Robust Beamforming Designs for Nonregenerative Multipair Two-Way Relaying Systems

Changick Song, Haewook Park, Hoon Lee, and Inkyu Lee, *Senior Member, IEEE*

***Abstract*—In this paper, we consider nonregenerative multipair two-way relaying systems, where a relay node supports K pairs of two-way communications in the presence of imperfect channel state information (CSI). We employ a stochastic approach to model the channel uncertainties and study a robust beamforming design at the relay. It is shown that the proposed beamformer provides a reduction in the computational complexity, as well as good robustness against channel errors, compared with conventional designs. To reduce the estimation overhead at the users, we also suggest an efficient downlink training method to inform the users of the channel-dependent self-interference cancellation (SIC) parameters. With the training method, the additive noise may incur a malfunction of the SIC. To address this issue, we develop an intelligent selection criterion that decides whether the SIC should be adopted or not. Finally, from simulation results, we demonstrate the efficiency of our proposed schemes.**

***Index Terms*—Imperfect channel-state information (CSI), minimum mean squared error (MMSE) beamforming, multi-antenna, multipair two-way relay, self interference cancellation.**

I. Introduction

Recently, relay-assisted cooperative networks have garnered a lot of interest due to such advantages as extended cell coverage and improved reliability [1]–[5]. In particular, two-way relay systems, in which a bidirectional communication occurs between two users with the help of relay(s), have been an active research area over the past few years [6]–[8]. Unlike one-way relaying, which requires four time slots to exchange information due to the half-duplex constraint, the communication in the two-way relaying system can be made in two time slots.

Depending on forwarding strategies, the two-way system may adopt either nonregenerative or regenerative relays [6]. The regenerative relay has a merit on channel estimation, because channels in the multiple-access channel (MAC) and broadcast channel (BC) phases are separated due to the decoding process at the relay. In practice, however, the nonregenerative relays, which do not perform data decoding, are more attractive, particularly when there is a complexity constraint at the relay. Thus, in this paper, we consider the nonregenerative relays. One major feature of the two-way nonregenerative systems is the self-interference, because a signal transmitted by one user may return to the same user. Since each user knows its own transmitted signal, the self-interference may be removed as long as proper channel state information (CSI) is known. Unlike the regenerative relays, however, it is not easy for the users to obtain the CSI, since the effective channel gain embraces all channels in the MAC phase, as well as its own channel in the BC phase [9].

In recent years, to further enhance the network throughput, two-way relaying, which serves a single pair of users, has been extended to multipair two-way (MPTW) systems, where multiple user pairs exchange messages using shared relay(s) [10]–[15]. In this scenario, simultaneously transmitted signals from multiple users incur interference, which degrades the performance of the system. One way to control the interference is to employ the spatial division multiple access (SDMA) utilizing multiantennas at the relay [10]. However, the SDMA scheme in MPTW suffers from a rate loss, because the relay spends unnecessary energy to remove all interference, including the self-interference that can be managed by users. To take the self-interference cancellation (SIC) into account, some zero-forcing (ZF) and minimum mean squared error (MMSE) beamforming designs have been proposed based on the block diagonalization (BD) technique, focusing on interpair interference rather than intrapair interference [11]–[15]. However, none of these works addressed the practical imperfect CSI scenarios that may seriously affect performance.

This paper contains two major contributions. In the first part of this paper, we present a new MMSE-BD beamforming design in MPTW relaying systems, taking both the SIC and the imperfect CSI into account. Whereas the conventional robust designs in relaying systems normally assume the same level of channel uncertainties at both the relay and the users [16], our method provides a flexible design that accounts for different channel uncertainties between the relay and the users and, thus, is practically more important. In particular, since we find the relay beamformer as a function of SIC parameters that will be used to subtract the self-interference at users, we are allowed to adjust the beamformer under various CSI circumstances of the users. It is also shown that our beamforming method achieves complexity gain from $\mathcal{O}(N_r^6)$ to $\mathcal{O}(N_r^3)$ compared with the conventional methods in [7], [15], and [16].

In the second part of this paper, we investigate an efficient training method to inform the users of the CSI being used for the SIC. We first show that the optimal SIC parameter can be approximated to the effective precoded channel in the BC phase, which gives rise to a significant reduction in the size of the training sequence. With the training method, however, the additive noise may incur malfunction of the SIC, i.e., the SIC operation may degrade the performance rather than improve it. To avoid such a loss, we develop an intelligent SIC mode selection (SMS) criterion that decides whether the SIC should be adopted or not. It is shown that the SMS scheme provides good robustness against channel errors at the users. Finally, simulation results demonstrate the efficiency of the proposed designs.

Notations: Normal letters represent scalar quantities, boldface letters indicate vectors, and boldface uppercase letters designate matrices. The superscripts $(\cdot)^T$, $(\cdot)^H$, and $(\cdot)^*$ stand for transpose, conjugate transpose, and element-wise conjugate, respectively. In addition, $\mathbf{I}_N$ is defined by an $N \times N$ identity matrix, and $E[\cdot]$ denotes the expectation operator. $\mathrm{Tr}(\mathbf{A})$ and $[\mathbf{A}]_{i,j}$ indicate the trace and the (i,j)th element of $\mathbf{A}$, respectively. The notation $\mathrm{diag}\{\cdot\}$ stands for a diagonal matrix with a vector on its diagonal.

II. System Model

As shown in Fig. 1, we consider an MPTW nonregenerative relay channel where a relay with N_r antennas helps in the communication of $2K$ mobile users having a single antenna. Here, $2K$ users result in K pairs of two users each performing two-way communications. It is

Manuscript received April 9, 2013; revised April 17, 2015 and July 8, 2015; accepted September 19, 2015. Date of publication October 1, 2015; date of current version September 15, 2016. This work was supported in part by the National Research Foundation of Korea funded by the Korean government (Ministry of Science, ICT, and Future Planning) under Grant 2014R1A2A1A10049769 and Grant 2015R1C1A1A02036927, by Korea Electric Power Corporation through Korea Electrical Engineering and Science Research Institute under Grant R14XA02-48, and by Korea National University of Transportation in 2015. The review of this paper was coordinated by Prof. M. Uysal.

C. Song was with the School of Electrical Engineering, Korea University, Seoul 136-713, Korea. He is now with the Department of Information and Communications Engineering, Korea National University of Transportation, Chungju 274 69, Korea (e-mail: c.song@ut.ac.kr).

H. Park is with LG Electronics, Seoul 150-721, Korea (e-mail: haewook.park@lge.com).

H. Lee and I. Lee are with the School of Electrical Engineering, Korea University, Seoul 151-744, Korea (e-mail: ihun1@korea.ac.kr; inkyu@korea.ac.kr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2015.2485498

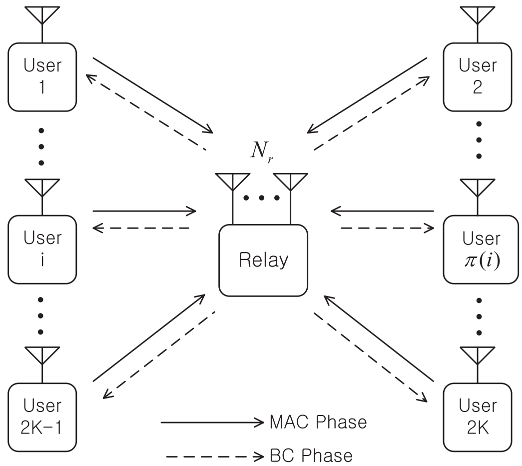

Fig. 1. System model for K-pair two-way nonregenerative relaying systems.

assumed that user i exchanges a message with user $\pi(i)$ using two consecutive time slots, where $\pi(i)$ is an index function indicating the partner of user i. Thus, if $i = 2k-1$, we have $\pi(i) = 2k$, and *vice versa*, for the kth user pair $k \in \{1, \ldots, K\}$. We denote the channel vectors from the ith user to the relay and from the relay to the ith user by $\mathbf{h}_i \in \mathbb{C}^{N_r \times 1}$ and $\mathbf{g}_i^T \in \mathbb{C}^{1 \times N_r}$, respectively, for $i = 1, \ldots, 2K$.

The relay operates in a half-duplex mode, and thus, the information exchange between user pairs occurs in two orthogonal phases. In the MAC phase, the ith user transmits signal x_i to the relay for all i. Then, the received signal at the relay is written by

$$\mathbf{y}_r = \mathbf{H}\mathbf{x} + \mathbf{n}_r$$

where $\mathbf{x} \triangleq [x_1, \ldots, x_{2K}]^T \in \mathbb{C}^{2K \times 1}$, $\mathbf{H} \triangleq [\mathbf{h}_1, \ldots, \mathbf{h}_{2K}] \in \mathbb{C}^{N_r \times 2K}$, and $\mathbf{n}_r \in \mathbb{C}^{N_r \times 1} \sim \mathcal{CN}(0, \mathbf{I}_{N_r})$ stands for the input signal vector with covariance $\mathbf{R}_x \triangleq E[\mathbf{x}\mathbf{x}^H] = P_s \mathbf{I}_{2K}$, the MAC phase channel matrix between the relay and users, and the circularly symmetric complex Gaussian (CSCG) noise at the relay, respectively.

