathbf{W}\mathbf{X}_{\mathrm{B}}\|_{\mathrm{F}}^2\right\} &= \frac{1}{T}\sum_{k=1}^{N} E\|\mathbf{x}_{\mathrm{B},k}\|^2\|\mathbf{w}_k\|^2 \\ &\le \sum_{k=1}^{N}\gamma_{\mathrm{B},k}\|\mathbf{w}_k\|^2 \le P_{\mathrm{B}},\end{aligned} \tag{20}$$

$$\frac{1}{T}E\left\{\|\mathbf{x}_{\mathrm{M},k}\|^2\right\} \le \gamma_{\mathrm{M},k} \le P_{\mathrm{M},k}. \tag{21}$$

2) Decoding at the RS: If we denote the permutation order in Phase 1 as $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_N\}$, the signal received at the RS can be written by

$$\begin{aligned}\mathbf{Y}_{\mathrm{R}} &= \mathbf{H}_{\mathrm{BR}}\mathbf{V}_{\mathrm{BR}}\mathbf{W}\mathbf{X}_{\mathrm{B}} + \mathbf{H}_{\mathrm{MR}}\mathbf{X}_{\mathrm{M}} + \mathbf{Z}_{\mathrm{R}} \\ &= \mathbf{H}_{\mathrm{MR}}(\mathbf{X}_{\mathrm{B}} + \mathbf{X}_{\mathrm{M}}) + (\mathbf{H}_{\mathrm{BR}}\mathbf{V}_{\mathrm{BR}}\mathbf{W} - \mathbf{H}_{\mathrm{MR}})\mathbf{X}_{\mathrm{B}} + \mathbf{Z}_{\mathrm{R}} \\ &= \mathbf{H}_{\mathrm{MR}}{\mathbf{\Pi}_{\boldsymbol{\theta}}}^{\mathrm{T}}(\mathbf{X}_{\mathrm{B}}^{[\boldsymbol{\theta}]} + \mathbf{X}_{\mathrm{M}}^{[\boldsymbol{\theta}]}) \\ &\quad + (\mathbf{H}_{\mathrm{BR}}\mathbf{V}_{\mathrm{BR}}\mathbf{W}{\mathbf{\Pi}_{\boldsymbol{\theta}}}^{\mathrm{T}} - \mathbf{H}_{\mathrm{MR}}{\mathbf{\Pi}_{\boldsymbol{\theta}}}^{\mathrm{T}})\mathbf{X}_{\mathrm{B}}^{[\boldsymbol{\theta}]} + \mathbf{Z}_{\mathrm{R}},\end{aligned} \tag{22}$$

where $\mathbf{X}_{\chi}^{[\boldsymbol{\theta}]} \triangleq \mathbf{\Pi}_{\boldsymbol{\theta}}\mathbf{X}_{\chi}$. The compact QR-decomposition for ${\mathbf{H}_{\mathrm{MR}}\mathbf{\Pi}_{\boldsymbol{\theta}}}^{\mathrm{T}}$ can be expressed as

$${\mathbf{H}_{\mathrm{MR}}\mathbf{\Pi}_{\boldsymbol{\theta}}}^{\mathrm{T}} = \mathbf{Q}\mathbf{R}, \tag{23}$$

where $\mathbf{Q} \in \mathbb{C}^{N_{\mathrm{R}} \times N}$ consists of N orthogonal columns and where $\mathbf{R} \in \mathbb{C}^{N \times N}$ is an upper-triangular matrix, the (n,m)th element of which is denoted by $r_{n,m}$. Multiplying $\mathbf{Q}^{\mathrm{H}}$ to $\mathbf{Y}_{\mathrm{R}}$ and normalizing the diagonal channel coefficients, we get

$$\begin{aligned}\mathbf{Y}_{\mathrm{R}}' &= \left[{\mathbf{y}_{\mathrm{R},1}'}^{\mathrm{T}}, \ldots, {\mathbf{y}_{\mathrm{R},N}'}^{\mathrm{T}}\right]^{\mathrm{T}} \triangleq {\mathbf{R}_{\mathrm{d}}}^{-1}\mathbf{Q}^{\mathrm{H}}\mathbf{Y}_{\mathrm{R}} \\ &= \left(\mathbf{X}_{\mathrm{B}}^{[\boldsymbol{\theta}]} + \mathbf{X}_{\mathrm{M}}^{[\boldsymbol{\theta}]}\right) + \underbrace{\mathbf{R}'\left(\mathbf{X}_{\mathrm{B}}^{[\boldsymbol{\theta}]} + \mathbf{X}_{\mathrm{M}}^{[\boldsymbol{\theta}]}\right)}_{\text{cancelled by SIC}} \\ &\quad + \underbrace{{\mathbf{R}_{\mathrm{d}}}^{-1}\mathbf{G}\mathbf{X}_{\mathrm{B}}^{[\boldsymbol{\theta}]}}_{\text{ISI}} + \mathbf{Z}_{\mathrm{R}}',\end{aligned} \tag{24}$$

where $\mathbf{R}_{\mathrm{d}} \triangleq \mathrm{diag}(r_{1,1}, \ldots, r_{N,N})$, $\mathbf{Z}_{\mathrm{R}}' = \left[{\mathbf{z}_{\mathrm{R},1}'}^{\mathrm{T}}, \ldots, {\mathbf{z}_{\mathrm{R},N}'}^{\mathrm{T}}\right]^{\mathrm{T}} \triangleq {\mathbf{R}_{\mathrm{d}}}^{-1}\mathbf{Q}^{\mathrm{H}}\mathbf{Z}_{\mathrm{R}}$, $\mathbf{R}' \triangleq {\mathbf{R}_{\mathrm{d}}}^{-1}\mathbf{R} - \mathbf{I}_N$, and

$$\mathbf{G} \triangleq \mathbf{Q}^{\mathrm{H}}\mathbf{H}_{\mathrm{BR}}\mathbf{V}_{\mathrm{BR}}\mathbf{W}{\mathbf{\Pi}_{\boldsymbol{\theta}}}^{\mathrm{T}} - \mathbf{R}. \tag{25}$$

Because $\mathbf{R}'$ is strictly upper-triangular, the term $\mathbf{R}'\left(\mathbf{X}_{\mathrm{B}}^{[\boldsymbol{\theta}]} + \mathbf{X}_{\mathrm{M}}^{[\boldsymbol{\theta}]}\right)$ will be canceled during the SIC decoding process. Note that if we choose $\mathbf{W}$ such that $\mathbf{W} = \left(\mathbf{Q}^{\mathrm{H}}\mathbf{H}_{\mathrm{BR}}\mathbf{V}_{\mathrm{BR}}\right)^{-1}\mathbf{R}\mathbf{\Pi}_{\boldsymbol{\theta}}$, the scheme becomes equivalent to ZF-based precoding described in Section III-A. We utilize the ISI of ${\mathbf{R}_{\mathrm{d}}}^{-1}\mathbf{G}\mathbf{X}_{\mathrm{B}}^{[\boldsymbol{\theta}]}$ to increase SINR. To this end, we need $\mathrm{diag}(\mathbf{G}) = \mathbf{0}$ such that the ISI serves as noise independent of the signal term $\mathbf{X}_{\mathrm{B}}^{[\boldsymbol{\theta}]} + \mathbf{X}_{\mathrm{M}}^{[\boldsymbol{\theta}]}$.

Note that $\mathbf{x}_{\mathrm{B},k} + \mathbf{x}_{\mathrm{M},k}$ is contained in the i_kth spatial stream due to the permutation, where i_k is the index that satisfies $\theta_{i_k} = k$ for given k. From (24), the i_kth spatial stream is given by

$$\begin{aligned}\mathbf{y}_{\mathrm{R},i_k}' &= \mathbf{x}_{\mathrm{B},k} + \mathbf{x}_{\mathrm{M},k} + \underbrace{\sum_{n=i_k+1}^{N}\frac{r_{i_k,n}}{r_{i_k,i_k}}\left(\mathbf{x}_{\mathrm{B},\theta_n} + \mathbf{x}_{\mathrm{M},\theta_n}\right)}_{\text{cancelled by SIC}} \\ &\quad + \underbrace{\sum_{m=1}^{N}\frac{g_{i_k,m}}{r_{i_k,i_k}}\mathbf{x}_{\mathrm{B},\theta_m}}_{\text{ISI}} + \mathbf{z}_{\mathrm{R},i_k}',\end{aligned} \tag{26}$$

where $g_{n,m}$ is the (n,m)th element of $\mathbf{G}$.

Now suppose that $\mathbf{x}_{\mathrm{B},\theta_n} + \mathbf{x}_{\mathrm{M},\theta_n}$, $n = i_k + 1, \ldots, N$, have been perfectly retrieved, i.e., $\hat{\mathbf{x}}_{\mathrm{add},\theta_n}^{(\mathrm{R})} = \mathbf{x}_{\mathrm{add},\theta_n}^{(\mathrm{R})}$ and $\hat{\mathbf{x}}_{\mathrm{add},\theta_n}^{(\mathrm{I})} = \mathbf{x}_{\mathrm{add},\theta_n}^{(\mathrm{I})}$, where $\hat{\mathbf{x}}_{\mathrm{add},\theta_n}^{(\mathrm{R})}$ and $\hat{\mathbf{x}}_{\mathrm{add},\theta_n}^{(\mathrm{I})}$ are the estimates of $\mathbf{x}_{\mathrm{add},\theta_n}^{(\mathrm{R})} \triangleq \mathbf{x}_{\mathrm{B},\theta_n}^{(\mathrm{R})} + \mathbf{x}_{\mathrm{M},\theta_n}^{(\mathrm{R})}$ and $\mathbf{x}_{\mathrm{add},\theta_n}^{(\mathrm{I})} \triangleq \mathbf{x}_{\mathrm{B},\theta_n}^{(\mathrm{I})} + \mathbf{x}_{\mathrm{M},\theta_n}^{(\mathrm{I})}$, respectively. The decoder input for $\mathbf{x}_{\mathrm{B},k}^{(\mathrm{R})} + \mathbf{x}_{\mathrm{M},k}^{(\mathrm{R})}$ is then given by

$$\begin{aligned}\bar{\mathbf{y}}_{\mathrm{R},i_k}^{(\mathrm{R})} &= \Re\left(\mathbf{y}_{\mathrm{R},i_k}' - \sum_{n=i_k+1}^{N}\frac{r_{i_k,n}}{r_{i_k,i_k}}\left(\hat{\mathbf{x}}_{\mathrm{add},\theta_n}^{(\mathrm{R})} + j\hat{\mathbf{x}}_{\mathrm{add},\theta_n}^{(\mathrm{I})}\right)\right) \\ &= \mathbf{x}_{\mathrm{B},k}^{(\mathrm{R})} + \mathbf{x}_{\mathrm{M},k}^{(\mathrm{R})} + \Re\left(\sum_{m=1}^{N}\frac{g_{i_k,m}}{r_{i_k,i_k}}\mathbf{x}_{\mathrm{B},\theta_m} + \mathbf{z}_{\mathrm{R},i_k}'\right) \\ &= \mathbf{x}_{\mathrm{B},k}^{(\mathrm{R})} + \mathbf{x}_{\mathrm{M},k}^{(\mathrm{R})} + \Re\left(\boldsymbol{\varepsilon}_k\right),\end{aligned} \tag{27}$$

where $\boldsymbol{\varepsilon}_k \triangleq \sum_{m=1}^{N}\frac{g_{i_k,m}}{r_{i_k,i_k}}\mathbf{x}_{\mathrm{B},\theta_m} + \mathbf{z}_{\mathrm{R},i_k}'$ is the effective noise. We get the decoder input for $\mathbf{x}_{\mathrm{B},k}^{(\mathrm{I})} + \mathbf{x}_{\mathrm{M},k}^{(\mathrm{I})}$ analogously with the effective noise $\Im\left(\boldsymbol{\varepsilon}_k\right)$.

The goal is now to retrieve $\mathbf{x}_{\mathrm{B},k}^{(\mathrm{R})} + \mathbf{x}_{\mathrm{M},k}^{(\mathrm{R})}$ with error probability vanishing as $T \to \infty$. We consider the *ambiguity decoder* [39], which exhibits optimal performance for lattice codes

without random dithering. Roughly, if $x \in \mathcal{C}$ is the channel input with the lattice code $\mathcal{C}$ and $y = x + z$ is the channel output with the additive noise z, the ambiguity decoder finds the unique codeword $\hat{x} \in \mathcal{C}$ if $y \in \hat{x}+\mathcal{E}$, where $\mathcal{E}$ is a decision region, or declares the ambiguity event if there exist $\hat{x}, \tilde{x} \in \mathcal{C}$, $\hat{x} \neq \tilde{x}$, such that $y \in (\hat{x}+\mathcal{E}) \cap (\tilde{x}+\mathcal{E})$. The main task in the development of the ambiguity decoder is the determination of the decision region and the calculation of the error probability that is defined by the incorrect estimate of the channel input and the ambiguity event.

In the channel model with the conventional lattice codes, the ISI can be rendered to be uniformly distributed over a bounded region with the aid of random dithering [17]–[19]. It was shown that this ISI becomes Gaussian as $T \to \infty$ by properly constructing the lattice codes and that the rate of $[\log(\mathrm{SINR})]^{+}$[3] is achievable. However, in our channel model of (27), the ISI of $\sum_{m=1}^{N} \frac{g_{i_k,m}}{r_{i_k,i_k}} \mathbf{x}_{\mathrm{B},\theta_m}$ for given $g_{i_k,m}$ and r_{i_k,i_k} is not Gaussian nor uniformly distributed over a bounded region, because $\mathbf{x}_{\mathrm{B},\theta_m} \in \mathcal{C}_{\mathrm{B},\theta_m}$ takes a discrete value. Moreover, the conventional lattice theory cannot be trivially modified to calculate the achievable rate of this channel model. We show in the following theorem that the ambiguity decoder can achieve the rate of $[\log(\mathrm{SINR})]^{+}$ even without random dithering.