During the BC phase, the relay generates the transmit signal $\mathbf{Q}\mathbf{y}_r$ with the beamforming matrix $\mathbf{Q} \in \mathcal{C}^{N_r \times N_r}$ and transmits to all users. In this case, $\mathbf{Q}$ must be designed to satisfy the relay power budget P_r as $\mathrm{Tr}(\mathbf{Q}\mathbf{R}_{\mathrm{yr}}\mathbf{Q}^H) \leq P_r$, where $\mathbf{R}_{\mathrm{yr}} \triangleq E[\mathbf{y}_r \mathbf{y}_r^H]$. Then, the received signal y_i at user i becomes

$$y_i = \frac{1}{\gamma}\left(\mathbf{g}_i^T \mathbf{Q}\mathbf{h}_{\pi(i)} x_{\pi(i)} + \mathbf{g}_i^T \mathbf{Q}\mathbf{h}_i x_i + \sum_{j \neq \{i, \pi(i)\}}^{2K} \mathbf{g}_i^T \mathbf{Q}\mathbf{h}_j x_j + \mathbf{g}_i^T \mathbf{Q}\mathbf{n}_r + n_{u_i}\right) \quad (1)$$

where $n_{u_i} \sim \mathcal{CN}(0, 1)$ denotes the noise at the ith user, and γ indicates the automatic gain control [7] being used to mitigate the user noise n_{u_i}.

The first and second terms in (1) indicate the desired signal for user i and the backpropagated self-interference, respectively. Since each user knows its own signal transmitted in the previous phase, it is often assumed that the self-interference can be removed. In practical systems, however, achieving perfect SIC is a hard problem due to channel estimation errors. Thus, the residual self-interference after the SIC may not be ignorable. To take the imperfect SIC into consideration, we newly define an SIC parameter α_i that will be used to subtract the self-interference at user i. Then, we obtain the final observation signal at user i as

$$\tilde{y}_i = \frac{1}{\gamma}\left(\mathbf{g}_i^T \mathbf{Q}\mathbf{h}_{\pi(i)} x_{\pi(i)} + \sum_{j \neq \{i, \pi(i)\}}^{2K} \mathbf{g}_i^T \mathbf{Q}\mathbf{h}_j x_j + \mathbf{g}_i^T \mathbf{Q}\mathbf{n}_r + n_{u_i}\right) + \left(\frac{1}{\gamma}\mathbf{g}_i^T \mathbf{Q}\mathbf{h}_i - \alpha_i\right) x_i \quad (2)$$

where the last term in (2) represents the residual self-interference after the SIC.

From (2), the signal-to-noise-plus-interference ratio (SINR) is computed by (3), shown at the bottom of the page, based on which we have the sum-rate maximization problem

$$\{\mathbf{Q}, \gamma, \alpha_i \ \ \forall i\} = \arg \max_{\mathbf{Q}, \gamma, \alpha_i \ \forall i} \sum_{i=1}^{2K} \frac{\zeta}{2} \log_2(1 + \mathrm{SINR}_i)$$
$$\text{s.t.} \quad \mathrm{Tr}(\mathbf{Q}\mathbf{R}_{\mathrm{yr}}\mathbf{Q}^H) \leq P_r. \quad (4)$$

Here, $\zeta = \max[1 - t/T, 0]$ denotes the spectral efficiency penalty factor, where T stands for the total number of time slots in each coherence block in which a training sequence occupies t time slots for channel estimation. The pre-log factor 1/2 is due to the half-duplex constraint of the relay.

III. Relay Beamforming Designs

Since problem (4) is neither convex nor concave with respect to the relay matrix $\mathbf{Q}$, a direct rate optimization may require high computational cost, even for finding a local optimal solution [13]. To tackle the problem, traditionally, low-complexity alternatives such as ZF and MMSE methods have been widely adopted in wireless communications. In the case of MPTW systems, the individual users can take care of the backpropagated self-interference, and thus, the relay needs only to eliminate the interpair interference rather than the intrapair interference. A natural way to take this into consideration is to exploit the BD techniques [17]. Some ZF-based BD schemes have been proposed in [11]–[14]. However, their performance is rather poor due to the transmit power boost issue at the relay. Here, we propose a new relay beamforming design based on the MMSE-based BD method. Our solution guarantees optimality with respect to minimizing the weighted sum of mean square error (MSE) and is also robust against channel errors at the relay.

We start with rewriting the received signal (1) in a vector form as

$$\mathbf{y} = \frac{1}{\gamma}(\mathbf{G}\mathbf{Q}\mathbf{H}\mathbf{x} + \mathbf{G}\mathbf{Q}\mathbf{n}_r + \mathbf{n}_u) \quad (5)$$

where $\mathbf{y} = [y_1, \ldots, y_{2K}]^T$, $\mathbf{G} = [\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_{2K}]^T$, and $\mathbf{n}_u \triangleq [n_{u_1}, \ldots, n_{u_{2K}}]^T$. We remind that the input signal $x_{\pi(i)}$ is desired

$$\mathrm{SINR}_i = \frac{P_s \left|\mathbf{g}_i^T \mathbf{Q}\mathbf{h}_{\pi(i)}\right|^2}{P_s \left(\left|\mathbf{g}_i^T \mathbf{Q}\mathbf{h}_i - \gamma\alpha_i\right|^2 + \sum_{j \neq \{i, \pi(i)\}}^{2K} \left|\mathbf{g}_i^T \mathbf{Q}\mathbf{h}_j\right|^2\right) + \left\|\mathbf{g}_i^T \mathbf{Q}\right\|^2 + 1} \quad (3)$$

for the ith user, and the ith user will subtract its own self-interference using $\alpha_i x_i$. Thus, (5) suggests the use of $(\mathbf{U}+\mathbf{F})\mathbf{x}$ as a target vector of an MMSE problem, where $\mathbf{U} \triangleq \text{diag}\{\alpha_1, \ldots, \alpha_{2K}\} \in \mathbb{C}^{2K\times 2K}$ designates the SIC matrix and $\mathbf{F} \in \mathbb{C}^{2K\times 2K}$ indicates a switching matrix matching the ith user with the $\pi(i)$th user as

$$\mathbf{F} \triangleq \text{diag}\left\{\begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}, \ldots, \begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}\right\}.$$

Then, we formulate the weighted MSE optimization problem as

$$\min_{\gamma,\mathbf{Q},\mathbf{U}} E\left[\|\mathbf{e}\|^2\right] \quad \text{s.t.} \quad \text{Tr}(\mathbf{Q}\mathbf{R}_{\text{yr}}\mathbf{Q}^H) \le P_r \tag{6}$$

where $\mathbf{e} = \mathbf{y} - (\mathbf{U}+\mathbf{F})\mathbf{x}$. Since the weight matrix $(\mathbf{U}+\mathbf{F})$ has a 2×2 blockwise diagonal structure, this approach tends to spend more resources to eliminate the interpair interference rather than the intrapair interference, and thus, we call it the *MMSE-BD* scheme.

The channel estimation at the relay can be made by training or by feedback signals from the users, as in the multiuser uplink cellular systems. Since mounting multiple antennas closely on the relay node introduces channel correlation, the Kronecker model [16] is used to describe the true channels $\mathbf{H}$ and $\mathbf{G}$ as $\mathbf{H} = \mathbf{T}_h^{1/2}\mathbf{H}_w$ and $\mathbf{G} = \mathbf{G}_w\mathbf{T}_g^{1/2}$, where $\mathbf{T}_h$ and $\mathbf{T}_g$, all being positive semidefinite, are the row and column correlation matrices of $\mathbf{H}$ and $\mathbf{G}$, respectively. The relay estimates $\mathbf{H}_w$ and $\mathbf{G}_w$ with estimation noise $\Delta\mathbf{H}_w$ and $\Delta\mathbf{G}_w$, of which the entries are independent and identically distributed (i.i.d.) CSCG with zero mean and variances $\sigma_{e,h}^2$ and $\sigma_{e,g}^2$, respectively. For simplicity, we assumed that all users experience the same antenna correlation, but the result can easily be applied to the individual correlation cases.