Theorem 1: For the interference channel (27), the probability error under the use of ambiguity decoding for $\mathbf{x}_{\mathrm{B},k}^{(\mathrm{R})} + \mathbf{x}_{\mathrm{M},k}^{(\mathrm{R})}$ vanishes as $T \to \infty$, if the rate pair $(R_{\mathrm{B},k}, R_{\mathrm{M},k})$ satisfies

$$R_{\mathrm{B},k} \leq \left[\log \frac{\gamma_{\mathrm{B},k}|r_{i_k,i_k}|^2}{\sum_{m=1}^{N}|g_{i_k,m}|^2\gamma_{\mathrm{B},\theta_m}+1}\right]^{+} \triangleq R_{\mathrm{BR},k}^{*},$$
$$R_{\mathrm{M},k} \leq \left[\log \frac{\gamma_{\mathrm{M},k}|r_{i_k,i_k}|^2}{\sum_{m=1}^{N}|g_{i_k,m}|^2\gamma_{\mathrm{B},\theta_m}+1}\right]^{+} \triangleq R_{\mathrm{MR},k}^{*}, \quad (28)$$

where i_k is obtained from $\theta_{i_k} = k$ for given k. Therefore, the rate region of (28) is achievable. The same argument holds true for decoding of $\mathbf{x}_{\mathrm{B},k}^{(\mathrm{I})} + \mathbf{x}_{\mathrm{M},k}^{(\mathrm{I})}$ from $\mathbf{x}_{\mathrm{B},k}^{(\mathrm{I})} + \mathbf{x}_{\mathrm{M},k}^{(\mathrm{I})} + \Im(\boldsymbol{\varepsilon}_k)$.

Proof: See Appendix B. ∎

Note that $R_{\mathrm{BR},k}^{*}$ and $R_{\mathrm{MR},k}^{*}$ of (28) are in the form of $\log(\mathrm{SINR})$, because it can be easily seen that the variances of $\boldsymbol{\varepsilon}_{\boldsymbol{k}}$ and $\mathbf{x}_{\chi,k}$ per dimension are bounded by $(\sum_{m=1}^{N}|g_{i_k,m}|^2\gamma_{\mathrm{B},\theta_m}+1)/|r_{i_k,i_k}|^2$ and $\gamma_{\chi,k}$, respectively (See (66) in Appendix B). In fact, Theorem 1 is valid for any lattice additive interference channel with bounded interference.

C. Phase 2: One-Sided DPC

1) Encoding at the RS: In Phase 2, the retrieved added codewords $\hat{\mathbf{x}}_{\mathrm{add},k}^{(\mathrm{R})}$ and $\hat{\mathbf{x}}_{\mathrm{add},k}^{(\mathrm{I})}$, $k = 1, \ldots, N$, are one-to-one mapped to the codewords of another lattice codes and then broadcasted to the BS and MSs. Unlike in Phase 1, the random dithering can be used by applying DPC at the RS. Specifically, for each MS to receive an interference-free signal, ZF-DPC is used for the channel $\mathbf{H}_{\mathrm{RM}}$.

Let us denote the cardinality of $\mathbf{x}_{\mathrm{add},k}^{(\mathrm{R})}$ and $\mathbf{x}_{\mathrm{add},k}^{(\mathrm{I})}$ as $2^{TR_{\mathrm{add},k}}$ using the rate $R_{\mathrm{add},k}$ ($\leq R_{\mathrm{B},k}{}' + R_{\mathrm{M},k}{}'$). Similarly to Phase 1, we develop a T-dimensional lattice code $\mathcal{C}_{\mathrm{R},k}$ by constructing two lattices $\Lambda_{D,k}$ and $\Lambda_{\mathrm{R},k}$ that satisfy $\Lambda_{\mathrm{R},k} \subseteq \Lambda_{D,k}$. The lattice code $\mathcal{C}_{\mathrm{R},k}$ is defined by the lattice points of $\Lambda_{D,k}$, nested in $\mathcal{R}(\Lambda_{\mathrm{R},k})$. The lattices $\Lambda_{D,k}$ and $\Lambda_{\mathrm{R},k}$ are jointly constructed such that $|\mathcal{C}_{\mathrm{R},k}| = 2^{TR_{\mathrm{add},k}}$. Because continuous signaling is used with the aid of random dithering, the main difference compared to the code construction in Phase 1 is that we develop the two lattices such that $\frac{1}{T}E\|\mathbf{x}\|^2 = \gamma_{\mathrm{R},k}/2$ for $\mathbf{x} \sim \mathrm{Unif}(\mathcal{R}(\Lambda_{\mathrm{R},k}))$, where $\gamma_{\mathrm{R},k}$ is the power constant at the RS (cf. (16)). Each pair of the retrieved codewords $\hat{\mathbf{x}}_{\mathrm{add},k}^{(\mathrm{R})}$ and $\hat{\mathbf{x}}_{\mathrm{add},k}^{(\mathrm{I})}$, $k = 1, \ldots, N$, is one-to-one mapped to $\mathbf{x}_{\mathrm{R},k}^{(\mathrm{R})} \in \mathcal{C}_{\mathrm{R},k}$ and $\mathbf{x}_{\mathrm{R},k}^{(\mathrm{I})} \in \mathcal{C}_{\mathrm{R},k}$, respectively.

Now, we consider DPC encoding for the given permutation order, denoted by $\boldsymbol{\phi} = [\phi_1, \ldots, \phi_N]$. The transposed QR-decomposition of $\boldsymbol{\Pi}_{\boldsymbol{\phi}}\mathbf{H}_{\mathrm{RM}}$ can be written by

$$\boldsymbol{\Pi}_{\boldsymbol{\phi}}\mathbf{H}_{\mathrm{RM}} = \mathbf{L}\boldsymbol{\Omega}, \quad (29)$$

where $\boldsymbol{\Omega} \in \mathbb{C}^{N \times N_{\mathrm{R}}}$ consists of orthogonal columns and where $\mathbf{L} \in \mathbb{C}^{N \times N}$ is lower-triangular, the (n,m)th element of which is denoted by $l_{n,m}$. If we denote the codeword matrix encoded by DPC as $\mathbf{X}_{\mathrm{DPC}} = \left[\mathbf{x}_{\mathrm{DPC},1}{}^{\mathrm{T}}, \ldots, \mathbf{x}_{\mathrm{DPC},N}{}^{\mathrm{T}}\right]^{\mathrm{T}}$, the transmit signal at the RS is given by

$$\mathbf{S}_{\mathrm{R}} = \boldsymbol{\Omega}^{\mathrm{H}}\mathbf{X}_{\mathrm{DPC}}. \quad (30)$$

The row-permutated received signal at the MSs is obtained by

$$\mathbf{Y}_{\mathrm{M}}^{[\boldsymbol{\phi}]} = \left[\mathbf{y}_{\mathrm{M},1}^{[\boldsymbol{\phi}]}{}^{\mathrm{T}}, \ldots, \mathbf{y}_{\mathrm{M},N}^{[\boldsymbol{\phi}]}{}^{\mathrm{T}}\right]^{\mathrm{T}} \triangleq \boldsymbol{\Pi}_{\boldsymbol{\phi}}\mathbf{Y}_{\mathrm{M}}$$
$$= \boldsymbol{\Pi}_{\boldsymbol{\phi}}\mathbf{H}_{\mathrm{RM}}\mathbf{S}_{\mathrm{R}} + \boldsymbol{\Pi}_{\boldsymbol{\phi}}\mathbf{Z}_{\mathrm{M}} = \mathbf{L}\mathbf{X}_{\mathrm{DPC}} + \boldsymbol{\Pi}_{\boldsymbol{\phi}}\mathbf{Z}_{\mathrm{M}}, \quad (31)$$

$$\mathbf{y}_{\mathrm{M},k} = \mathbf{y}_{\mathrm{M},q_k}^{[\boldsymbol{\phi}]}$$
$$= l_{q_k,q_k}\mathbf{x}_{\mathrm{DPC},q_k} + \underbrace{\sum_{n=1}^{q_k-1} l_{q_k,n}\mathbf{x}_{\mathrm{DPC},n}}_{\text{cancelled by DPC}} + \mathbf{z}_{\mathrm{M},k}, \quad (32)$$

where $\mathbf{Z}_{\mathrm{M}} \triangleq \left[\mathbf{z}_{\mathrm{M},1}{}^{\mathrm{T}}, \ldots, \mathbf{z}_{\mathrm{M},N}{}^{\mathrm{T}}\right]^{\mathrm{T}}$, and q_k is the index that satisfies $\phi_{q_k} = k$ for given k. Therefore, $\mathbf{x}_{\mathrm{DPC},k}$ should be obtained in ascending order of k, where $\mathbf{x}_{\mathrm{DPC},q_k}$ is constructed as

$$\mathbf{x}_{\mathrm{DPC},q_k} = \mathbf{x}_{\mathrm{DPC},q_k}^{(\mathrm{R})} + j\mathbf{x}_{\mathrm{DPC},q_k}^{(\mathrm{I})}, \quad (33)$$

$$\mathbf{x}_{\mathrm{DPC},q_k}^{(\mathrm{R})} \triangleq \left[\mathbf{x}_{\mathrm{R},k}^{(\mathrm{R})} - \Re\left(\alpha_k \sum_{n=1}^{q_k-1} \frac{l_{q_k,n}}{l_{q_k,q_k}}\mathbf{x}_{\mathrm{DPC},n}\right) - \mathbf{u}_{\mathrm{R},k}^{(\mathrm{R})}\right]_{\Lambda_{\mathrm{R},k}}, \quad (34)$$

$$\mathbf{x}_{\mathrm{DPC},q_k}^{(\mathrm{I})} \triangleq \left[\mathbf{x}_{\mathrm{R},k}^{(\mathrm{I})} - \Im\left(\alpha_k \sum_{n=1}^{q_k-1} \frac{l_{q_k,n}}{l_{q_k,q_k}}\mathbf{x}_{\mathrm{DPC},n}\right) - \mathbf{u}_{\mathrm{R},k}^{(\mathrm{I})}\right]_{\Lambda_{\mathrm{R},k}}. \quad (35)$$

Here, $\mathbf{u}_{\mathrm{R},k}^{(\mathrm{R})}, \mathbf{u}_{\mathrm{R},k}^{(\mathrm{I})} \in \mathbb{R}^{1\times T}$ are the random dithers distributed uniformly over $\mathcal{R}(\Lambda_{\mathrm{R},k})$, and $\alpha_k = \gamma_{\mathrm{R},k}/(\gamma_{\mathrm{R},k} + |l_{q_k,q_k}|^{-2})$ is the minimum mean square error (MMSE) coefficient [17] for decoding at the kth MS. From [17, Lemma 1], $\mathbf{x}_{\mathrm{DPC},q_k}^{(\mathrm{R})}$ is uniformly distributed over $\mathcal{R}(\Lambda_{\mathrm{R},k})$ due to the random dithering.

[3] In [17], [18], the ISI is chosen such that SINR $= 1 +$ SNR. For the SISO TWRC [19], the maximum SINR is given by SINR $= \alpha +$ SNR, where $0 < \alpha < 1$ is a constant value determined by power constants.

2) Decoding at the MSs: Decoding of $\mathbf{x}_{\mathrm{R},k}^{(\mathrm{R})}$ or $\mathbf{x}_{\mathrm{R},k}^{(\mathrm{I})}$ at the kth MS is analogous to that of the traditional DPC scheme [40], i.e., the decoder input for $\mathbf{x}_{\mathrm{R},k}^{(\mathrm{R})}$ is given by $\left[\Re\left(\alpha_k \frac{\mathbf{y}_{\mathrm{M},k}}{l_{q_k,q_k}}\right)+\mathbf{u}_{\mathrm{R},k}^{(\mathrm{R})}\right]_{\Lambda_{\mathrm{R},k}}$. We omit the detailed derivation of decoding at the MSs to avoid duplication.

It is important to note that, while the cardinality of $\mathcal{C}_{\mathrm{R},k}$ is $2^{TR_{\mathrm{add},k}}$, the cardinality of the codebook which the kth MS should search to decode $\mathbf{x}_{\mathrm{R},k}^{(\mathrm{R})}$ is $2^{TR_{\mathrm{B},k}{}'}$, as $\mathbf{x}_{\mathrm{M},k}$ is used as side-information. Therefore, it is easy to show that the achievable sum-rate at the kth MS is given in terms of the condition for $R_{\mathrm{B},k}$ instead of $R_{\mathrm{add},k}$, as follows:

$$R_{\mathrm{B},k} \le \log\left(1+{l_{q_k,q_k}}^2\gamma_{\mathrm{R},k}\right) \triangleq R_{\mathrm{RM},k}^{*}. \tag{36}$$

3) Decoding at the BS: Inserting (30) into (7), the received signal matrix at the BS can be rewritten by

$$\mathbf{Y}_{\mathrm{B}} \;=\; \mathbf{H}_{\mathrm{RB}}\mathbf{\Omega}^{\mathrm{H}}\mathbf{X}_{\mathrm{DPC}}+\mathbf{Z}_{\mathrm{B}}. \tag{37}$$

The QL-decomposition of $\mathbf{H}_{\mathrm{RB}}\mathbf{\Omega}^{\mathrm{H}}$ can be denoted as

$$\mathbf{H}_{\mathrm{RB}}\mathbf{\Omega}^{\mathrm{H}}=\tilde{\mathbf{Q}}\tilde{\mathbf{L}}, \tag{38}$$

where $\tilde{\mathbf{Q}} \in \mathbb{C}^{N_{\mathrm{B}}\times N}$ consists of N orthogonal columns and where $\tilde{\mathbf{L}} \in \mathbb{C}^{N\times N}$ is lower-triangular, the (n,m)th element of which is denoted as $\tilde{l}_{n,m}$. Equalizing $\tilde{\mathbf{Q}}$ from $\mathbf{Y}_{\mathrm{B}}$ gives

$$\mathbf{Y}_{\mathrm{B}}'=\left[{\mathbf{y}_{\mathrm{B},1}'}^{\mathrm{T}},\ldots,{\mathbf{y}_{\mathrm{B},N}'}^{\mathrm{T}}\right]^{\mathrm{T}} \triangleq \tilde{\mathbf{Q}}^{\mathrm{H}}\mathbf{Y}_{\mathrm{B}}=\tilde{\mathbf{L}}\mathbf{X}_{\mathrm{DPC}}+\mathbf{Z}_{\mathrm{B}}', \tag{39}$$

$$\mathbf{y}_{\mathrm{B},q_k}'=\tilde{l}_{q_k,q_k}\mathbf{x}_{\mathrm{DPC},q_k}+\sum_{n=1}^{q_k-1}\tilde{l}_{q_k,n}\mathbf{x}_{\mathrm{DPC},n}+\mathbf{z}_{\mathrm{B},q_k}', \tag{40}$$

where $\mathbf{Z}_{\mathrm{B}}'=\left[{\mathbf{z}_{\mathrm{B},1}'}^{\mathrm{T}},\ldots,{\mathbf{z}_{\mathrm{B},N}'}^{\mathrm{T}}\right]^{\mathrm{T}} \triangleq \tilde{\mathbf{Q}}^{\mathrm{H}}\mathbf{Z}_{\mathrm{B}}$. From (40), we get the decoder input for $\mathbf{x}_{\mathrm{R},k}^{(\mathrm{R})}$ as

$$\begin{aligned}&\Bigg[\Re\Bigg(\tilde{\alpha}_k\Bigg(\frac{\mathbf{y}_{\mathrm{B},q_k}'}{\tilde{l}_{q_k,q_k}}-\underbrace{\sum_{n=1}^{q_k-1}\frac{\tilde{l}_{q_k,n}}{\tilde{l}_{q_k,q_k}}\mathbf{x}_{\mathrm{DPC},n}}_{\text{SIC}}\Bigg)\Bigg)\\&+\underbrace{\Re\left(\alpha_k\sum_{n=1}^{q_k-1}\frac{l_{q_k,n}}{l_{q_k,q_k}}\mathbf{x}_{\mathrm{DPC},n}\right)}_{\text{DPC cancellation}}+\mathbf{u}_{\mathrm{R},k}^{(\mathrm{R})}+\mathbf{x}_{\mathrm{DPC},q_k}^{(\mathrm{R})}-\mathbf{x}_{\mathrm{DPC},q_k}^{(\mathrm{R})}\Bigg]_{\Lambda_{\mathrm{R},k}}\\&=\Big[\mathbf{x}_{\mathrm{R},k}^{(\mathrm{R})}-\underbrace{(1-\tilde{\alpha}_k)\mathbf{x}_{\mathrm{DPC},q_k}^{(\mathrm{R})}}_{\text{ISI}}+\Re\big(\tilde{\alpha}_k\mathbf{z}_{\mathrm{B},q_k}'/\tilde{l}_{q_k,q_k}\big)\Big]_{\Lambda_{\mathrm{R},k}},\end{aligned} \tag{41}$$

where $\tilde{\alpha}_k=\gamma_{\mathrm{R},k}/(\gamma_{\mathrm{R},k}+|\tilde{l}_{q_k,q_k}|^{-2})$ is the MMSE coefficient at the BS. The subtraction of $\sum_{n=1}^{q_k-1}\frac{\tilde{l}_{q_k,n}}{\tilde{l}_{q_k,q_k}}\mathbf{x}_{\mathrm{DPC},n}$ is intended to cancel the retrieved codewords in an SIC manner. This term can be modified for any other decoding order at the BS. However, decoding at the BS must be performed in order from $\mathbf{x}_{\mathrm{DPC},1}$ to $\mathbf{x}_{\mathrm{DPC},N}$, because the information of $\mathbf{x}_{\mathrm{DPC},n}$, $n=1,\ldots,q_k$, is required to cancel the term $\Re\left(\alpha_k\sum_{n=1}^{q_k-1}\frac{l_{q_k,n}}{l_{q_k,q_k}}\mathbf{x}_{\mathrm{DPC},n}\right)$ which results from DPC encoding for the kth MS (See (34)). Due to this restriction, the rate lower bound for Phase 2 of [8], [24], [35], which allows the independent optimization of ϕ and the decoding order at the BS, cannot be achieved in general.

Note that $(1-\tilde{\alpha}_k)\mathbf{x}_{\mathrm{DPC},q_k}^{(\mathrm{R})}$ serves as ISI, because $\mathbf{x}_{\mathrm{DPC},q_k}^{(\mathrm{R})}$ is statistically independent of $\mathbf{x}_{\mathrm{R},k}^{(\mathrm{R})}$ due to the random dithering [17, Lemma 1]. Using this ISI and following the decoding order that corresponds to ϕ, the BS can achieve the following rate (See [17]):

$$R_{\mathrm{M},k} \le \log\left(1+|\tilde{l}_{q_k,q_k}|^2\gamma_{\mathrm{R},k}\right) \triangleq R_{\mathrm{RB},k}^{*}, \tag{42}$$

where q_k is obtained from $\phi_{q_k}=k$ for given k.

From (28), (36), and (42), we get the new lower bound on the sum-rate R_{sum} as follows:

$$\begin{aligned}R_{\mathrm{sum}} &\le \sum_{k=1}^{N}\left(\min\left\{R_{\mathrm{BR},k}^{*},R_{\mathrm{RM},k}^{*}\right\}+\min\left\{R_{\mathrm{MR},k}^{*},R_{\mathrm{RB},k}^{*}\right\}\right)\\&\triangleq R_{\mathrm{sum}}^{*}\end{aligned} \tag{43}$$

The overall block diagram of the proposed scheme is depicted in Fig. 2.

Remark 1: It can be easily shown [33] that the sum-rate lower bound R_{sum}^{*} achieves the full multiplexing gain of $2N$ regardless of N_{B} or N_{R}, as in the cut-set bound.

D. Optimal Precoder and Power Constraints

We now provide the optimization problem for the precoding matrix $\mathbf{W}$ and power constants $\gamma_{\mathrm{B},k}$, $\gamma_{\mathrm{M},k}$, and $\gamma_{\mathrm{R},k}$ that maximize the sum-rate. Because $\mathbf{W}$ can be uniquely determined from (25) as

$$\mathbf{W}=\left(\mathbf{Q}^{\mathrm{H}}\mathbf{H}_{\mathrm{BR}}\mathbf{V}_{\mathrm{BR}}\right)^{-1}(\mathbf{G}+\mathbf{R})\,\mathbf{\Pi}_{\theta}, \tag{44}$$

we find the optimal $\mathbf{G}$ instead of $\mathbf{W}$ without loss of generality. In addition, to make the problem remain in the real-domain, the complex-valued matrix $\mathbf{G}$ is decomposed as

$$\mathbf{G}=\mathbf{G}^{(\mathrm{R})}+j\mathbf{G}^{(\mathrm{I})}, \tag{45}$$

where $\mathbf{G}^{(\mathrm{R})},\mathbf{G}^{(\mathrm{I})} \in \mathbb{R}^{N\times N}$. The (n,m)th elements of $\mathbf{G}^{(\mathrm{R})}$ and $\mathbf{G}^{(\mathrm{I})}$ are denoted as $g_{n,m}^{(\mathrm{R})}$ and $g_{n,m}^{(\mathrm{I})}$, respectively. The optimization problem to maximize the achievable sum-rate can be formulated as (46) at the bottom of the next page, where $\langle\mathbf{A}\rangle_k$ denotes the kth column of the matrix $\mathbf{A}$ and

$$\rho_{\mathrm{BR},k} \triangleq \frac{\gamma_{\mathrm{B},k}|r_{i_k,i_k}|^2}{\sum_{m=1}^{N}\left({g_{i_k,m}^{(\mathrm{R})}}^2+{g_{i_k,m}^{(\mathrm{I})}}^2\right)\gamma_{\mathrm{B},\theta_m}+1}, \tag{47}$$

$$\rho_{\mathrm{RM},k} \triangleq |l_{q_k,q_k}|^2\gamma_{\mathrm{R},k}, \tag{48}$$

$$\rho_{\mathrm{MR},k} \triangleq \frac{\gamma_{\mathrm{M},k}|r_{i_k,i_k}|^2}{\sum_{m=1}^{N}\left({g_{i_k,m}^{(\mathrm{R})}}^2+{g_{i_k,m}^{(\mathrm{I})}}^2\right)\gamma_{\mathrm{B},\theta_m}+1}, \tag{49}$$

$$\rho_{\mathrm{RB},k} \triangleq |\tilde{l}_{q_k,q_k}|^2\gamma_{\mathrm{R},k}. \tag{50}$$

In (48) and (50), to convert the problem into a signomial problem as shown in Appendix C, the lower bounds of the rates for Phase 2 are considered by ignoring the additive term '1' both in $R_{\mathrm{RM},k}^{*}$ and $R_{\mathrm{RB},k}^{*}$. The converted signomial optimization problem guarantees a global optimal solution which can be obtained by well-known signomial programming techniques such as reverse geometric programming [41]. The sum-rate loss due to the use of these lower bounds is minor

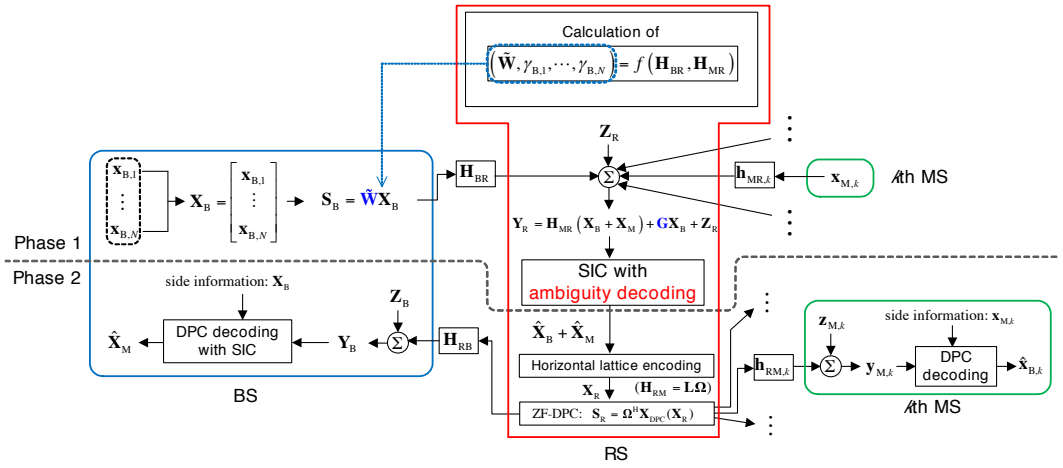


Fig. 2. Overall block diagram of the proposed TWR scheme

and becomes negligible in a high-SNR regime [22], [38]. The problem can be further simplified, because it is apparent that the optimal $\gamma_{\mathrm{M},k}$, which is proportional to $\rho_{\mathrm{MR},k}$ defined in (50), is given as $\gamma_{\mathrm{M},k} = P_{\mathrm{M},k}$. Therefore, each MS can determine $\gamma_{\mathrm{M},k}$ in a distributed manner.

Different choices of $\boldsymbol{\theta}$ and ϕ result in different diagonal values of $\mathbf{R}$ in (23) and $\mathbf{L}$ in (29), respectively, and eventually different achievable sum-rates. The optimal $\boldsymbol{\theta}$ and ϕ are found through an exhaustive search such that the sum-rate achieved from (46) is maximized. It is easy to foresee that the optimal $\boldsymbol{\theta}$ and ϕ minimize the differences between $R^*_{\mathrm{BR},k}$ and $R^*_{\mathrm{RM},k}$ and also between $R^*_{\mathrm{MR},k}$ and $R^*_{\mathrm{RB},k}$. In other words, the spatial permutation performs a rates-matching function.

Remark 2: The QR-SIC at the RS and BS can be replaced with the MMSE-SIC [3]. However, unlike the optimization problem (46), the optimization with this MMSE-SIC cannot be solved systematically to find a global optimal solution. Moreover, the performance loss due to the suboptimality of the QR-SIC is minor [38] and the QR-SIC is sufficient to achieve the cut-set bound in the high SNR regime [21], [22]. Our main focus is the asymptotic behavior in the high SNR regime, in which the achievable rate of the QR-SIC rapidly approaches to that of the MMSE-SIC [3], [38], [42].

E. Asymptotic Achievability of the Cut-set Bound for Asymmetric SNR TWRC

From (10), the cut-set bound on the sum-capacity can be rewritten as (51) at the bottom of the next page, where $\sigma_{\bar{\chi},k}$ is the kth singular value of $\mathbf{H}_{\bar{\chi}}$, $\bar{\chi} \in \{\mathrm{BR}, \mathrm{MR}, \mathrm{RB}, \mathrm{RM}\}$, and $\gamma^{\mathrm{c}}_{\mathrm{B},k}$, $\gamma^{\mathrm{c}}_{\mathrm{M},k}$, $\gamma^{\mathrm{c}}_{\mathrm{RM},k}$, and $\gamma^{\mathrm{c}}_{\mathrm{RB},k}$ are the optimal power constants that satisfy $\sum_{k=1}^{N} \gamma^{\mathrm{c}}_{\mathrm{B},k} \le P_{\mathrm{B}}$, $\sum_{k=1}^{N} \gamma^{\mathrm{c}}_{\mathrm{RM},k} \le P_{\mathrm{R}}$, $\sum_{k=1}^{N} \gamma^{\mathrm{c}}_{\mathrm{RB},k} \le P_{\mathrm{R}}$, $\gamma^{\mathrm{c}}_{\mathrm{M},k} \le P_{\mathrm{M},k}$, $k = 1, \ldots, N$. Here, $\gamma^{\mathrm{c}}_{\mathrm{B},k}$, $\gamma^{\mathrm{c}}_{\mathrm{RM},k}$, and $\gamma^{\mathrm{c}}_{\mathrm{RB},k}$ can be determined from the waterfilling optimizations[4] [3], and $\gamma^{\mathrm{c}}_{\mathrm{M},k} = P_{\mathrm{M},k}$. The following theorem shows that the sum-rate difference between the cut-set bound and the proposed scheme is bounded by $\sum_{k=1}^{N} o_{P_{\mathrm{M},k}}(1) + o_{P_{\mathrm{R}}}(1)$ or $o_{P_{\mathrm{R}}}(1)$, where $o_x(1) \to 0$ as $x \to \infty$, if SNR asymmetry is sufficiently large.