Denoting the estimated channels at the relay by $\widehat{\mathbf{H}}_w$ and $\widehat{\mathbf{G}}_w$, it follows that $\mathbf{H}_w = \widehat{\mathbf{H}}_w + \Delta\mathbf{H}_w$ and $\mathbf{G}_w = \widehat{\mathbf{G}}_w + \Delta\mathbf{G}_w$. Thus, we can rewrite the true channels $\mathbf{H}$ and $\mathbf{G}$ as

$$\mathbf{H} = \widehat{\mathbf{H}} + \Delta\mathbf{H} \;\text{ and }\; \mathbf{G} = \widehat{\mathbf{G}} + \Delta\mathbf{G} \tag{7}$$

where $\widehat{\mathbf{H}} = \mathbf{T}_h^{1/2}\widehat{\mathbf{H}}_w$, $\widehat{\mathbf{G}} = \widehat{\mathbf{G}}_w\mathbf{T}_g^{1/2}$, $\Delta\mathbf{H} \sim \mathcal{CN}(\mathbf{0}_{2K}, \sigma_{e,h}^2\mathbf{I}_{2K}\otimes\mathbf{T}_h)$, and $\Delta\mathbf{G} \sim \mathcal{CN}(\mathbf{0}_{2K}, \sigma_{e,g}^2\mathbf{T}_g^T\otimes\mathbf{I}_{2K})$. With our estimation error model, the expectation in (6) is taken over $\mathbf{x}$, $\mathbf{n}_r$, $\mathbf{n}_u$, $\Delta\mathbf{H}$, and $\Delta\mathbf{G}$. Then, the MSE is equivalently $E[\|\mathbf{e}\|^2] = \text{Tr}(\mathbf{R}_e)$, where[1]

$$\begin{aligned}\mathbf{R}_e &\triangleq E[\mathbf{e}\mathbf{e}^H]\\
&= \gamma^{-2}\left(\widehat{\mathbf{G}}\mathbf{Q}\widehat{\mathbf{H}}\mathbf{R}_x\widehat{\mathbf{H}}^H\mathbf{Q}^H\widehat{\mathbf{G}}^H + \widehat{\mathbf{G}}\mathbf{Q}\mathbf{Q}^H\widehat{\mathbf{G}}^H + \mathbf{I}_{2K}\right)\\
&\quad - 2\gamma^{-1}\Re\left\{\widehat{\mathbf{G}}\mathbf{Q}\widehat{\mathbf{H}}\mathbf{R}_x(\mathbf{U}+\mathbf{F})^H\right\} + (\mathbf{U}+\mathbf{F})\mathbf{R}_x(\mathbf{U}+\mathbf{F})^H\\
&\quad + \sigma_{e,h}^2\text{Tr}(\mathbf{R}_x)\widehat{\mathbf{G}}\mathbf{Q}\mathbf{T}_h\mathbf{Q}^H\widehat{\mathbf{G}}^H\\
&\quad + \sigma_{e,g}^2\text{Tr}\left(\mathbf{Q}\widehat{\mathbf{H}}\mathbf{R}_x\widehat{\mathbf{H}}^H\mathbf{Q}^H\mathbf{T}_g\right)\mathbf{I}_{2K}\\
&\quad + \sigma_{e,h}^2\sigma_{e,g}^2\text{Tr}(\mathbf{R}_x)\text{Tr}(\mathbf{Q}\mathbf{T}_h\mathbf{Q}^H\mathbf{T}_g)\mathbf{I}_{2K}\\
&\quad + \sigma_{e,g}^2\text{Tr}(\mathbf{Q}\mathbf{Q}^H\mathbf{T}_g)\mathbf{I}_{2K}.\end{aligned}$$

Let us first solve problem (6) in terms of γ and $\mathbf{Q}$ for given $\mathbf{U}$. Using the Lagrange multiplier μ, the cost function is set up by $\mathcal{C} = \text{Tr}(\mathbf{R}_e(\gamma,\mathbf{Q},\mathbf{U})) + \mu\text{Tr}(\mathbf{Q}\mathbf{R}_{\text{yr}}\mathbf{Q}^H)$, where

$$\mathbf{R}_{\text{yr}} = E\left[\mathbf{y}_r\mathbf{y}_r^H\right] = \widehat{\mathbf{H}}\mathbf{R}_x\widehat{\mathbf{H}}^H + \mathbf{R}_{n,h}$$

[1]Here, we have used the fact that $E[\Delta\mathbf{H}\mathbf{M}\Delta\mathbf{H}^H] = \text{Tr}(\mathbf{M}\mathbf{T})\mathbf{R}$ for an arbitrary matrix $\mathbf{M}$ with $\Delta\mathbf{H} \sim \mathcal{CN}(\mathbf{0}, \mathbf{T}^T\otimes\mathbf{R})$.

with $\mathbf{R}_{n,h} = \sigma_{e,h}^2\text{Tr}(\mathbf{R}_x)\mathbf{T}_h + \mathbf{I}_{N_r}$ representing the effective noise covariance in the MAC phase. Then, the Karush–Kuhn–Tucker conditions are given by

$$\frac{\partial\mathcal{C}}{\partial\mathbf{Q}^*} = 0,\; \frac{\partial\mathcal{C}}{\partial\gamma} = 0, \quad \mu \ge 0, \quad \text{Tr}(\mathbf{Q}\mathbf{R}_{\text{yr}}\mathbf{Q}^H) \le P_R$$
$$\text{and } \mu\left(\text{Tr}(\mathbf{Q}\mathbf{R}_{\text{yr}}\mathbf{Q}^H) - P_R\right) = 0. \tag{8}$$

From the first two zero-gradient conditions, we have the following two equations:

$$\begin{aligned}\mathbf{Q}_{\text{opt}} &= \gamma\left(\widehat{\mathbf{G}}^H\widehat{\mathbf{G}} + 2K\sigma_{e,g}^2\mathbf{T}_g + \mu\gamma^2\mathbf{I}_{N_r}\right)^{-1}\widehat{\mathbf{G}}^H\\
&\quad\times(\mathbf{U}+\mathbf{F})\mathbf{R}_x\widehat{\mathbf{H}}^H\mathbf{R}_{\text{yr}}^{-1}\end{aligned} \tag{9}$$

$$\begin{aligned}&\text{Tr}\left(\widehat{\mathbf{G}}\mathbf{Q}\mathbf{R}_{\text{yr}}\mathbf{Q}^H\widehat{\mathbf{G}}^H + 2K\sigma_{e,g}^2\mathbf{Q}\mathbf{R}_{\text{yr}}\mathbf{Q}^H\mathbf{T}_g + \mathbf{I}_{2K}\right)\\
&\quad= \hat{\gamma}\left(\text{Tr}(\mathbf{X}) + \text{Tr}(\mathbf{X}^H)\right)\end{aligned} \tag{10}$$

where $\mathbf{X} \triangleq (\mathbf{U}+\mathbf{F})\mathbf{R}_x\widehat{\mathbf{H}}^H\mathbf{Q}^H\widehat{\mathbf{G}}^H$.

In addition, by applying (9) to (10), $\gamma\text{Tr}(\mathbf{X})$ in the right-hand side of (10) can be further manipulated as

$$\begin{aligned}&\gamma\text{Tr}\left((\mathbf{U}+\mathbf{F})\mathbf{R}_x\widehat{\mathbf{H}}^H\mathbf{Q}_{\text{opt}}^H\widehat{\mathbf{G}}^H\right)\\
&\quad= \text{Tr}\left(\gamma\mathbf{G}^H(\mathbf{U}+\mathbf{F})\mathbf{R}_x\mathbf{H}^H\mathbf{Q}_{\text{opt}}^H\right)\\
&\quad= \text{Tr}\left(\left(\mathbf{G}^H\mathbf{G} + 2\sigma_{e,g}^2\mathbf{T}_g + \mu\gamma^2\mathbf{I}_{N_r}\right)\mathbf{Q}_{\text{opt}}\mathbf{R}_{\text{yr}}\mathbf{Q}_{\text{opt}}^H\right)\\
&\quad= \text{Tr}\left(\left(\mathbf{G}^H\mathbf{G} + 2\sigma_{e,g}^2\mathbf{T}_g\right)\left(\mathbf{Q}_{\text{opt}}\mathbf{R}_{\text{yr}}\mathbf{Q}_{\text{opt}}^H\right)\right)\\
&\qquad + \mu\gamma^2\text{Tr}\left(\mathbf{Q}_{\text{opt}}\mathbf{R}_{\text{yr}}\mathbf{Q}_{\text{opt}}^H\right).\end{aligned} \tag{11}$$

We first see from (11) that the equality $\text{Tr}(\mathbf{X}) = \text{Tr}(\mathbf{X}^H)$ holds. Thus, combining (10) and (11), and considering the relay power constraint $\text{Tr}(\mathbf{Q}\mathbf{R}_{\text{yr}}\mathbf{Q}^H) \le P_r$, we obtain $\mu\gamma^2 = (2K/\text{Tr}(\mathbf{Q}\mathbf{R}_{\text{yr}}\mathbf{Q}^H)) \ge (2K/P_r)$, which implies that $\mu > 0$. Therefore, from the slackness condition $\mu(\text{Tr}(\mathbf{Q}\mathbf{R}_{\text{yr}}\mathbf{Q}^H) - P_r) = 0$ in (8), we arrive at $\mu\hat{\gamma}^2 = (2K/P_r)$ and $\gamma = \sqrt{P_r/\text{Tr}(\widetilde{\mathbf{Q}}\mathbf{R}_{\text{yr}}\widetilde{\mathbf{Q}}^H)}$, where

$$\widetilde{\mathbf{Q}} = \left(\widehat{\mathbf{G}}^H\widehat{\mathbf{G}} + 2K\sigma_{e,g}^2\mathbf{T}_g + \frac{2K}{P_r}\mathbf{I}_{N_r}\right)^{-1}\widehat{\mathbf{G}}^H\times(\mathbf{U}+\mathbf{F})\mathbf{R}_x\widehat{\mathbf{H}}^H\mathbf{R}_{\text{yr}}^{-1}.$$

Finally, we obtain the optimal relay beamformer as

$$\mathbf{Q}_{\text{opt}} = \gamma\mathbf{B}(\mathbf{U}+\mathbf{F})\mathbf{L} \tag{12}$$

where $\mathbf{B} = (\widehat{\mathbf{G}}^H\widehat{\mathbf{G}} + 2K\sigma_{e,g}^2\mathbf{T}_g + (2K/P_r)\mathbf{I}_{N_r})^{-1}\widehat{\mathbf{G}}^H$, and $\mathbf{L} = \mathbf{R}_x\widehat{\mathbf{H}}^H\mathbf{R}_{\text{yr}}^{-1}$. It is seen from (12) that $\mathbf{L}$ and $\mathbf{B}$ form the transmit and receive MMSE filter structures for the BC phase channel $\mathbf{G}$ and the MAC phase channel $\mathbf{H}$, respectively, on either side of $(\mathbf{U}+\mathbf{F})$. Therefore, as SNR increases, the resulting effective channel in the BC phase will have a 2×2 blockwise diagonal structure. The automatic gain control factor γ plays a role of power normalization at the relay.

As proper solutions of γ and $\mathbf{Q}$ are found in (12), the remaining work is to identify the SIC parameters in $\mathbf{U}$. To this end, one may substitute (12) into problem (6) and solve it again with respect to $\mathbf{U}$. However, a direct substitution leads to a complicated problem that is not easy to compute. In what follows, we show that when γ and $\mathbf{Q}$ are given in the form of (12), the optimization problem of $\mathbf{U}$ can be formulated as a simple convex problem.

Let us define $\overline{\mathbf{y}}_r \triangleq (\mathbf{U}+\mathbf{F})\tilde{\mathbf{y}}_r$, where $\widetilde{\mathbf{y}}_r = \mathbf{L}\mathbf{y}_r$. Then, it is seen that since $(\mathbf{U}+\mathbf{F})\mathbf{L}$ amounts to the Wiener filter for the virtual input signal $(\mathbf{U}+\mathbf{F})\mathbf{x}$, its output signal $\overline{\mathbf{y}}_r$ satisfies the orthogonality principle $E[\overline{\mathbf{y}}_r^H(\overline{\mathbf{y}}_r - (\mathbf{U}+\mathbf{F})\mathbf{x})] = 0$. Now, using $\overline{\mathbf{y}}_r$, the MSE in

(6) is equivalently $E[\|\mathbf{e}\|^2] = E[\|\mathbf{y} - \overline{\mathbf{y}}_r + \overline{\mathbf{y}}_r - (\mathbf{U}+\mathbf{F})\mathbf{x}\|^2]$. Due to the orthogonality principle again, we check that the signal $\overline{\mathbf{y}}_r - (\mathbf{U}+\mathbf{F})\mathbf{x}$ becomes orthogonal to $\mathbf{y}$ and $\overline{\mathbf{y}}_r$, since $\mathbf{y} = \gamma\mathbf{G}\mathbf{B}\overline{\mathbf{y}}_r + \mathbf{n}_u$ is also a function of $\overline{\mathbf{y}}_r$ and the independent noise $\mathbf{n}_u$ [see (5)]. Therefore, the overall MSE in (6) is equivalently given by a sum of two MSE terms, each of which corresponding to the MAC and BC phases as

$$E\left[\|\mathbf{e}\|^2\right] = \mathrm{MSE}_{\mathrm{BC}} + \mathrm{MSE}_{\mathrm{MAC}} \tag{13}$$

where $\mathrm{MSE}_{\mathrm{BC}} \triangleq E[\|\mathbf{y} - \overline{\mathbf{y}}_r\|^2]$, and $\mathrm{MSE}_{\mathrm{MAC}} \triangleq E[\|\overline{\mathbf{y}}_r - (\mathbf{U}+\mathbf{F})\mathbf{x}\|^2]$.

Let us further evaluate $\mathrm{MSE}_{\mathrm{BC}}$ and $\mathrm{MSE}_{\mathrm{MAC}}$ in (13) in matrix forms. First, we have

$$\begin{aligned}\mathrm{MSE}_{\mathrm{BC}} &= E\left[\left\|\left(\mathbf{G}\mathbf{B}\overline{\mathbf{y}}_r + \frac{1}{\gamma}\mathbf{n}_u\right) - \overline{\mathbf{y}}_r\right\|^2\right]\\ &= \mathrm{Tr}\left(\widehat{\mathbf{G}}\mathbf{B}\overline{\mathbf{R}}_{\mathrm{yr}}\mathbf{B}^H\widehat{\mathbf{G}}^H + 2K\sigma_{e,g}^2\mathbf{B}\overline{\mathbf{R}}_{\mathrm{yr}}\mathbf{B}^H\mathbf{T}_g + \frac{1}{\gamma^2}\mathbf{I}_{2K}\right.\\ &\qquad\left. - \widehat{\mathbf{G}}\mathbf{B}\overline{\mathbf{R}}_{\mathrm{yr}} - \overline{\mathbf{R}}_{\mathrm{yr}}\mathbf{B}^H\widehat{\mathbf{G}}^H + \overline{\mathbf{R}}_{\mathrm{yr}}\right)\end{aligned}$$

where $\overline{\mathbf{R}}_{\mathrm{yr}} \triangleq E[\overline{\mathbf{y}}_r\overline{\mathbf{y}}_r^H] = (\mathbf{U}+\mathbf{F})\widetilde{\mathbf{R}}_{\mathrm{yr}}(\mathbf{U}+\mathbf{F})^H$, and

$$\begin{aligned}\widetilde{\mathbf{R}}_{\mathrm{yr}} &= \mathbf{R}_x\widehat{\mathbf{H}}^H\mathbf{R}_{\mathrm{yr}}^{-1}\mathbf{H}\mathbf{R}_x\\ &= \mathbf{R}_x\mathbf{H}^H\mathbf{R}_{n,h}^{-1}\mathbf{H}\left(\mathbf{H}^H\mathbf{R}_{n,h}^{-1}\mathbf{H} + \mathbf{R}_x^{-1}\right)^{-1}.\end{aligned} \tag{14}$$

From (10), we have $\mathbf{G}\mathbf{B}\overline{\mathbf{R}}_{\mathrm{yr}}\mathbf{B}^H\mathbf{G}^H + 2K\sigma_{e,g}^2\mathbf{B}\overline{\mathbf{R}}_{\mathrm{yr}}\mathbf{B}^H\mathbf{T}_g + (1/\gamma^2)\mathbf{I}_{2K} = \mathbf{G}\mathbf{B}\overline{\mathbf{R}}_{\mathrm{yr}}$. Thus, applying the relay precoder $\mathbf{B}$ in (12) and invoking some matrix inversion lemma, it follows that

$$\begin{aligned}\mathrm{MSE}_{\mathrm{BC}} &= \mathrm{Tr}\left(\overline{\mathbf{R}}_{\mathrm{yr}}\left(\mathbf{I} - \mathbf{B}^H\widehat{\mathbf{G}}^H\right)\right)\\ &= \mathrm{Tr}\left(\overline{\mathbf{R}}_{\mathrm{yr}}\left(\mathbf{I} - \widehat{\mathbf{G}}\left(\widehat{\mathbf{G}}^H\widehat{\mathbf{G}} + 2K\sigma_{e,g}^2\mathbf{T}_g + \frac{2K}{P_r}\mathbf{I}_{N_r}\right)^{-1}\widehat{\mathbf{G}}^H\right)\right)\\ &= \mathrm{Tr}\left(\frac{2K}{P_r}\overline{\mathbf{R}}_{\mathrm{yr}}\left(\widehat{\mathbf{G}}\mathbf{R}_{n,g}^{-1}\widehat{\mathbf{G}}^H + \frac{2K}{P_r}\mathbf{I}_{2K}\right)^{-1}\right)\end{aligned}$$

where $\mathbf{R}_{n,g} = P_r\sigma_{e,g}^2\mathbf{T}_g + \mathbf{I}_{N_r}$ amounts to the effective noise covariance in the BC phase.