Theorem 2: For given $\mathbf{H}_{\bar{\chi}}$, $\boldsymbol{\theta}$, and ϕ, the sum-rate gap $C_{\text{cut-set}} - R_{\text{sum}}$ is bounded as

$$C_{\text{cut-set}} - R_{\text{sum}} \le \sum_{k=1}^{N} \frac{1}{{\sigma_{\mathrm{MR},k}}^2} \cdot \frac{1}{P_{\mathrm{M},k}} + \left(\frac{N^2}{{\sigma_{\mathrm{RM,min}}}^2} - \frac{N^2}{{\sigma_{\mathrm{RM,max}}}^2} + \sum_{i=1}^{N} \frac{N}{{\sigma_{\mathrm{RM},i}}^2} \right) \frac{1}{P_{\mathrm{R}}}, \tag{52}$$

if the inequalities E1 and E2 hold true, or as

$$C_{\text{cut-set}} - R_{\text{sum}} \le \left(\frac{N^2}{{\sigma_{\mathrm{RM,min}}}^2} - \frac{N^2}{{\sigma_{\mathrm{RM,max}}}^2} + \frac{N^2}{{\sigma_{\mathrm{RB,min}}}^2} - \frac{N^2}{{\sigma_{\mathrm{RB,max}}}^2} + \sum_{i=1}^{N} \left(\frac{N}{{\sigma_{\mathrm{RM},i}}^2} + \frac{N}{{\sigma_{\mathrm{RB},i}}^2} \right) \right) \frac{1}{P_{\mathrm{R}}}, \tag{53}$$

[4]Different power allocations at the RS for the channels $\mathbf{H}_{\mathrm{RM}}$ and $\mathbf{H}_{\mathrm{RB}}$ are not physically possible, however, respective optimal power allocations are separately assumed according to the cut-set theorem.

$$\max_{\mathbf{G},\gamma_{\mathrm{B}},\gamma_{\mathrm{M}},\gamma_{\mathrm{R}}} \sum_{k=1}^{N} \left(\min\left\{ [\log \rho_{\mathrm{BR},k}]^+, [\log \rho_{\mathrm{RM},k}]^+ \right\} + \min\left\{ [\log \rho_{\mathrm{MR},k}]^+, [\log \rho_{\mathrm{RB},k}]^+ \right\} \right) \tag{46a}$$

$$\text{s.t.} \sum_{n=1}^{N} \gamma_{\mathrm{B},\theta_n} \left\| \langle \left(\mathbf{Q}^{\mathrm{H}} \mathbf{H}_{\mathrm{BR}} \mathbf{V}_{\mathrm{BR}} \right)^{-1} \left(\mathbf{G}^{(\mathrm{R})} + j\mathbf{G}^{(\mathrm{I})} + \mathbf{R} \right) \boldsymbol{\Pi}_{\boldsymbol{\theta}} \rangle_n \right\|^2 \le P_{\mathrm{B}}, \tag{46b}$$

$$\gamma_{\mathrm{M},k} \le P_{\mathrm{M},k}, \quad k = 1, 2, \ldots, N, \quad \sum_{n=1}^{N} \gamma_{\mathrm{R},n} \le P_{\mathrm{R}}, \quad \mathrm{diag}\left(\mathbf{G}^{(\mathrm{R})} \right) = \mathrm{diag}\left(\mathbf{G}^{(\mathrm{I})} \right) = \mathbf{0}, \tag{46c}$$

if the inequalities E1 and E3 hold true, where E1, E2, and E3 are given by (54), (55), and (56), respectively, at the bottom of the page. Here, $\mathbf{W}_{\mathrm{ZF}} \triangleq \left(\mathbf{Q}^{\mathrm{H}}\mathbf{H}_{\mathrm{BR}}\mathbf{V}_{\mathrm{BR}}\right)^{-1}\mathbf{R}\boldsymbol{\Pi}_{\boldsymbol{\theta}}$, $P_{\mathrm{M,max}} \triangleq \max_k\{P_{\mathrm{M},k}\}$, $P_{\mathrm{M,min}} \triangleq \min_k\{P_{\mathrm{M},k}\}$, $\sigma_{\bar{\chi},\max} \triangleq \max_k\{|\sigma_{\bar{\chi},k}|\}$, $\sigma_{\bar{\chi},\min} \triangleq \min_k\{|\sigma_{\bar{\chi},k}|\}$, and $P_{\mathrm{B}}^0 < \infty$ and $P_{\mathrm{R}}^0 < \infty$ are minimum positive values for all spatial streams of $\mathbf{H}_{\mathrm{BR}}$ and $\mathbf{H}_{\mathrm{RB}}$, respectively, to be active; i.e., $\gamma_{\mathrm{B},k}^{\mathrm{c}}, \gamma_{\mathrm{RB},k}^{\mathrm{c}} > 0$, $\forall k$ [38].

Proof: See Appendix D. ∎

Therefore, with a certain level of SNR asymmetry derived in Theorem 2, the proposed scheme achieves the cut-set bound in high SNR regime.

E1 is the condition for the SNR from the BS to RS to be large enough compared to the SNR from the RS to MSs such that the end-to-end rate from the BS to MSs is given as $\sum_{k=1}^{N}\min\{R_{\mathrm{BR},k}, R_{\mathrm{RM},k}\} = \sum_{k=1}^{N} R_{\mathrm{RM},k}$. In other words, E1 is the condition for the rate-loss due to precoding at the BS to be negligible.

On the other hand, precoding for the channels from the MSs to RS and from the RS to BS is lossless, because only orthogonal transformation is performed for $\mathbf{H}_{\mathrm{MR}}$ and $\mathbf{H}_{\mathrm{RB}}$ (See (22) and (37)). However, the end-to-end rate from the MSs to BS of the cut-set bound is given by the minimum of the sums as $\min\left\{\sum_{k=1}^{N} R_{\mathrm{MR},k}, \sum_{k=1}^{N} R_{\mathrm{RB},k}\right\}$ from (51), whereas that of the proposed scheme is given by the sum of the minima as $\sum_{k=1}^{N}\min\{R_{\mathrm{MR},k}, R_{\mathrm{RB},k}\}$ due to non-cooperation among the MSs (See (43)). E2 or E3 is needed for the proposed scheme to overcome the rate loss due to this restriction. E2 is the condition for $\sum_{k=1}^{N}\min\{R_{\mathrm{MR},k}, R_{\mathrm{RB},k}\} = \sum_{k=1}^{N} R_{\mathrm{MR},k}$ and E3 is the condition for $\sum_{k=1}^{N}\min\{R_{\mathrm{MR},k}, R_{\mathrm{RB},k}\} = \sum_{k=1}^{N} R_{\mathrm{RB},k}$, both of which achieve the cut-set bound on the end-to-end rate from the MSs to BS.

We provide an intuition for the SNR asymmetry in the following remark.

Remark 3: If all of the channel matrices are identity matrices and equal power allocation is optimal, i.e., the AWGN channel in a high-SNR regime, E1 to E3 are simplified to E1: $P_{\mathrm{B}} \geq P_{\mathrm{R}} + N$, E2: $P_{\mathrm{R}}/N \geq P_{\mathrm{M,max}}$, and E3: $P_{\mathrm{M,min}} \geq 1 + P_{\mathrm{R}}/N$, which clearly shows the SNR asymmetry. Here, the additive N in E1 and 1 in E3 result from the loss of '1' of $\log(1+\mathrm{SNR})$ in $R_{\mathrm{BR},k}^{*}$ and $R_{\mathrm{MR},k}^{*}$, respectively.

IV. Simulation Results

The achievable sum-rate of the proposed scheme is evaluated in i.i.d. Rayleigh fading channels. The channel matrix $\mathbf{H}_{\bar{\chi}}$ was generated from $\mathbf{H}_{\bar{\chi}} = \upsilon_{\bar{\chi}}\tilde{\mathbf{H}}_{\bar{\chi}}$, where $\upsilon_{\bar{\chi}} > 0$ and $\tilde{\mathbf{H}}_{\bar{\chi}}$ is a random matrix, each element of which is i.i.d. according to $\mathcal{CN}(0,1)$. It was assumed that $N_{\mathrm{B}} = N_{\mathrm{R}} = N$ and that all of the SNRs between the RS and the kth MS, $k = 1, \ldots, N$, are identical with $P_{\mathrm{M},k} = P_{\mathrm{M}}$ for all k. The average SNRs are defined by $\mathrm{SNR}_{\mathrm{BR}} \triangleq \upsilon_{\mathrm{BR}}^2 N_{\mathrm{B}} P_{\mathrm{B}}$, $\mathrm{SNR}_{\mathrm{MR}} \triangleq \upsilon_{\mathrm{MR}}^2 P_{\mathrm{M}}$, $\mathrm{SNR}_{\mathrm{RB}} \triangleq \upsilon_{\mathrm{RB}}^2 N_{\mathrm{R}} P_{\mathrm{R}}$, and $\mathrm{SNR}_{\mathrm{RM}} \triangleq \upsilon_{\mathrm{RM}}^2 N_{\mathrm{R}} P_{\mathrm{R}}$.

We compare the achievable rate of optimal precoding described in Section III-B, denoted by 'Opt. Prec.' in the figures, with the achievable rates of ZF-based precoding and DF TWR [8] based on the MAC bound of (11), which are labeled by 'ZFP' and 'MAC TWR', respectively, in the figures. The proposed ZF-DPC scheme was used in Phase 2 for all these schemes, because the decoding procedure at the BS of [8] does not work (See Section III-C1). For a fair comparison, we consider a tighter upper bound on the sum of the bounds for each MS's rate, because the cut-set bound given by (10) is the bound on the sum of all MSs' rates. The maximum sum-rate for each MS is achieved by applying DPC with optimal permutation for each of the four channels $\mathbf{H}_{\mathrm{BR}}$, $\mathbf{H}_{\mathrm{MR}}$, $\mathbf{H}_{\mathrm{RB}}$, and $\mathbf{H}_{\mathrm{RM}}$ assuming that all these channels are separated according to the cut-set theorem. We denote this upper bound as 'Separate DPC bound'.

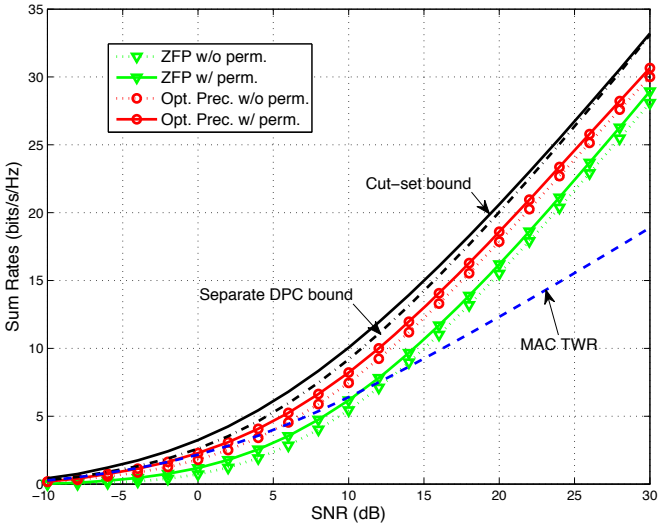
Figures 3(a) and 3(b) illustrate the achievable sum-rates with respect to SNRs when $N = 2$ and $N = 4$, respectively. The average SNRs were all assumed to be identical (symmetric SNRs). Figure 3(a) shows that the proposed scheme with optimal precoding exhibits higher sum-rates than the other schemes and achieves the cut-set bound within 3 bits. From the results of the separate DPC bound, we can understand the trend of the rate loss due to non-cooperation among the MSs, which vanishes as the SNRs increase [38]. The sum-rate gain of utilizing the ISI is observable from the sum-rate gap between the ZF-based precoding scheme and the optimal precoding scheme, and the permutations provide a sum-rate

$$C_{\text{cut-set}} = \min\left\{\sum_{k=1}^{N}\log\left(1+\gamma_{\mathrm{B},k}^{\mathrm{c}}{\sigma_{\mathrm{BR},k}}^2\right), \sum_{k=1}^{N}\log\left(1+\gamma_{\mathrm{RM},k}^{\mathrm{c}}{\sigma_{\mathrm{RM},k}}^2\right)\right\} + \min\left\{\sum_{k=1}^{N}\log\left(1+\gamma_{\mathrm{M},k}^{\mathrm{c}}{\sigma_{\mathrm{MR},k}}^2\right), \sum_{k=1}^{N}\log\left(1+\gamma_{\mathrm{RB},k}^{\mathrm{c}}{\sigma_{\mathrm{RB},k}}^2\right)\right\} \tag{51}$$

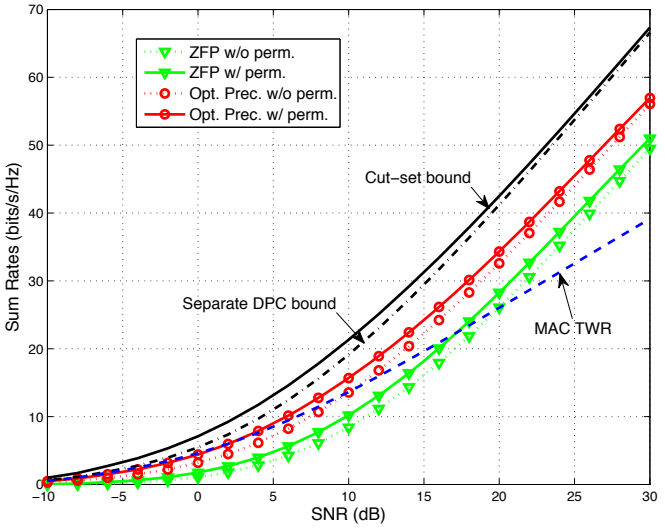
$$\text{E1: } P_{\mathrm{B}} \geq \max\left\{\frac{{\sigma_{\mathrm{RM,max}}}^2}{{\sigma_{\mathrm{BR,min}}}^2}\left(P_{\mathrm{R}} + \frac{N}{{\sigma_{\mathrm{RM,min}}}^2}\right) - \sum_{i=1}^{N}\frac{1}{{\sigma_{\mathrm{BR},i}}^2}, \max_k \frac{\|\mathbf{W}_{\mathrm{ZF}}\|_{\mathrm{F}}^2}{|r_{i_k,i_k}|^2}\left(1+|l_{q_k,q_k}|^2\frac{P_{\mathrm{R}}}{N}\right), P_{\mathrm{B}}^0\right\}, \tag{54}$$

$$\text{E2: } P_{\mathrm{R}} \geq \max\left\{\frac{N\left(P_{\mathrm{M,max}}{\sigma_{\mathrm{MR,max}}}^2 + 1\right)}{{\sigma_{\mathrm{RB,min}}}^2} - \sum_{i=1}^{N}\frac{1}{{\sigma_{\mathrm{RB},i}}^2}, \max_k \frac{N}{|\tilde{l}_{q_k,q_k}|^2}\left(P_{\mathrm{M},k}|r_{i_k,i_k}|^2 - 1\right), P_{\mathrm{R}}^0\right\}, \tag{55}$$

$$\text{E3: } P_{\mathrm{M,min}} \geq \max\left\{\frac{1}{{\sigma_{\mathrm{MR,min}}}^2}\left(\frac{{\sigma_{\mathrm{RB,max}}}^2 P_{\mathrm{R}}}{N} + \frac{{\sigma_{\mathrm{RB,max}}}^2}{{\sigma_{\mathrm{RB,min}}}^2} - 1\right), \max_k \frac{1+|\tilde{l}_{q_k,q_k}|^2 P_{\mathrm{R}}/N}{|r_{i_k,i_k}|^2}\right\}. \tag{56}$$



(a)



(b)

Fig. 3. Sum-rates vs. SNR for the MU-MIMO cTWRC when (a) $N_{\mathrm{B}} = N_{\mathrm{R}} = N = 2$ and (b) $N_{\mathrm{B}} = N_{\mathrm{R}} = N = 4$. The SNRs of the channels a assumed to be all the same (symmetric SNRs).

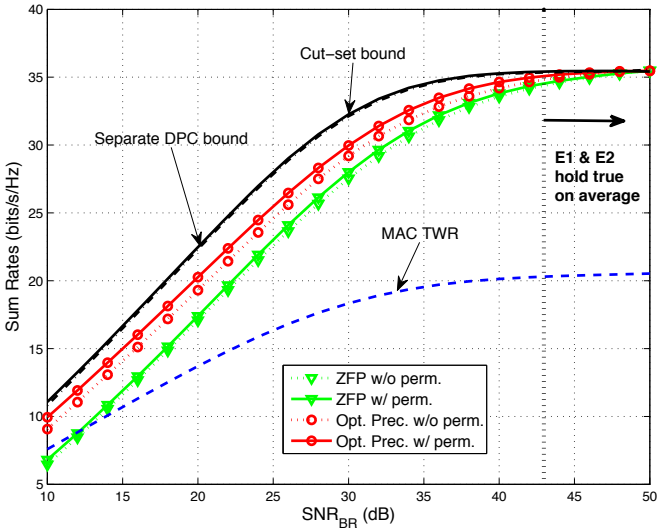


Fig. 4. Sum-rates vs. SNRs between the BS and RS for the MU-MIMO cTWRC with $N_{\mathrm{B}} = N_{\mathrm{R}} = N = 2$ when the average SNRs between the RS and MSs are assumed to be fixed at 30dB.

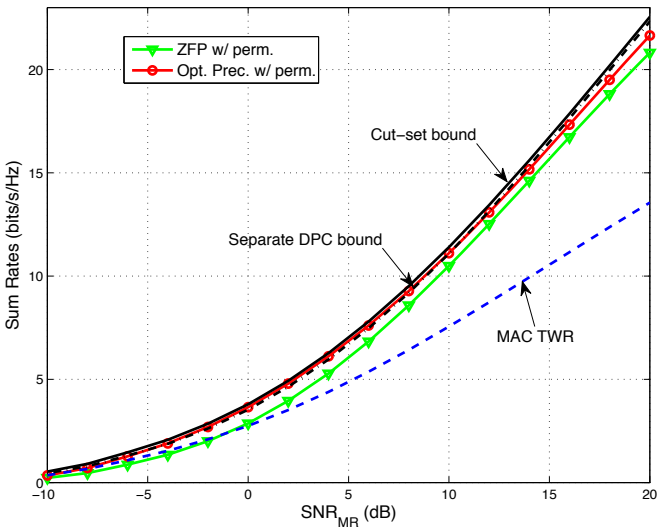


Fig. 5. Sum-rates vs. SNRs between the MSs and RS for the MU-MIMO cTWRC with $N_{\mathrm{B}} = N_{\mathrm{R}} = N = 2$. The average SNRs between the BS and RS are assumed to be fixed at 30dB.

gain of about 1 bit for $N = 2$, which is significant in low SNR regime. The MAC TWR scheme suffers from a loss of multiplexing gain in high SNR regime. Figgure 3(b) shows that the gap between the cut-set bound and the proposed schemes becomes greater as N increases, because proposed optimal precoding can be understood as a regularized ZF-based precoding (See (44)), and because the performance of optimal precoding as well as ZF-based precoding depends on the conditioning of $\mathbf{H}_{\mathrm{BR}}$ which becomes worse as N increases.

The permutation gain for $N = 4$ is increased up to 2 bits, because the gain of the rates-matching becomes large if the differences among the spatial rates are significant or if the number of spatial streams increases. It is clear from Fig. 3 that the proposed scheme with optimal precoding and permutation shows a higher sum-rate than the MAC TWR in all SNR regimes. Figures 4 and 5 illustrate the achievability of the cut-set bound in asymmetric SNR scenarios. Figure 4 demonstrates the sum-rates with respect to $\mathrm{SNR}_{\mathrm{BR}}$ for $N = 2$, where $\mathrm{SNR}_{\mathrm{MR}} = \mathrm{SNR}_{\mathrm{RM}} = 30$dB, and where $\mathrm{SNR}_{\mathrm{RB}} = \mathrm{SNR}_{\mathrm{BR}}$. It is seen from Fig. 4 that if $\mathrm{SNR}_{\mathrm{BR}}$ is large enough compared to $\mathrm{SNR}_{\mathrm{MR}}$ such that E1 and E2 hold true in an *average sense*, the cut-set bound can be achieved within one bit. On the other hand, Fig. 5 illustrates the sum-rates with respect to $\mathrm{SNR}_{\mathrm{MR}}$ for $N = 2$, where $\mathrm{SNR}_{\mathrm{RM}} = \mathrm{SNR}_{\mathrm{MR}}$ and $\mathbf{SNR}_{\mathrm{BR}} = \mathbf{SNR}_{\mathrm{RB}} = 30$dB. It can be seen that the proposed scheme with optimal precoding achieves the cut-set bound within one bit if $\mathrm{SNR}_{\mathrm{MR}} \leq 20$dB.

V. Conclusion

We have proposed an achievable scheme for the MU-MIMO cTWRC based on lattice code-aided linear precoding at the BS. Although linear precoding cannot be made lossless unlike the three-node case, it turned out that the proposed lattice coding and linear precoding schemes asymptotically achieve the sum-rate upper bound in the presence of SNR asymmetry. The proposed scheme with ZF-based precoding without spatial permutations is enough to prove the achievability of the cut-set bound, and optimal precoding can considerably increase the sum-rate in a symmetric SNR scenario or in a low-to-mid SNR regime of an asymmetric SNR scenario. The capacity region for the MU-MIMO cTWRC with symmetric SNRs remains to be open.

Although it has been assumed here that each MS has a single antenna, an extension to a case in which MSs have multiple antennas but with single spatial stream can be readily accomplished by applying the independent SVD precoding scheme at each MS.

Appendix A
Lattice Preliminaries and Construction of $\mathcal{C}_{\chi,k}$

A. Lattice Preliminaries

We briefly recall the relevant lattice theory which is used throughout the paper. For a comprehensive introduction to lattices, readers are referred to [17], [43]–[47] and the references therein. A T-dimensional lattice is a discrete additive group $\Lambda \subseteq \mathbb{R}^{1\times T}$, which is defined by its basis $\{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_m\}$, where $\mathbf{b}_i \in \mathbb{R}^{1\times T}$. That is, the lattice Λ is defined as $\Lambda = \{\boldsymbol{\lambda} : \boldsymbol{\lambda} = \sum_{i=1}^{m} \nu_i \mathbf{b}_i, \nu_i \in \mathbb{Z}\}$. The fundamental Voronoi region $\mathcal{R}(\Lambda)$ is the set of points $\mathbf{x} \in \mathbb{R}^{1\times T}$ closest to $\boldsymbol{\lambda} = \mathbf{0}$ compared to any other point $\boldsymbol{\lambda}'(\neq \boldsymbol{\lambda}) \in \Lambda$, and is isomorphic for all $\boldsymbol{\lambda} \in \Lambda$. Every $\mathbf{x} \in \mathbb{R}^{1\times T}$ can be uniquely written as

$$\mathbf{x} = \boldsymbol{\lambda} + \eta, \quad \boldsymbol{\lambda} \in \Lambda, \eta \in \mathcal{R}(\Lambda), \tag{57}$$

where $\boldsymbol{\lambda} = Q_\Lambda(\mathbf{x})$ is a nearest neighbor of $\mathbf{x}$ in Λ. The modulo-Λ is defined as

$$[\mathbf{x}]_\Lambda \triangleq \mathbf{x} - Q_\Lambda(\mathbf{x}) = \eta, \tag{58}$$

and the volume of $\mathcal{R}(\Lambda)$, $V(\mathcal{R}(\Lambda))$, and the average energy or the second moment per dimension $\rho^2(\Lambda)$ are given as

$$V(\mathcal{R}(\Lambda)) = \int_{\mathcal{R}} d\mathbf{x}, \quad \rho^2(\Lambda) = \frac{1}{T}\frac{\int_{\mathcal{R}} \|\mathbf{x}\|^2 d\mathbf{x}}{V(\mathcal{R})}. \tag{59}$$

The normalized second moment of Λ is

$$G(\Lambda) \triangleq \frac{\rho^2(\Lambda)}{V(\mathcal{R}(\Lambda))^{2/T}}. \tag{60}$$

Geometrically, $G(\Lambda)$ is always greater than $\frac{1}{2\pi e}$ and the lattice Λ is said to be good for MSE quantization or Poltyrev-good if $G(\Lambda) \to \frac{1}{2\pi e}$. The covering radius r_{cov} is the radius of the smallest sphere that contains $\mathcal{R}(\Lambda)$. The effective radius r_{eff} is the radius of the sphere with a volume equal to $V(\mathcal{R}(\Lambda))$. The lattice is good for covering or Rogers-good if it satisfies $r_{\text{cov}}/r_{\text{eff}} \to 1$.