The matrix form of $\mathrm{MSE}_{\mathrm{MAC}}$ in (13) is similarly expressed as

$$\begin{aligned}\mathrm{MSE}_{\mathrm{MAC}} &= E\left[\left\|(\mathbf{U}+\mathbf{F})\left(\mathbf{L}(\widehat{\mathbf{H}}\mathbf{x} + \Delta\mathbf{H}\mathbf{x} + \mathbf{n}_r) - \mathbf{x}\right)\right\|^2\right]\\ &= \mathrm{Tr}\left((\mathbf{U}+\mathbf{F})(\mathbf{L}\mathbf{R}_{\mathrm{yr}}\mathbf{L}^H - \mathbf{L}\widehat{\mathbf{H}}\mathbf{R}_x\right.\\ &\qquad\left. - \mathbf{R}_x\widehat{\mathbf{H}}^H\mathbf{L}^H + \mathbf{R}_x)(\mathbf{U}+\mathbf{F})^H\right)\\ &\overset{(a)}{=} \mathrm{Tr}\left((\mathbf{U}+\mathbf{F})\left(\mathbf{R}_x - \mathbf{R}_x\widehat{\mathbf{H}}^H\mathbf{R}_{\mathrm{yr}}^{-1}\widehat{\mathbf{H}}\mathbf{R}_x\right)(\mathbf{U}+\mathbf{F})^H\right)\\ &\overset{(b)}{=} \mathrm{Tr}\left((\mathbf{U}+\mathbf{F})\left(\widehat{\mathbf{H}}^H\mathbf{R}_{n,h}^{-1}\widehat{\mathbf{H}} + \mathbf{R}_x^{-1}\right)^{-1}(\mathbf{U}+\mathbf{F})^H\right)\end{aligned} \tag{15}$$

where (a) is obtained from $\mathbf{L} = \mathbf{R}_x\widehat{\mathbf{H}}^H\mathbf{R}_{\mathrm{yr}}^{-1}$, and (b) follows from the matrix inversion lemma.

Therefore, it is now readily shown that the optimization problem for $\mathbf{U}$ is formulated as an unconstrained $2K$-dimensional convex problem with respect to $\alpha_i\ \forall i$ as

$$\mathbf{U}_{\mathrm{opt}} = \arg\min_{\alpha_1,\ldots,\alpha_{2K}} \mathrm{Tr}\left(\mathbf{R}_e(\mathbf{U})\right) \tag{16}$$

where $\mathbf{R}_e(\mathbf{U}) \triangleq (2K/P_r)\widetilde{\mathbf{R}}_{\mathrm{yr}}(\mathbf{U}+\mathbf{F})^H\mathbf{E}_B(\mathbf{U}+\mathbf{F}) + (\mathbf{U}+\mathbf{F})\mathbf{E}_M(\mathbf{U}+\mathbf{F})^H$ with

$$\mathbf{E}_M \triangleq \left(\widehat{\mathbf{H}}^H\mathbf{R}_{n,h}^{-1}\widehat{\mathbf{H}} + \mathbf{R}_x^{-1}\right)^{-1}$$

$$\mathbf{E}_B \triangleq \left(\widehat{\mathbf{G}}\mathbf{R}_{n,g}^{-1}\widehat{\mathbf{G}}^H + \frac{2K}{P_r}\mathbf{I}_{2K}\right)^{-1}.$$

Now, we take a derivative of $\mathrm{Tr}(\mathbf{R}_e(\mathbf{U}))$ with respect to α_i and set it to zero. Then, this results in

$$\alpha_{i,\mathrm{opt}} = \frac{C_iB_{\pi(i)} - A_{\pi(i)}B_i}{A_iA_{\pi(i)} - |C_k|^2}, \qquad \text{for } i = 1,\ldots,2K \tag{17}$$

where $A_i = (2K/P_r)[\widetilde{\mathbf{R}}_{\mathrm{yr}}]_{i,i}[\mathbf{E}_B]_{i,i} + [\mathbf{E}_M]_{i,i}$, $B_i = (2K/P_r)[\widetilde{\mathbf{R}}_{\mathrm{yr}}]_{i,i}[\mathbf{E}_B]_{i,\pi(i)} + (2K/P_r)[\widetilde{\mathbf{R}}_{\mathrm{yr}}]_{\pi(i),i}[\mathbf{E}_B]_{i,i} + [\mathbf{E}_M]_{\pi(i),i}$, and $C_i = (2K/P_r)[\widetilde{\mathbf{R}}_{\mathrm{yr}}]_{\pi(i),i}[\mathbf{E}_B]_{i,\pi(i)}$. Finally, combining (17) with the results in (12), we obtain the beamforming solution $\mathbf{Q}_{\mathrm{opt}}$.

We note that the derived SIC parameters (17) may not achieve the perfect SIC at the users due to the estimation errors, but they provide stochastically the best choice for minimizing the MSE. The remaining problem is how the relay inform the ith user of $\alpha_{i,\mathrm{opt}}$, during which some additional noise is typically imposed. This will be discussed in more detail in the following section. Finally, we would like to raise a few remarks for the proposed beamforming scheme.

Remark 1: If perfect CSI of all links, i.e., genie-aided channel estimation, is available at the users, the ith user would set $\alpha_i = \gamma^{-1}\mathbf{g}_i^T\mathbf{Q}_{\mathrm{opt}}\mathbf{h}_i$ to completely remove the self-interference, whereas the relay would still use $\mathbf{U}_{\mathrm{opt}}$ due to the channel errors at the relay. This situation is unrealistic but useful as a performance upper bound of the system. In contrast, when no-CSI is allowed at the users, one may adopt the SDMA scheme by setting $\alpha_i = 0\forall i$ at both the relay and the users, which serves as a performance lower bound of the proposed scheme.

Remark 2: Since our solution is MMSE optimal in the general MPTW systems with imperfect CSI, it includes the previous MMSE schemes such as the SDMA [10], the two-way MMSE [7], and the MPTW MMSE [15] as special cases with $\alpha_i = 0\ \forall i$, $K = 1$, and $\sigma_{e,h}^2 = \sigma_{e,g}^2 = 0$, respectively.

Remark 3: The overall complexity of the proposed scheme is dominated by the N_r-dimensional matrix inversion whose computational efforts grow with an order of $\mathcal{O}(N_r^3)$. On the contrary, the conventional two-way or MPTW beamformer designs based on the Kronecker product in [7], [15], and [16] require N_r^2-dimensional matrix inversions whose complexity is $\mathcal{O}(N_r^6)$. Therefore, our method offers an efficient way to calculate the solution and the robustness against channel errors. We note that the complexity gain will become more significant as K and N_r increase, as in the large-scale array regime [18].

IV. Self-Interference Cancellation at Users

With the MMSE strategy, the maximum-likelihood decoder at the ith user makes a decision according to the rule as $\hat{x}_i = \arg\min_{x_{\pi(i)}\in\mathbb{S}_{\pi(i)}} |(1/\gamma)y_i - \alpha_ix_i - x_{\pi(i)}|^2$, where $\mathbb{S}_{\pi(i)}$ indicates a set of modulated signals at user $\pi(i)$. Therefore, for the proposed MPTW systems to operate properly, accurate estimation of α_i is essential for the users.[2] It may be achieved by the relay feeding forward the quantized information of α_i to the ith user. However, this is practically less attractive, because transporting arbitrary complex values by control channels might lead to inefficient bandwidth usage

[2]We assume that γ is known to all users because it is a common real value that is easy to broadcast.

and heavy distortions. Moreover, as K becomes large, setting up individual control channels with each of $2K$ users may increase the system overhead significantly. To tackle the problem, here, we suggest an efficient downlink training method from the relay whose signaling overhead is independent of the system parameters K and N_r.

A. Downlink Training for SIC

Before we proceed further, we introduce the following two lemmas that show that the effective channel estimation in the BC phase may be sufficient for the ith user to perform the SIC.

Lemma 1: In the low-input SNR region, i.e., $P_s \to 0^+$, the SDMA scheme is optimal, i.e., $\mathbf{U}_{\rm opt} = \mathbf{0}_{2K}$.

Proof: See Appendix A. ■

Lemma 2: When the channel estimation at the relay is small, i.e., $\sigma_{e,h}^2 \ll 1$ and $\sigma_{e,g}^2 \ll 1$, the SIC parameter $\alpha_{i,\rm opt}$ in (17) is approximated as $\alpha_{i,\rm opt} \simeq [\overline{\mathbf{G}}]_{i,i}$ in both high-SNR ($P_s \to \infty$ or $P_r \to \infty$) and low-SNR ($P_s \to 0^+$ and $P_r \to 0^+$) regimes, where $\overline{\mathbf{G}} \triangleq \widehat{\mathbf{G}}\overline{\mathbf{B}}$ represents an effective downlink channel with an effective precoder $\overline{\mathbf{B}} = \mathbf{B}(\mathbf{U}_{\rm opt} + \mathbf{F})$ at the relay.