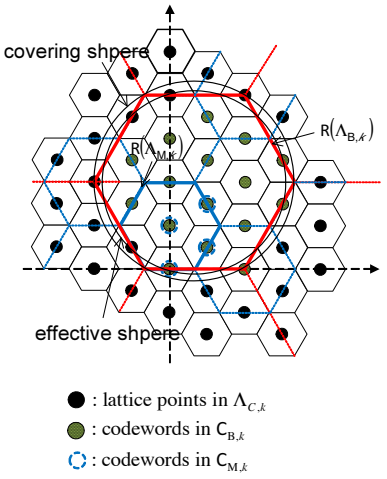


Fig. 6. Example of a 2-dimensional nested lattice code when $|\mathcal{C}_{\mathrm{B},k}| = 16$ and $|\mathcal{C}_{\mathrm{M},k}| = 4$

B. Construction of $\mathcal{C}_{\chi,k}$

For every $k \in \{1, \ldots, N\}$, we consider T-dimensional lattice codes $\mathcal{C}_{\mathrm{B},k}$ and $\mathcal{C}_{\mathrm{M},k}$ composed of the T-dimensional lattices $\Lambda_{C,k}$, $\Lambda_{\mathrm{B},k}$, and $\Lambda_{\mathrm{M},k}$, where $\Lambda_{C,k}$ is Poltyrev-good while $\Lambda_{\mathrm{B},k}$ and $\Lambda_{\mathrm{M},k}$ are simultaneously Poltyrev-good and Rogers-good. The lattice code $\mathcal{C}_{\chi,k} \subset \mathbb{R}^{1\times T}$ is defined by

$$\mathcal{C}_{\chi,k} \triangleq \{(\mathbf{v}_{\chi,k} + \Lambda_{C,k}) \cap \mathcal{R}(\Lambda_{\chi,k})\}, \tag{61}$$

where $\mathbf{v}_{\chi,k} \in \mathbb{R}^{1\times T}$ is the translate vector. From the Rogers-goodness, $\mathcal{R}(\Lambda_{\chi,k})$ is a T-dimensional sphere with a radius of $\sqrt{T \cdot \gamma_{\chi,k}/2}$, where $\gamma_{\chi,k}$ is the power constant, and $\mathbf{v}_{\chi,k}$ and $\Lambda_{C,k}$ are chosen for a given $\mathcal{R}(\Lambda_{\chi,k})$ such that $|\mathcal{C}_{\chi,k}| = 2^{TR_{\chi,k}{}'}$ [39]. Because $\mathbf{x}_{\chi,k}^{(\mathrm{R})}$ and $\mathbf{x}_{\chi,k}^{(\mathrm{I})}$ are inside the spherical-region of $\mathcal{R}(\Lambda_{\chi,k})$, we have

$$\frac{1}{T}\left\|\mathbf{x}_{\chi,k}^{(\mathrm{R})}\right\|^2 \le \frac{\gamma_{\chi,k}}{2}, \quad \frac{1}{T}\left\|\mathbf{x}_{\chi,k}^{(\mathrm{I})}\right\|^2 \le \frac{\gamma_{\chi,k}}{2}. \tag{62}$$

Assuming $\gamma_{\mathrm{B},k} \ge \gamma_{\mathrm{M},k}$ without loss of generality, $\Lambda_{\mathrm{B},k}$, $\Lambda_{\mathrm{M},k}$, and $\Lambda_{C,k}$ should be constructed such that $\Lambda_{\mathrm{B},k} \subseteq \Lambda_{\mathrm{M},k} \subseteq \Lambda_{C,k}$ according to [19, Theorem 2]. As a consequence, the code rates satisfy

$$\begin{aligned} R_{\mathrm{B},k}{}' &= \frac{1}{T}\log|\mathcal{C}_{\mathrm{B},k}| = \frac{1}{T}\log\left(\frac{V(\Lambda_{\mathrm{B},k})}{V(\Lambda_{C,k})}\right) \\ &= R_{\mathrm{M},k}{}' + \frac{1}{2}\log\left(\frac{\gamma_{\mathrm{B},k}}{\gamma_{\mathrm{M},k}}\right) + o_T(1), \end{aligned} \tag{63}$$

where $o_T(1) \to 0$ as $T \to \infty$. An example of a 2-dimensional nested lattice code is depicted in Fig. 6 for $|\mathcal{C}_{\mathrm{B},k}| = 16$ and $|\mathcal{C}_{\mathrm{M},k}| = 4$.

Appendix B
Proof of Theorem 1

The RS decodes $\mathbf{x}_{\mathrm{B},k}^{(\mathrm{R})} + \mathbf{x}_{\mathrm{M},k}^{(\mathrm{R})}$ from (27) using the ambiguity decoder. By construction, i.e., $\Lambda_{\mathrm{B},k} \subseteq \Lambda_{\mathrm{M},k} \subseteq \Lambda_{C,k}$ if $\gamma_{\mathrm{B},k} \ge \gamma_{\mathrm{M},k}$, or $\Lambda_{\mathrm{M},k} \subseteq \Lambda_{\mathrm{B},k} \subseteq \Lambda_{C,k}$ if $\gamma_{\mathrm{B},k} \le \gamma_{\mathrm{M},k}$, we have $\mathbf{x}_{\mathrm{B},k}^{(\mathrm{R})} + \mathbf{x}_{\mathrm{M},k}^{(\mathrm{R})} \in \Lambda_{C,k}$ as $\mathbf{x}_{\mathrm{B},k}^{(\mathrm{R})}, \mathbf{x}_{\mathrm{M},k}^{(\mathrm{R})} \in \Lambda_{C,k}$. The decoder outputs $\hat{\mathbf{x}}_{\mathrm{add},k}^{(\mathrm{R})} \in \Lambda_{C,k}$ if it is the only lattice point such that $\bar{\mathbf{y}}_{\mathrm{R},i_k}^{(\mathrm{R})} \in \hat{\mathbf{x}}_{\mathrm{add},k}^{(\mathrm{R})} + \mathcal{E}$, where $\mathcal{E} \subset \mathbb{R}^{1\times T}$ is the decision

region. If $\bar{\mathbf{y}}_{\mathrm{R},i_k}^{(\mathrm{R})} \in \{\hat{\mathbf{x}}_{\mathrm{add},k}^{(\mathrm{R})} + \mathcal{E}\} \cap \{\tilde{\mathbf{x}}_{\mathrm{add},k}^{(\mathrm{R})} + \mathcal{E}\}$ for some $\hat{\mathbf{x}}_{\mathrm{add},k}^{(\mathrm{R})}, \tilde{\mathbf{x}}_{\mathrm{add},k}^{(\mathrm{R})} \in \Lambda_{C,k}$, where $\hat{\mathbf{x}}_{\mathrm{add},k}^{(\mathrm{R})} \neq \tilde{\mathbf{x}}_{\mathrm{add},k}^{(\mathrm{R})}$, the ambiguity event $\mathcal{A}$ is declared. The error probability is given by [39]

$$\begin{aligned} p_e(\mathcal{E}|\Lambda_{C,k}) &= \Pr\left\{\hat{\mathbf{x}}_{\mathrm{add},k}^{(\mathrm{R})} \neq (\mathbf{x}_{\mathrm{B},k}^{(\mathrm{R})} + \mathbf{x}_{\mathrm{M},k}^{(\mathrm{R})})\right\} \\ &= \Pr\{\Re(\boldsymbol{\varepsilon}_k) \notin \mathcal{E} \text{ or } \mathcal{A}\} \\ &\leq \Pr\{\Re(\boldsymbol{\varepsilon}_k) \notin \mathcal{E}\} + \Pr\{\mathcal{A}\}. \end{aligned} \tag{64}$$

For some $\delta > 0$, we consider the following decision region:

$$\mathcal{E} = \left\{\mathbf{x} \in \mathbb{R}^{1\times T} : \|\mathbf{x}\|^2 \leq T(1+\delta){\nu_k}^2\right\}, \tag{65}$$

where ${\nu_k}^2$ is defined from

$$\begin{aligned} \frac{1}{T}E\left\|\Re(\boldsymbol{\varepsilon}_k)\right\|^2 &= \frac{1}{T}E\sum_{m=1}^{N}\left\|\Re\left(\frac{g_{i_k,m}}{r_{i_k,i_k}}\mathbf{x}_{\mathrm{B},\theta_m}\right)\right\|^2 \\ &\quad + \frac{1}{T}E\left\|\Re\left(\frac{\mathbf{z}_{\mathrm{R},i_k}}{r_{i_k,i_k}}\right)\right\|^2 \\ &\leq \sum_{m=1}^{N}\frac{|g_{i_k,m}|^2}{|r_{i_k,i_k}|^2}\cdot\frac{{\gamma_{\mathrm{B},\theta_m}}^2}{2} + \frac{1}{2\,|r_{i_k,i_k}|^2} \triangleq {\nu_k}^2. \end{aligned} \tag{66}$$

From [39, Theorem 4], we can get the error probability averaged over $\Lambda_{C,k}$ as

$$\begin{aligned} \bar{p}_e(\mathcal{E}) &\triangleq E_{\Lambda_{C,k}}\left[p_e(\mathcal{E}|\Lambda_{C,k})\right] \\ &\leq \Pr\{\Re(\boldsymbol{\varepsilon}_k) \notin \mathcal{E}\} + (1+\delta')V(\mathcal{E})/V(\Lambda_{C,k}), \end{aligned} \tag{67}$$

where $\delta' \to 0$ as $T \to \infty$.

We first consider the last term of (67), which is the probability of the ambiguity event. Since $V(\Lambda_{\mathrm{B},k})/V(\Lambda_{C,k}) = 2^{TR_{\mathrm{B},k}{}'}$ from (63), we have

$$\begin{aligned} \frac{(1+\delta')\cdot V(\mathcal{E})}{V(\Lambda_{C,k})} &= (1+\delta')\frac{V\left(B(\sqrt{T})\right)(1+\delta)^{\frac{T}{2}}{\nu_k}^T}{V(\Lambda_{\mathrm{B},k})}2^{TR_{\mathrm{B},k}{}'} \\ &= (1+\delta')\cdot 2^{-T\cdot\kappa\left(R_{\mathrm{B},k}{}',T\right)}, \end{aligned} \tag{68}$$

where $B(\sqrt{T})$ is a T-dimensional sphere of radius $\sqrt{T}$ and the error exponent κ is given by

$$\begin{aligned} \kappa\left(R_{\mathrm{B},k}{}',T\right) \triangleq &-\frac{1}{T}\log\frac{V\left(B(\sqrt{T})\right)}{V(\Lambda_{\mathrm{B},k})} \\ &-\frac{1}{2}\log(1+\delta) - \frac{1}{2}\log{\nu_k}^2 - R_{\mathrm{B},k}{}'. \end{aligned} \tag{69}$$

From the Rogers-goodness of $\Lambda_{\mathrm{B},k}$, i.e., $\lim_{T\to\infty}\frac{V(\Lambda_{\mathrm{B},k})}{V(B(\sqrt{T}))} = (\frac{\gamma_{\mathrm{B},k}}{2})^{T/2}$, κ is simplified as

$$\kappa\left(R_{\mathrm{B},k}{}',T\right) \to \frac{1}{2}\log\frac{\gamma_{\mathrm{B},k}/2}{{\nu_k}^2} - \frac{1}{2}\log(1+\delta) - R_{\mathrm{B},k}{}' \tag{70}$$

as $T \to \infty$. Therefore, we have $(1+\delta')V(\mathcal{E})/V(\Lambda_{C,k}) \leq \epsilon$, where $\epsilon \to 0$ as $T \to \infty$, if

$$R_{\mathrm{B},k}{}' \leq \frac{1}{2}\log\frac{\gamma_{\mathrm{B},k}/2}{{\nu_k}^2} = \frac{1}{2}\log\frac{\gamma_{\mathrm{B},k}\,|r_{i_k,i_k}|^2}{\sum_{m=1}^{N}|g_{i_k,m}|^2\gamma_{\mathrm{B},\theta_m}+1} \tag{71}$$

From (63), (71) is equivalent to the condition

$$R_{\mathrm{M},k}{}' \leq \frac{1}{2}\log\frac{\gamma_{\mathrm{M},k}/2}{{\nu_k}^2} = \frac{1}{2}\log\frac{\gamma_{\mathrm{M},k}\,|r_{i_k,i_k}|^2}{\sum_{m=1}^{N}|g_{i_k,m}|^2\gamma_{\mathrm{B},\theta_m}+1}. \tag{72}$$

We now move our attention to $\Pr\{\Re(\boldsymbol{\varepsilon}_k) \notin \mathcal{E}\}$. Since $\|\mathbf{x}_{\mathrm{B},k}^{(\mathrm{R})}\|^2, \|\mathbf{x}_{\mathrm{B},k}^{(\mathrm{I})}\|^2 \leq T(\gamma_{\mathrm{B},k}/2)$, we have

$$\begin{aligned} \left\|\sum_{m=1}^{N}\Re\left(\frac{g_{i_k,m}}{r_{i_k,i_k}}\mathbf{x}_{\mathrm{B},\theta_m}\right)\right\|^2 &\leq \sum_{m=1}^{N}\left\|\Re\left(\frac{g_{i_k,m}}{r_{i_k,i_k}}\mathbf{x}_{\mathrm{B},\theta_m}\right)\right\|^2 \\ &\leq T\sum_{m=1}^{N}\left(\frac{\gamma_{\mathrm{B},\theta_m}}{2}\cdot\frac{|g_{i_k,m}|^2}{|r_{i_k,i_k}|^2}\right) \end{aligned} \tag{73}$$

Furthermore, for any $\xi < 2\,|r_{i_k,i_k}|^2$, the moment-generating function of $\Re(\boldsymbol{\varepsilon}_k)$ is given by (74) to (76) at the bottom of the next page, where $P_N \triangleq \left(2\,|r_{i_k,i_k}|^2\right)^{-1}$ and $P_I \triangleq \sum_{m=1}^{N}\left(\gamma_{\mathrm{B},\theta_m}/2\cdot|g_{i_k,m}|^2/|r_{i_k,i_k}|^2\right)$. Here, (74) follows from the fact that $2\,|r_{i_k,i_k}|^2\cdot\|\Re(\boldsymbol{\varepsilon}_k)\|^2$, conditioned on $\mathbf{x}_{\mathrm{B},n}$, $\forall n \neq k$, is a noncentral chi-square random variable and (75) follows from (73). For any $\delta > 0$, we get

$$\begin{aligned} \Pr\{\Re(\boldsymbol{\varepsilon}_k) \notin \mathcal{E}\} &= \Pr\left\{\|\Re(\boldsymbol{\varepsilon}_k)\|^2 \geq T(1+\delta){\nu_k}^2\right\} \\ &\leq \exp\left(-T\xi(1+\delta){\nu_k}^2 + \ln M_k(\xi)\right) \qquad (77) \\ &\leq \exp\Bigg(-T\Bigg(\xi(1+\delta){\nu_k}^2 + \frac{1}{2}\ln(1-2P_N\xi) \\ &\qquad - \frac{\xi}{1-2P_N\xi}P_I\Bigg)\Bigg) \qquad (78) \\ &\leq \exp\Bigg(-\frac{T}{2}\Bigg((P_I/P_N+1) \\ &\quad\times\left(1+(1+\delta)-\frac{2(1+\delta)}{\tau}\right)-\ln\tau\Bigg)\Bigg), \qquad (79) \end{aligned}$$

where

$$\tau \triangleq \frac{\sqrt{(4P_I^2/P_N^2 + 4P_I/P_N)(1+\delta)+1}-1}{2P_I/P_N}. \tag{80}$$

Here, (77) follows from the Chernoff bound [48] on $\Re(\boldsymbol{\varepsilon}_k)$, (78) follows from (76), and (79) follows from choosing ξ such that the righthand-side of (78) is maximized (See [49, Lemma 6]). The error exponent of (79) is positive for any $\delta > 0$ and $\Pr\{\Re(\boldsymbol{\varepsilon}_k) \notin \mathcal{E}\}$ vanishes as $T \to \infty$. The proof of Theorem 1 is completed from (71), (72), and (79). It is clear that decoding $\mathbf{x}_{\mathrm{B},k}^{(\mathrm{I})}+\mathbf{x}_{\mathrm{M},k}^{(\mathrm{I})}$ from $\mathbf{x}_{\mathrm{B},k}^{(\mathrm{I})}+\mathbf{x}_{\mathrm{M},k}^{(\mathrm{I})}+\Im(\boldsymbol{\varepsilon}_k)$ gives us the same results.

APPENDIX C
CONVERSION OF (46) INTO A SIGNOMIAL PROBLEM

A function f is termed a *signomial* function if f is defined as [41]

$$f(x_1,\ldots,x_n) = \sum_{i=1}^{m}a_i\prod_{l=1}^{n}{x_l}^{\tau_i^{(l)}}, \tag{81}$$

where $x_l \geq 0$, $a_i \in \mathbb{R}$, and $\tau_i^{(l)} \in \mathbb{R}$, $i = 1,\ldots,m$, $l = 1,\ldots,n$. If a problem consists of a signomial cost function and upper bound inequalities with signomial functions, the problem is known as a signomial problem.

Now, let us consider the conversion of (46) into a signomial problem. Ignoring the $[\cdot]^+$ operations which can be taken into consideration inductively as in the waterfilling problem [3], we can equivalently write the maximization of (46a) as

$$\max \sum_{n=1}^{N} (u_n + v_n) \tag{82a}$$

$$\text{s.t. } u_k = \log(\tilde{u}_k), \;\; v_k = \log(\tilde{v}_k) \tag{82b}$$

$$\rho_{\mathrm{BR},k} \geq \tilde{u}_k, \rho_{\mathrm{RM},k} \geq \tilde{u}_k, \rho_{\mathrm{MR},k} \geq \tilde{v}_k, \rho_{\mathrm{RB},k} \geq \tilde{v}_k, \tag{82c}$$

$$k = 1, \ldots, N$$

with the auxiliary variables $u_k, v_k, \tilde{u}_k, \tilde{v}_k \geq 0$. Furthermore, because $g_{n,m}$ can have negative values, we define $\bar{\mathbf{G}}^{(\mathrm{R})}, \underline{\mathbf{G}}^{(\mathrm{R})}, \bar{\mathbf{G}}^{(\mathrm{I})}, \underline{\mathbf{G}}^{(\mathrm{I})} \in \mathbb{R}_{++}^{N \times N}$ such that $\mathbf{G}^{(\mathrm{R})} = \bar{\mathbf{G}}^{(\mathrm{R})} - \underline{\mathbf{G}}^{(\mathrm{R})}$ and $\mathbf{G}^{(\mathrm{I})} = \bar{\mathbf{G}}^{(\mathrm{I})} - \underline{\mathbf{G}}^{(\mathrm{I})}$. Now, all the variables are represented by non-negative variables, and it is easy to verify that the inequalities in (82c), (46b), and (46c) are signomials. Note that the equality in (46c) can be removed by fixing $\mathrm{diag}(\mathbf{G}) = \mathbf{0}$ and that the equalities in (82b) vanish in the exponential variable change step of the signomial programming process.

Appendix D
Proof of Theorem 2

We only show that $C_{\text{cut-set}} - R_{\text{sum}}$ is bounded by (52) if E1 and E2 hold true. The proof of (53) where E1 and E3 hold true is analogous. Let us first consider the sum-capacity of the cut-set bound. From the inequality E1 with the first term of the outmost $\max$ operation, we have

$$\frac{{\sigma_{\mathrm{BR,min}}}^2}{N}\left(P_{\mathrm{B}} + \sum_{i=1}^{N} \frac{1}{{\sigma_{\mathrm{BR},i}}^2}\right) \geq \frac{P_{\mathrm{R}}{\sigma_{\mathrm{RM,max}}}^2}{N} + \frac{{\sigma_{\mathrm{RM,max}}}^2}{{\sigma_{\mathrm{RM,min}}}^2}. \tag{83}$$

Given that $P_{\mathrm{B}} \geq P_{\mathrm{B}}^0$, we have the waterfilling solutions of

$$\gamma_{\mathrm{B},k}^{\mathrm{c}} = \frac{P_{\mathrm{B}}}{N} + \frac{1}{N}\sum_{i=1}^{N} \frac{1}{{\sigma_{\mathrm{BR},i}}^2} - \frac{1}{{\sigma_{\mathrm{BR},k}}^2}. \tag{84}$$

This gives

$$\begin{aligned}\gamma_{\mathrm{B},k}^{\mathrm{c}}{\sigma_{\mathrm{BR},k}}^2 &= \frac{{\sigma_{\mathrm{BR},k}}^2}{N}\left(P_{\mathrm{B}} + \sum_{i=1}^{N} \frac{1}{{\sigma_{\mathrm{BR},i}}^2}\right) - 1 \\ &\geq \underbrace{\frac{{\sigma_{\mathrm{BR,min}}}^2}{N}\left(P_{\mathrm{B}} + \sum_{i=1}^{N} \frac{1}{{\sigma_{\mathrm{BR},i}}^2}\right)}_{A} - 1 \end{aligned} \tag{85}$$

On the other hand, we have

$$\begin{aligned}\gamma_{\mathrm{RM},k}^{\mathrm{c}}{\sigma_{\mathrm{RM},k}}^2 &\leq \gamma_{\mathrm{RM,max}}^{\mathrm{c}}{\sigma_{\mathrm{RM,max}}}^2 && (86)\\ &\leq \left(\frac{P_{\mathrm{R}}}{N} + \frac{1}{{\sigma_{\mathrm{RM,min}}}^2} - \frac{1}{{\sigma_{\mathrm{RM,max}}}^2}\right){\sigma_{\mathrm{RM,max}}}^2 && (87)\\ &= \underbrace{\frac{P_{\mathrm{R}}{\sigma_{\mathrm{RM,max}}}^2}{N} + \frac{{\sigma_{\mathrm{RM,max}}}^2}{{\sigma_{\mathrm{RM,min}}}^2}}_{B} - 1, && (88)\end{aligned}$$

where $\gamma_{\mathrm{RM,max}}^{\mathrm{c}} \triangleq \max_k\{\gamma_{\mathrm{RM},k}^{\mathrm{c}}\}$, and (87) follows from the fact that $\gamma_{\mathrm{RM,max}}^{\mathrm{c}} \leq P_{\mathrm{R}}/N + {\sigma_{\mathrm{RM,min}}}^{-2} - {\sigma_{\mathrm{RM,max}}}^{-2}$ [50, Theorem 1]. From (83), (85), and (88), we have

$$\gamma_{\mathrm{B},k}^{\mathrm{c}}{\sigma_{\mathrm{BR},k}}^2 \geq A - 1 \geq B - 1 \geq \gamma_{\mathrm{RM},k}^{\mathrm{c}}{\sigma_{\mathrm{RM},k}}^2. \tag{89}$$

Similarly, from E2 with the first term of the outmost $\max$ operation, it is easy to show that

$$\gamma_{\mathrm{RB},k}^{\mathrm{c}}{\sigma_{\mathrm{RB},k}}^2 \geq \gamma_{\mathrm{M},k}^{\mathrm{c}}{\sigma_{\mathrm{MR},k}}^2 = P_{\mathrm{M},k}{\sigma_{\mathrm{MR},k}}^2. \tag{90}$$

Because the inequalities (89) and (90) hold true for all k, the sum-capacity is given by

$$\begin{aligned}C_{\text{cut-set}} = &\sum_{k=1}^{N} \log\left(1 + \gamma_{\mathrm{RM},k}^{\mathrm{c}} \cdot {\sigma_{\mathrm{RM},k}}^2\right) \\ &+ \sum_{k=1}^{N} \log\left(1 + P_{\mathrm{M},k} \cdot {\sigma_{\mathrm{MR},k}}^2\right). \end{aligned} \tag{91}$$

Let us now consider the sum-rate of the proposed scheme. For simplicity, we restrict the proposed scheme to ZF-based precoding such that $\mathbf{W} = \left(\mathbf{Q}^{\mathrm{H}}\mathbf{H}_{\mathrm{BR}}\mathbf{V}_{\mathrm{BR}}\right)^{-1}\mathbf{R}\boldsymbol{\Pi}_{\boldsymbol{\theta}} \triangleq \mathbf{W}_{\mathrm{ZF}}$ and $\mathbf{G} = \mathbf{0}$. We further restrict the proposed scheme by applying the equal power allocation at the BS and RS such that $\gamma_{\mathrm{B},k} = P_{\mathrm{B}}/\|\mathbf{W}_{\mathrm{ZF}}\|_{\mathrm{F}}^2$ and $\gamma_{\mathrm{R},k} = P_{\mathrm{R}}/N$. Note again that the maximum power allocation at each MS is always optimal, i.e. $\gamma_{\mathrm{M},k} = P_{\mathrm{M},k}$. We denote the sum-rate of this simplified TWR scheme based on ZF-based precoding as $R_{\text{sum}}^{\mathrm{ZF}}$.

From E1 and E2, we have

$$P_{\mathrm{B}}\left|r_{i_k,i_k}\right|^2 / \|\mathbf{W}\|_{\mathrm{F}}^2 \geq 1 + \left|l_{q_k,q_k}\right|^2 P_{\mathrm{R}}/N, \tag{92}$$

$$1 + \left|\tilde{l}_{q_k,q_k}\right|^2 P_{\mathrm{R}}/N \geq P_{\mathrm{M},k}\left|r_{i_k,i_k}\right|^2, \tag{93}$$

for $k = 1, \ldots, N$. From (28), (36), and (42), the sum-rate is given by (94) to (96) at the bottom of the next page, where $\boldsymbol{\Omega}_{\mathrm{M}} \triangleq \mathrm{diag}(P_{\mathrm{M},1}, \ldots, P_{\mathrm{M},N})$, and (94) follows from (92) and (93).

$$\begin{aligned}M_k(\xi) \triangleq E\left\{\exp\left(\xi\|\Re(\boldsymbol{\varepsilon}_k)\|^2\right)\right\} &= E\left\{E\left\{\exp\left(\xi\|\Re(\boldsymbol{\varepsilon}_k)\|^2\right)\middle|\mathbf{x}_{\mathrm{B},n}, \forall n \neq k\right\}\right\} \\ &= E\left\{\exp\left(-\frac{T}{2}\ln\left(1 - 2\left(2\left|r_{i_k,i_k}\right|^2\right)^{-1}\xi\right) + \frac{\xi\left\|\sum_{m=1}^{N}\Re\left(\frac{g_{i_k,m}}{r_{i_k,i_k}}\mathbf{x}_{\mathrm{B},\theta_m}\right)\right\|^2}{1 - 2\left(2\left|r_{i_k,i_k}\right|^2\right)^{-1}\xi}\right)\right\} && (74)\\ &\leq \exp\left(-\frac{T}{2}\ln\left(1 - 2\left(2\left|r_{i_k,i_k}\right|^2\right)^{-1}\xi\right) + \frac{T\xi\sum_{m=1}^{N}\left(\gamma_{\mathrm{B},\theta_m}/2 \cdot \left|g_{i_k,m}\right|^2/\left|r_{i_k,i_k}\right|^2\right)}{1 - 2\left(2\left|r_{i_k,i_k}\right|^2\right)^{-1}\xi}\right) && (75)\\ &= \exp\left(-T\left[1/2 \cdot \ln\left(1 - 2P_N\xi\right) - \xi P_I/(1 - 2P_N\xi)\right]\right), && (76)\end{aligned}$$

As R_{sum} is the maximum sum-rate of the proposed scheme that is obtained from the optimal $\mathbf{W}$ and power allocation, $R_{\text{sum}} \geq R_{\text{sum}}^{\text{ZF}}$. Therefore, the sum-rate difference is given from (91) and (96) as

$$\begin{aligned} C_{\text{cut-set}} - R_{\text{sum}} &\leq C_{\text{cut-set}} - R_{\text{sum}}^{\text{ZF}} \\ &\leq \sum_{k=1}^{N} \Big(\log\left(1 + \gamma_{\text{RM},k}^{\text{c}} \cdot {\sigma_{\text{RM},k}}^2\right) \\ &\qquad - \log\left(P_{\text{R}}/N \cdot {\sigma_{\text{RM},k}}^2\right) \Big) \\ &\quad + \sum_{k=1}^{N} \log\left({P_{\text{M},k}}^{-1} \cdot {\sigma_{\text{MR},k}}^{-2} + 1\right) \end{aligned} \tag{97}$$

From the fact that $P_{\text{R}}/N \leq \gamma_{\text{RM,max}}^{\text{c}} \leq P_{\text{R}}/N + {\sigma_{\text{RM,min}}}^{-2} - {\sigma_{\text{RM,max}}}^{-2}$, we have

$$\begin{aligned} \log(1+\gamma_{\text{RM},k}^{\text{c}}{\sigma_{\text{RM},k}}^2) &\leq \log(1 + \gamma_{\text{RM,max}}^{\text{c}}{\sigma_{\text{RM},k}}^2) \\ &= \log(\gamma_{\text{RM,max}}^{\text{c}}) + \log\left(\frac{1}{\gamma_{\text{RM,max}}^{\text{c}}} + {\sigma_{\text{RM},k}}^2\right) \\ &\leq \log(P_{\text{R}}/N + {\sigma_{\text{RM,min}}}^{-2} - {\sigma_{\text{RM,max}}}^{-2}) \\ &\quad + \log(N/P_{\text{R}} + {\sigma_{\text{RM},k}}^2). \end{aligned} \tag{98}$$

Inserting (98) into (97) and using the fact that $\log(1+x) \leq x$ for any $x \geq 0$, we get

$$\begin{aligned} C_{\text{cut-set}} - R_{\text{sum}} \leq &\left(\frac{N^2}{{\sigma_{\text{RM,min}}}^2} - \frac{N^2}{{\sigma_{\text{RM,max}}}^2} + \sum_{i=1}^{N} \frac{N}{{\sigma_{\text{RM},i}}^2} \right) \cdot \frac{1}{P_{\text{R}}} \\ &+ \sum_{k=1}^{N} \frac{1}{{\sigma_{\text{MR},k}}^2} \cdot \frac{1}{P_{\text{M},k}}. \end{aligned} \tag{99}$$

REFERENCES

[1] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3037–3063, Sept. 2005.