Proof: See Appendix B. ■

Based on the aforementioned two lemmas, we consider a scheme where a scaled unitary training matrix $\mathbf{X}_{\rm tr} \in \mathbb{C}^{2\times 2}$ is sent from the relay simultaneously to all user pairs in the common downlink training phase. Define $\mathbf{G}_k \in \mathbb{C}^{2\times N_r}$ and $\mathbf{B}_k \mathbb{C}^{N_r\times 2}$ as the kth submatrix of $\mathbf{G}$ and $\mathbf{B}$, i.e., $\mathbf{G} = [\mathbf{G}_1^T, \ldots, \mathbf{G}_K^T]^T$ and $\mathbf{B} = [\mathbf{B}_1, \ldots, \mathbf{B}_K]$, respectively, and $\mathbf{P}_k \in \mathbb{C}^{2\times 2}$ as the kth block diagonal component of $\mathbf{U}_{\rm opt} + \mathbf{F}$. Then, the corresponding received signals at the kth user pair is given by

$$\mathbf{Z}_k = \gamma \mathbf{G}_k \mathbf{B}_k \mathbf{P}_k \mathbf{X}_{\rm tr} + \sum_{j\neq k} \gamma \mathbf{G}_k \mathbf{B}_j \mathbf{P}_j \mathbf{X}_{\rm tr} + \mathbf{N}_{u_k}. \tag{18}$$

Multiplying from the right by $\mathbf{X}_{\rm tr}^H$ and using the fact that, by design, $\mathbf{X}_{\rm tr}\mathbf{X}_{\rm tr}^H = P_{\rm tr}\mathbf{I}_2$, where $P_{\rm tr}$ is the power allocated to training, we obtain

$$\begin{aligned}\mathbf{Z}_k\mathbf{X}_{\rm tr}^H &= P_{\rm tr}\gamma \mathbf{G}_k \mathbf{B}_k \mathbf{P}_k + P_{\rm tr}\sum_{j\neq k} \gamma \mathbf{G}_k \mathbf{B}_j \mathbf{P}_j + \mathbf{N}_{u_k}\mathbf{X}_{\rm tr}^H \\ &= P_{\rm tr}\gamma \widehat{\mathbf{G}}_k \mathbf{B}_k \mathbf{P}_k + P_{\rm tr}\sum_{j\neq k} \gamma \widehat{\mathbf{G}}_k \mathbf{B}_j \mathbf{P}_j \\ &\quad + \mathbf{N}_{u_k}\mathbf{X}_{\rm tr}^H + P_{\rm tr}\gamma\Delta\mathbf{G}_k \sum_j \mathbf{B}_j\mathbf{P}_j \end{aligned} \tag{19}$$

where the first term in (19) represents the desired signal, and the remaining terms correspond to the estimation noise.

Dividing (19) by $\gamma P_{\rm tr}$ and extracting (l,l)th element for $l \in \{1,2\}$ from a matrix $\mathbf{Z}_k\mathbf{X}_{\rm tr}$, and applying the approximation in Lemma 2, the lth user in the kth pair finds a noisy observation of $\alpha_{l,\rm opt}$ as

$$\widehat{\alpha}_l \simeq \alpha_{l,\rm opt} + \Delta\alpha_l \tag{20}$$

where $\Delta\alpha_l \triangleq [\sum_{j\neq k} \widehat{\mathbf{G}}_k\mathbf{B}_j\mathbf{P}_j + P_{\rm tr}^{-1}\gamma^{-1}\mathbf{N}_{u_k}\mathbf{X}_{\rm tr}^H + \Delta\mathbf{G}_k\sum_j \mathbf{B}_j \mathbf{P}_j]_{l,l}$. Note that the first summation term $\sum_{j\neq k}\gamma\widehat{\mathbf{G}}_k\mathbf{B}_j\mathbf{P}_j$ is negligible since we have $\widehat{\mathbf{G}}_k\mathbf{B}_j \simeq \mathbf{0}_2$ for $j \neq k$ (see Appendix B). Thus, $\Delta\alpha_l \sim \mathcal{CN}(0, \sigma_{e,u}^2)$ is approximately CSCG with $\sigma_{e,u}^2 = (1/P_{\rm tr}\gamma^2) + \sigma_{e,g}^2 \mathrm{Tr}(\overline{\mathbf{B}}\overline{\mathbf{B}}^H\mathbf{T}_g)$.[3]

First, we observe that $\sigma_{e,h}^2$ does not affect $\sigma_{e,u}^2$, which means that, with our training method, the estimation noise for the MAC phase channel $\mathbf{H}$ will not be propagated to the users. In addition, assuming that each coherence block spans T time slots, the training phase occupies only two slots, and the remaining $T-2$ slots are still available for data transmission. Therefore, the spectral efficiency penalty factor ζ in (4) is negligible as T increases. Moreover, with the proposed training method, the system can be made more practical, because the training sequence and the data signal $\tilde{\mathbf{y}}$ will go through the same precoder $\overline{\mathbf{B}}$ so that we can put them together in the same data processing line.

[3]A proper use of the MMSE estimator at each user may lead to performance gain, but it is not investigated further here.

B. SIC Mode Selection

Here, we treat the estimation errors at the user sides, which may cause malfunction of the SIC operation. Suppose that each user obtains the erroneous SIC parameter $\widehat{\alpha}_i$ from the downlink training proposed in the previous section. We also define the noisy SIC matrix by $\widehat{\mathbf{U}} = \mathbf{U}_{\rm opt} + \Delta\mathbf{U}$, where $\Delta\mathbf{U} = \mathrm{diag}\{\Delta\alpha_1, \ldots, \Delta\alpha_{2K}\}$. Then, the sum MSE in (6) can be rewritten by

$$\begin{aligned} E\left[\|\mathbf{e}\|^2\right] &= E\left[\left\|\frac{1}{\gamma}\mathbf{y} - (\mathbf{U}_{\rm opt}+\mathbf{F})\mathbf{x} - \Delta\mathbf{U}\mathbf{x}\right\|^2\right] \\ &= \mathrm{Tr}\left(\mathbf{R}_e(\mathbf{U}_{\rm opt})\right) + \mathrm{Tr}\left(P_s\sigma_{e,u}^2\mathbf{I}_{2K}\right) \end{aligned} \tag{21}$$

where the expectation is taken over all random factors, including $\mathbf{n}_r$, $\mathbf{z}$, $\mathbf{x}$, $\Delta\mathbf{H}$, $\Delta\mathbf{G}$, and $\Delta\mathbf{U}$.

Now, we see from (21) that for given $\sigma_{e,u}^2$, the MSE is unbounded above with increasing P_s or K. Therefore, for large SNR P_s and network size K, it might be better for the users not to perform the SIC. Motivated by this fact, we suggest an intelligent selection criterion for choosing the SIC matrix between $\mathbf{U} = \mathbf{U}_{\rm opt}$ and $\mathbf{U} = \mathbf{0}_{2K}$, which lead to MMSE-BD and SDMA schemes, respectively. Denoting $\sigma_f^2 = (1/2KP_s)\mathrm{Tr}(\mathbf{R}_e(\mathbf{0}_{2K}) - \mathbf{R}_e(\mathbf{U}_{\rm opt}))$, the selection can be simply made by comparing their sum-MSEs as follows:

$$\mathbf{Q}_{\rm SMS} = \begin{cases}\mathbf{Q}_{\rm opt}, & \text{if } \sigma_{e,u}^2 \le \sigma_f^2 \\ \mathbf{Q}_{\rm SDMA}, & \text{if } \sigma_{e,u}^2 > \sigma_f^2.\end{cases} \tag{22}$$

From (22), we find several interesting observations. First, it is shown that if the channel estimation error for $\mathbf{G}$ at the relay, i.e., $\sigma_{e,g}^2$ is greater than $\sigma_f^2/\mathrm{Tr}(\overline{\mathbf{B}}\overline{\mathbf{B}}^H\mathbf{T}_g)$, there will be no gain from the SIC no matter how much training power the relay uses. Therefore, the system must be switched to the SDMA mode so that the relay can stop transmitting the training sequence. On the contrary, the estimation error $\sigma_{e,g}^2$ is sufficiently small such that $\sigma_{e,g}^2 < \sigma_f^2/\mathrm{Tr}(\overline{\mathbf{B}}\overline{\mathbf{B}}^H\mathbf{T}_g)$ and the relay uses enough training power $P_{\rm tr} \ge 1/(\gamma^2\sigma_f^2 - \gamma^2\sigma_{e,g}^2\mathrm{Tr}(\overline{\mathbf{B}}\overline{\mathbf{B}}^H\mathbf{T}_g))$, some SIC gain may be available. In this case, thus, the system can stay in the MMSE-BD mode, and the relay transmits the training sequence. Once the relay chooses the mode, each user can recognize it by checking the existence of the training signal.

V. Numerical Results

Here, we demonstrate the efficiency of our proposed scheme using Monte Carlo simulations. For ease of presentation, we assume that all users are distributed approximately in the same distance from the relay throughout the simulations. We consider the time-division duplexing operation at the relay; thus, $\sigma_{e,h}^2 = \sigma_{e,g}^2$, $\mathbf{T}_g = \mathbf{T}_h^T$, and $\mathbf{G}_w = \mathbf{H}_w^T$, where all entries of $\mathbf{H}_w$ are i.i.d. CSCG with zero mean and unit variance. The entries of the correlation matrix $\mathbf{T}_h$ is generated using the exponential model [16] by $[\mathbf{T}_h]_{k,l} = \beta^{|k-l|}$ for $\forall\, 1 \le k,l \le N_r$.

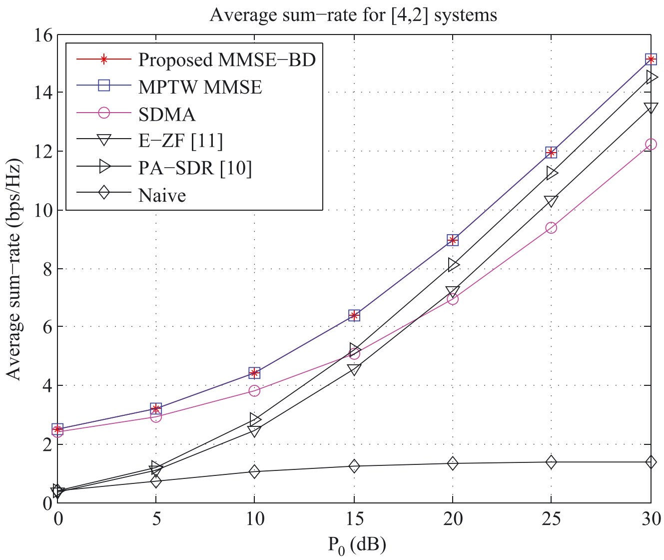

Fig. 2. Sum-rate performance comparison for [4, 2] systems with perfect CSI ($\sigma_{e,h}^2 = \sigma_{e,u}^2 = 0$).

Throughout the simulations, we set $P_s = P_r = P_0$, $\beta = 0.5$, and $\zeta = 1$. We denote the MPTW systems with N_r and K by $[N_r, K]$.

First, in Fig. 2, we present the comparison of the average sum-rate performance of various beamforming schemes in [4, 2] systems assuming that the full CSI is available at all nodes, i.e., $\sigma_{e,h}^2 = \sigma_{e,u}^2 = 0$. For the proposed scheme, as the target signal power equals P_s, the SINR is computed instead of (3) by

$$\mathrm{SINR}_i = \frac{P_s}{\mathrm{IN}} \tag{23}$$

where IN represents the interference plus noise term which is defined by

$$\mathrm{IN} = P_s \left|1 - \mathbf{g}_i^T \widetilde{\mathbf{Q}} \mathbf{h}_{\pi(i)}\right|^2 + P_s \left|\mathbf{g}_i^T \widetilde{\mathbf{Q}} \mathbf{h}_i - \widehat{\alpha}_i\right|^2 + P_s \sum_{j \neq \{i, \pi(i)\}}^{2K} \left|\mathbf{g}_i^T \widetilde{\mathbf{Q}} \mathbf{h}_j\right|^2 + \left\|\mathbf{g}_i^T \widetilde{\mathbf{Q}}\right\|^2 + \gamma^{-1}.$$

It is clear that the proposed MMSE-BD outperforms the conventional ZF-based BD methods such as the enhanced ZF (E-ZF) [13] and the pair-aware ZF (PA-SDR) [12]. From the circle line for the SDMA [10], we confirm that no SIC is optimum, i.e., $\mathbf{U}_{\mathrm{opt}} = \mathbf{0}_{2K}$, at low SNR as proved in Lemma 1. On the contrary, as SNR increases, the SDMA may suffer from rate loss due to no SIC gain. The figure also shows that the proposed MMSE-BD achieves the same performance as the MPTW MMSE [15] with reduced complexity.

Fig. 3 compares the average computation time of the previously mentioned MMSE beamformers with $P_0 = 20$ dB and $K = N_r/2$. The running time is measured by the *tic-toc* operation in MATLAB and averaged over 100 independent channel realizations. It is shown that the proposed MMSE-BD keeps the complexity in the same order with the SDMA with respect to N_r, whereas the required computation time for the MPTW MMSE grows with a higher order. Thus, for large N_r or K, a complexity gain of our scheme will be more significant.

Next, in Fig. 4, we demonstrate the robustness of the proposed MMSE-BD in [12, 6] systems. For fair comparison, we assume that the genie-aided channel estimation is available at the users, i.e., $P_s|1 - \mathbf{g}_i^T \widetilde{\mathbf{Q}} \mathbf{h}_{\pi(i)}|^2 = 0$ in (23). This is because the conventional designs are not defined in the case of imperfect CSI at the users. We confirm

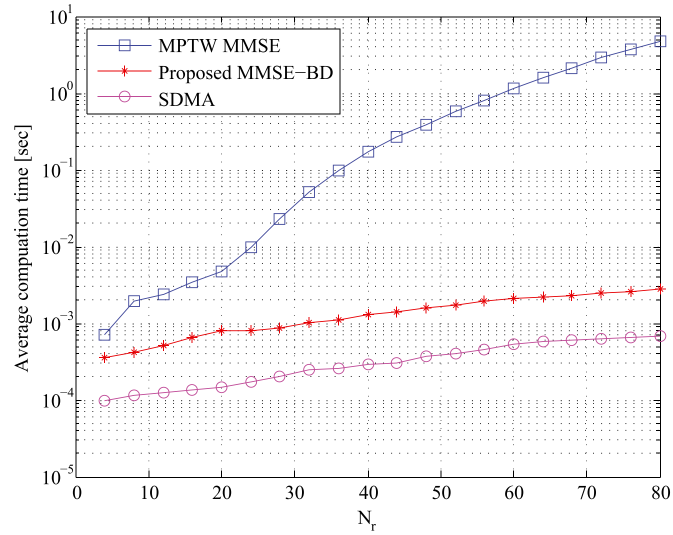

Fig. 3. Computational complexity comparison with $\sigma_{e,h}^2 = \sigma_{e,u}^2 = 0$, and $P_0 = 20$ dB.

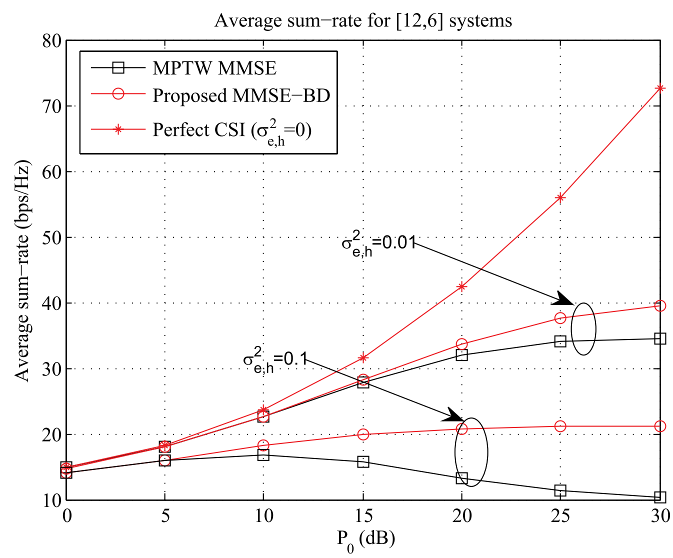

Fig. 4. Sum-rate performance comparison for [12, 6] systems with imperfect CSI at the relay ($\sigma_{e,h}^2 = 0.1$ and 0.01).

that the proposed beamformer shows better robustness against channel estimation errors at the relay over the conventional scheme.

Fig. 5 shows the efficiency of the proposed SMS in [6, 3] systems. Here, we consider channel estimation errors at both the relay and the users. We assume that the channel uncertainty level at the relay is fixed to $\sigma_{e,h}^2 = 0.001$. In addition, we consider a situation where the 2×2 unitary training sequence is transmitted simultaneously to all users with $P_{\mathrm{tr}} \in \{10, 100\}$ for channel estimation at users. As predicted in Section IV-B, we see that a sufficient training power is essential to achieve some performance gain from the SIC. As SNR increases, however, the MMSE-BD experiences inevitable performance loss even with the high training power. This is because the increase of P_s also amplifies the estimation noise $\sigma_{e,u}^2$, which may incur the malfunction of the SIC at users. It is shown that a simple switching operation according to the SMS criterion in (22) will resolve the problem and provide good robustness.