[2] C.-B. Chae and R. W. Heath, Jr., "MIMO relaying with linear processing for multiuser transmission in fixed relay networks," *IEEE Trans. Signal Process.*, vol. 56, no. 2, pp. 727–738, Feb. 2008.

[3] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[4] W. Xu, X. Dong, and W.-S. Lu, "MIMO relaying broadcast channels with linear precoding and quantized channel state information feedback," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5233–5245, 2010.

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

[6] S. Zhang, S. C. Liew, and P. P. Lam, "Hot topic: Physical-layer network coding," in *Proc. ACM MobiCom*, Los Angenles, CA, Sept. 2006.

[7] B. Rankov and A. Wittneben, "Achievable rate regions for the two-way relay channel," in *Proc. IEEE ISIT*, Seattle, WA, July 2006.

[8] C. Eşli and A. Wittneben, "Multiuser MIMO two-way relaying for cellular communications," in *Proc. IEEE PIMRC*, Cannes, France, Sept. 2008, pp. 1–6.

[9] Z. Ding, I. Krikidis, J. Thompson, and K. K. Leung, "Physical layer network coding and precoding for the two-way relay channel in cellular systems," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 696–712, 2011.

[10] S. Xu and Y. Hua, "Optimal design of spatial source-and-relay matrices for a non-regenerative two-way MIMO relay system," *IEEE Trans. Wireless Commun.*, vol. 10, no. 5, pp. 1645–1655, May 2011.

[11] Y. Rong, "Joint source and relay optimization for two-way MIMO multi-relay networks," *IEEE Commun. Lett.*, vol. 15, no. 12, pp. 1329–1331, Dec. 2011.

[12] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiely, 2006.

[13] C. E. Shannon, "Two-way communication channels," in *Proc. 4th Berkely Symp. Math. Stat. Prob.*, vol. 1, Berkely, CA, 1961, pp. 611–644.

[14] 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, July 2000.

[15] T. Koike-Akino, P. Popovski, and V. Tarokh, "Optimized constellations for two-way wireless relaying with physical network coding," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 5, pp. 773–787, June 2009.

[16] K. Narayanan, M. P. Wilson, and A. Sprintson, "Joint physical layer coding and network coding for bi-directional relaying," in *Proc. Annu. Allerton Conf. Commun., Control, Comput.*, Monticello, IL, Sept. 2007.

[17] U. Erez and R. Zamir, "Achieving 1/2 log(1 + SNR) on the AWGN channel with lattice encoding and decoding," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2293–2314, Oct. 2004.

[18] H. E. Gamal, G. Caire, and M. O. Damen, "Lattice coding and decoding achieve the optimal diversity-multiplexing tradeoff of MIMO channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 968–985, June 2004.

[19] W. Nam, S.-Y. Chung, and 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.

[20] D. Gündüz, E. Tuncel, and J. Nayak, "Rate regions for the separated two-way relay channel," in *Proc. Annu. Allerton Conf. Commun., Control, Comput.*, Moticello, IL, Sept. 2008.

[21] 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.

[22] ——, "Asymptotic capacity of the separated MIMO two-way relay channel with linear precoding," in *Proc. Annu. Allerton Conf. Commun., Control, Comput.*, Moticello, IL, Sept.-Oct. 2010.

[23] A. Khina, Y. Kochman, and U. Erez, "Physical-layer MIMO relaying," *arXiv*, 2011, available at http://arxiv.org/abs/1102.5357.

[24] C. Eşli and A. Wittneben, "One- and two-way decode-and-forward relaying for wireless multiuser MIMO networks," in *Proc. IEEE GLOBECOM*, New Orleans, LO, Nov.-Dec. 2008, pp. 1–6.

[25] A. U. T. Amah, A. Klein, Y. C. B. Silva, and A. Fernekeß, "Multi-group multicast beamforming for multi-user two-way relaying," in *Proc. Int'l ITG Worksh. Smart Antennas*, Berlin, Germany, Feb. 2009.

[26] A. Sezgin, H. Boche, and S. A. Avestimehr, "Bidirectional multi-pair network with a MIMO relay: Beamforming strategies and lack of duality," *arXiv*, 2010, available at http://arxiv.org/abs/1008.2857v1.

$$\begin{aligned} R_{\text{sum}}^{\text{ZF}} &= \sum_{k=1}^{N} \left[\min\left\{ \left[\log\left(\frac{P_{\text{B}} \left| r_{i_k,i_k} \right|^2}{\|\mathbf{W}\|_{\text{F}}^2} \right) \right]^+, \log\left(1 + \frac{\left| l_{q_k,q_k} \right|^2 P_{\text{R}}}{N}\right) \right\} + \min\left\{ \left[\log\left(\left| r_{i_k,i_k} \right|^2 P_{\text{M},k} \right) \right]^+, \log\left(1 + \frac{\left| \tilde{l}_{q_k,q_k} \right|^2 P_{\text{R}}}{N}\right) \right\} \right] \\ &= \sum_{k=1}^{N} \log\left(1 + \left| l_{q_k,q_k} \right|^2 P_{\text{R}}/N\right) + \sum_{k=1}^{N} \left[\log\left(\left| r_{i_k,i_k} \right|^2 P_{\text{M},k} \right) \right]^+ \end{aligned} \tag{94}$$

$$\geq \log\left(\prod_{k=1}^{N} \left| l_{q_k,q_k} \right|^2 P_{\text{R}}/N \right) + \log\left(\prod_{k=1}^{N} \left| r_{i_k,i_k} \right|^2 P_{\text{M},k} \right) = \log\det\left(\frac{P_{\text{R}}}{N} \mathbf{H}_{\text{RM}} {\mathbf{H}_{\text{RM}}}^{\text{H}} \right) + \log\det\left(\mathbf{\Omega}_{\text{M}} \mathbf{H}_{\text{MR}} {\mathbf{H}_{\text{MR}}}^{\text{H}} \right) \tag{95}$$

$$= \sum_{k=1}^{N} \log\left({\sigma_{\text{RM},k}}^2 P_{\text{R}}/N \right) + \sum_{k=1}^{N} \log\left({\sigma_{\text{MR},k}}^2 P_{\text{M},k} \right) \tag{96}$$

[27] A. S. Avestimehr, M. A. Khajehnejad, A. Sezgin, and B. Hassibi, "Capacity region of the deterministic multi-pair bi-directional relay network," in *Proc. IEEE ITW*, Volos, Greece, June 2009.
[28] M. Chen and A. Yener, "Multiuser two-way relaying: detection and interference management strategies," *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 4296–4305, Aug. 2009.
[29] J. Joung and A. H. 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.
[30] E. Yılmaz, R. Zakhour, D. Gesbert, and R. Knopp, "Multi-pair two-way relay channel with multiple antenna relay station," in *Proc. IEEE ICC*, Cape Town, South Africa, May 2010.
[31] D. Gündüz, A. Yener, A. Goldsmith, and H. V. Poor, "The multi-way relay channel," in *Proc. IEEE ITW*, Volos, Greece, June 2009.
[32] L. Ong, C. M. Kellett, and S. J. Johnson, "Capacity theorems for the AWGN multi-way relay channel," in *Proc. IEEE ISIT*, Austin, Texas, June 2010, pp. 664–668.
[33] L. Zheng and D. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.
[34] J. Zhang, F. Roemer, and M. Haardt, "Beamforming design for multi-user two-way relaying with MIMO amplify and forward relays," in *Proc. IEEE ICASSP*, Prague, Czech Republic, May 2011.
[35] G. Peng, X. Jie, and Q. Ling, "Zero-forcing dirty paper coding aided physical-layer network coding for MIMO two-way relaying channels with multiple users," *The Journal of China Universities of Posts and Telecommunications (Elsevier)*, vol. 18, no. 2, pp. 53–58, Apr. 2011.
[36] H. J. Yang, Y. Choi, and J. Chun, "Modified high-order PAMs for binary coded physical-layer network coding," *IEEE Commun. Lett.*, vol. 14, no. 8, pp. 689–691, Aug. 2010.
[37] A. Edelman, "Eigenvalues and condition numbers of random matrices," Ph.D. dissertation, Messachusetts Institute of Technology, 1989.
[38] G. Caire and S. Shamai, "On the achievable throughput of a multi-antenna Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1691–1706, July 2003.
[39] H.-A. Loeliger, "Averaging bounds for lattices and linear codes," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1767–1773, Nov. 1997.
[40] U. Erez and S. t. Brink, "A close-to-capacity dirty paper coding scheme," *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3417–3432, Oct. 2005.
[41] M. Chiang, "Geometric programming for communication systems," *Short monograph in Foundations and Trends in Commun. and Inf. Theory*, vol. 2, no. 1-2, pp. 1–154, Aug. 2005.
[42] A. D. Dabbagh and D. J. Love, "Precoding for multiple antenna Gaussian broadcast channels with successive zero-forcing," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3837–3850, July 2007.
[43] C. A. Rogers, *Packing and Covering*. Cambridge, U.k.: Cambridge Univ. Press, 1964.
[44] U. Erez, R. Zamir, and S. Litsyn, "Lattices which are good for (almost) everything," in *Proc. IEEE ITW*, Paris, France, Apr.-May 2003, pp. 271–274.
[45] G. D. Forney, Jr., "Coset codes-part 1: Introduction and geometrical classification," *IEEE Trans. Inf. Theory*, vol. 34, no. 5, pp. 1123–1151, Sept. 1988.
[46] G. D. Forney, Jr., M. D. Trott, and S.-Y. Chung, "Sphere-bound-achieving coset codes and multilevel coset codes," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 820–850, May 2000.
[47] G. D. Forney, Jr. and G. Ungerboeck, "Modulation and codings for linear Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2384–2415, Oct. 1998.
[48] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *Journal of the American Statistical Association*, vol. 58, no. 301, pp. 13–30, Mar. 1963.
[49] T. Liu, P. Moulin, and R. Koetter, "On error exponents of modulo lattice additive noise channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 454–471, Feb. 2006.
[50] E. Martinian, "Waterfilling gains O(1/SNR) at high SNR," *Unpublished note*, available at http://www.csua.berkeley.edu/ emin/research/wfill.pdf.

Hyun Jong Yang received the B.S. degree in electrical engineering from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea, in 2004, and the M.S. and Ph.D. degrees in electrical engineering from KAIST, in 2006 and 2010, respectively. From Aug. 2010 to Aug. 2011, he was a research fellow at Korea Ocean Research & Development Institute, Daejoen, Republic of Korea. Since Oct. 2011, he has been a post-doctoral researcher in the Electrical Engineering Department, Stanford University, Stanford, CA. His fields of interests are MIMO information theory, relay communication theory, and limited feedback theory.

Youngchol Choi received the B.S., M.S., and Ph.D. degrees in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 1998, 2000, and 2011, respectively. Since 2000, he has been with the Korea Institute of Ocean Science and Technology (KIOST), Daejeon, Korea, where he is a Senior Research Scientist. His research interests include MIMO systems, two-phase relaying systems, ad-hoc network protocols for underwater acoustic communications, and inverse scattering theory.

Namyoon Lee received the B.S. degree in radio and communication engineering from Korea University, Seoul, Korea, in 2006 and the M.S. degree in electrical engineering from the KAIST, Daejeon, Korea, in 2008. From 2008 to 2011, he was a member of technical staff at Samsung Advanced Institute of Technology (SAIT) and Samsung Electronics Co. Ltd. in Korea, where he investigated next generation device-to-device (D2D) wireless communication systems and involved standardization activities of the 3GPP LTE-Adv., especially for femto-cell deployment. He is currently a Ph.D. student at the University of Texas at Austin. His current research interests are multiuser and multiway-communication theory using interference alignment, neutralization, and network coding. Mr. Lee was a recipient of the 2009 Samsung Best Paper Award. He was also awarded several fellowships, including the Graduate Student Research Fellowship from the Korea Science and Engineering Foundation (KOSEF) in 2006; the Korea Government Fellowship from 2006 to 2007; the Kwanjeong Educational Foundation Fellowship in 2011.

Arogyaswami Paulraj received the B.E. degree from Naval College of Engineering, Lonavala, India, in 1966 and the Ph.D. degree from the Indian Institute of Technology, Delhi, India, in 1973. He is a pioneer of MIMO wireless communications. He joined Stanford University, Stanford, CA, in 1972, after a 20-year industry/military career in India. He founded two successful semiconductor companies - Iospan Wireless acquired by Intel Corp. in 2003 and Beceem Communications acquired by Broadcom Corp. in 2010. He has authored over 400 research publications and two text books, and is a co-inventor in 52 US patents. Dr. Paulraj is a member of the U.S. National Academy of Engineering, the Royal Swedish Academy of Engineering Sciences, the Academy of Sciences for the Developing World (TWAS), and the Indian National Academy of Engineering. He was the recipient of a number of national awards in India for sonar development and of several IEEE awards, including the Technical Achievement Award from the Signal Processing Society, fellow grade, and several best paper awards. He was awarded the 2011 IEEE Alexander Graham Bell Medal for his work in MIMO.