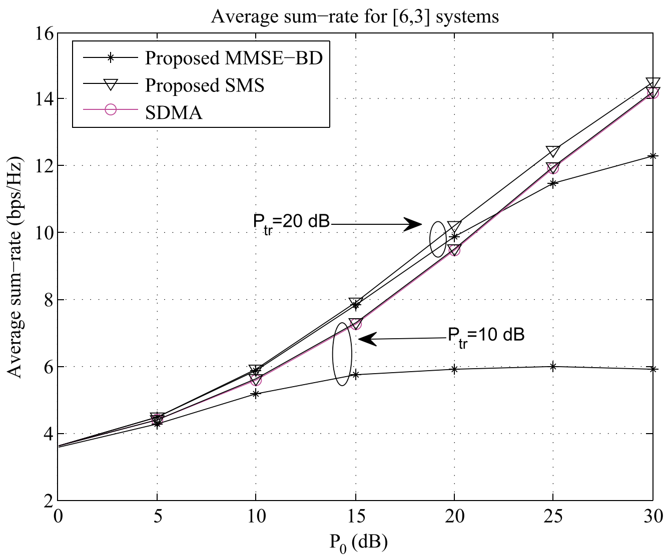

Fig. 5. Sum-rate performance comparison for [6, 3] systems with imperfect CSI at both the relay and the users.

VI. Conclusion

In this paper, we have proposed the robust beamforming designs in MPTW nonregenerative relaying systems where a multiantenna relay supports $2K$ users having a single antenna. By considering the SIC parameters as unknown variables that need to be optimized together with the relay beamformer, we could obtain an insightful closed-form solution that is not only robust to the imperfect CSI but also adjustable to various CSI circumstances at the users and the relay. We have also investigated the training method to inform the users of the SIC parameters and suggested the SMS scheme to improve the robustness performance. Finally, numerical simulations demonstrate the efficiency of the proposed schemes.

Appendix A
Proof of Lemma 1

Recalling that $\mathbf{R}_x = P_s\mathbf{I}_{2K}$ with $P_s \to 0^+$. Then, we can show that the MSE matrix $\mathbf{R}_e(\mathbf{U})$ in (16) is alternatively written by

$$\mathbf{R}_e(\mathbf{U}) = \mathbf{R}_x\left(\mathbf{R}_x^{-1}\overline{\mathbf{R}}_{\text{yr}}\left(\frac{P_r}{2K}\mathbf{G}\mathbf{R}_{n,g}^{-1}\mathbf{G}^H + \mathbf{I}_{2K}\right)^{-1}\right.$$
$$\left. + (\mathbf{U}+\mathbf{F})\left(\mathbf{R}_x\mathbf{H}^H\mathbf{R}_{n,h}^{-1}\mathbf{H} + \mathbf{I}_{2K}\right)^{-1}(\mathbf{U}+\mathbf{F})^H\right)$$
$$\simeq \mathbf{R}_x(\mathbf{U}+\mathbf{F})(\mathbf{U}+\mathbf{F})^H \tag{24}$$

where (24) holds because $\mathbf{R}_x^{-1}\overline{\mathbf{R}}_{\text{yr}} \approx \mathbf{0}_{2K}$ and $(\mathbf{R}_x\mathbf{H}^H\mathbf{H} + \mathbf{I}_{2K})^{-1} \approx \mathbf{I}_{2K}$ as $P_s \to 0^+$. Therefore, it is immediate from (24) that $\mathbf{U}_{\text{opt}} = \mathbf{0}_{2K}$, which means that the SDMA is optimum at low P_s.

Appendix B
Proof of Lemma 2

Since the MSE is a decreasing function of the SI power, it is obvious that the derived solution in (17) is equal to $\alpha_{i,\text{opt}} = \gamma^{-1}\hat{\mathbf{g}}_i^T\mathbf{Q}_{\text{opt}}\hat{\mathbf{h}}_i$, where $\hat{\mathbf{h}}_i$ and $\hat{\mathbf{g}}_i^T$ indicate the ith column and row of the estimated channels $\widehat{\mathbf{H}}$ and $\widehat{\mathbf{G}}$, respectively. Therefore, as $P_s \to \infty$, we have $\alpha_{i,\text{opt}} = \hat{\mathbf{g}}_i^T\overline{\mathbf{B}}\mathbf{L}\hat{\mathbf{h}}_i \simeq [\overline{\mathbf{G}}]_{i,i}$ for $\sigma_{e,h}^2 \ll 1$, because $\mathbf{L}$ reduces to the ZF receiver. Similarly, when $P_r \to \infty$ and $\sigma_{e,g}^2 \ll 1$, we obtain $[\overline{\mathbf{G}}]_{i,i} \simeq [\hat{\mathbf{U}} + \mathbf{F}]_{i,i} = \alpha_{i,\text{opt}}$. This is because $\mathbf{B}$ plays a role of the ZF channel inverter at high SNR. Hence, Lemma 2 holds in either case $P_s \to \infty$ or $P_r \to \infty$.

Meanwhile, from Lemma 1, we know that $\mathbf{U}_{\text{opt}} \simeq \mathbf{0}_{2K}$ for low P_s. It is readily shown that all components in $\mathbf{B}$ converge to zero as P_r becomes low, as shown in (12), which means that $[\overline{\mathbf{G}}]_{i,i} \simeq 0$. Therefore, for low-SNR regimes with $P_s, P_r \to 0^+$, we have $\alpha_{\text{opt}} \simeq [\overline{\mathbf{G}}]_{i,i}$, and the proof is completed.

References

[1] J. Laneman, D. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.

[2] C. Song, K.-J. Lee, and I. Lee, "MMSE based transceiver designs in closed-loop non-regenerative MIMO relaying systems," *IEEE Trans. Wireless Commun.*, vol. 9, no. 7, pp. 2310–2319, Jul. 2010.

[3] C. Song and I. Lee, "Diversity analysis of coded beamforming in MIMO-OFDM amplify-and-forward relaying systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2445–2450, Aug. 2011.

[4] C. Song, K.-J. Lee, and I. Lee, "MMSE-based MIMO cooperative relaying systems: Closed-form designs and outage behavior," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 8, pp. 1390–1401, Sep. 2012.

[5] H.-J. Choi, C. Song, H. Park, and I. Lee, "Transceiver designs for multipoint-to-multipoint MIMO amplify-and-forward relaying systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 1, pp. 198–209, Jan. 2014.

[6] B. Rankov and A. Wittneben, "Spectral efficient protocols for half-duplex fading relay channels," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 2, pp. 379–389, Feb. 2007.

[7] G. Li, Y. Wang, and P. Zhang, "Optimal linear MMSE beamforming for two way multi-antenna relay systems," *IEEE Commun. Lett.*, vol. 15, no. 5, pp. 533–535, May 2011.

[8] K.-J. Lee, H. Sung, E. Park, and I. Lee, "Joint optimization for one and two-way MIMO AF multiple-relay systems," *IEEE Trans. Wireless Commun.*, vol. 9, no. 12, pp. 3671–3681, Dec. 2010.

[9] R. Wang, M. Tao, and Z. Xiang, "Nonlinear precoding design for MIMO amplify-and-forward two-way relay systems," *IEEE Trans. Veh. Technol.*, vol. 61, no. 9, pp. 3984–3995, Nov. 2012.

[10] 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.

[11] E. Yilmaz, R. Zakhour, D. Gesbert, and R. Knopp, "Multi-pair two-way relay channel with multiple antenna relay station," in *Proc. IEEE ICC*, May 2010, pp. 1–5.

[12] A. U. T. Amah and A. Klein, "Pair-aware transceive beamforming for non-regenerative multi-user two way relaying," in *Proc. IEEE ICASSP*, 2010, pp. 2506–2609.

[13] H. Chung, N. Lee, B. Shim, and T. Oh, "On the beamforming design for MIMO multipair two-way relay channels," *IEEE Trans. Veh. Technol.*, vol. 61, no. 7, pp. 3301–3306, Sep. 2012.

[14] C. Y. Leow, Z. Ding, K. K. Leung, and D. L. Goeckel, "On the study of analogue network coding for multi-pair, bidirectional relay channels," *IEEE Trans. Wireless Commun.*, vol. 10, no. 2, pp. 670–681, Feb. 2011.

[15] H. Degenhardt and A. Klein, "Self-interference aware MIMO filter design for non-regenerative multi-pair two-way relaying," in *Proc. IEEE WCNC*, 2012, pp. 272–276.

[16] J. Zou, H. Luo, M. Tao, and R. Wang, "Joint source and relay optimization for non-regenerative MIMO two-way relay systems with imperfect CSI," *IEEE Trans. Wireless Commun.*, vol. 11, no. 9, pp. 3305–3315, Sep. 2012.

[17] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461–471, Feb. 2004.

[18] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing—The large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.