Successive Two-Way Relaying for Full-Duplex Users With Generalized Self-Interference Mitigation

Chao Ren, *Member, IEEE*, Haijun Zhang, *Senior Member, IEEE*, Jinming Wen, Jian Chen, *Member, IEEE*, and Chintha Tellambura, *Fellow, IEEE*

***Abstract*— In this paper, we propose a novel successive two-way relaying (STWR) system that uses a pair of conventional half-duplex (HD) relays to mimic a full-duplex two-way relay (FD-TWR). Although classical FD-TWR is spectral efficient and expands cell coverage, the proposed STWR utilizes the existing HD infrastructure to boost the FD implementation and offers bi-directional data exchange and low-complexity residual self-interference (RSI) mitigation. To formulate STWR, we develop a unified signal model to facilitate the mitigation of the generalized self-interference (GSI). GSI consists of back-propagating interference due to two-way relaying, RSI of FD sources and inter-relay interference caused by the pairs of HD relays. Because the GSI channel matrix has a distinct row linearity, we propose an efficient digital approach to remove the GSI and design two low-complexity algorithms. These algorithms avoid RSI channel estimation, full-rank matrix, and complex matrix computation. Our analysis and simulations show that: 1) the proposed STWR achieves the multiplexing gain of the true FD-TWR; 2) the distance between the two HD relays should be optimized to achieve the highest spectral efficiency; and 3) the STWR system with two algorithms can achieve a diversity order of one or two, respectively. Therefore, the STWR concept achieves a flexible tradeoff between performance and complexity, potentially enabling large-scale relay deployments.**

***Index Terms*— Full-duplex, residual self-interference cancellation, two-way relaying, successive relaying, cooperative communication.**

Manuscript received November 30, 2017; revised April 22, 2018 and August 4, 2018; accepted October 3, 2018. Date of publication December 4, 2018; date of current version January 8, 2019. This work was supported in part by the National Natural Science Foundation of China under Grants 61822104, 61471025, 61771044, and 11871248, in part by the Young Elite Scientist Sponsorship Program by CAST under Grant 2016QNRC001, in part by the Research Foundation of the Ministry of Education of China and China Mobile under Grant MCM20170108, in part by the Beijing Natural Science Foundation under Grants L172025 and L172049, in part by the 111 Project under Grant B170003, in part by the China Postdoctoral Science Foundation under Grant 2018M630072, and in part by the Fundamental Research Funds for the Central Universities under Grants FRF-GF-17-A6, RC1631, FRF-BD-17-015A, and FRF-TP-18-015A1. The associate editor coordinating the review of this paper and approving it for publication was M. R. Nakhai. *(Corresponding author: Haijun Zhang.)*

C. Ren and H. Zhang are with the Beijing Advanced Innovation Center for Materials Genome Engineering, University of Science and Technology Beijing, Beijing 100083, China, and also with the Beijing Engineering and Technology Research Center for Convergence Networks and Ubiquitous Services, Institute of Artificial Intelligence, University of Science and Technology Beijing, Beijing 100083, China (e-mail: chaoren@ustb.edu.cn; haijunzhang@ieee.org).

J. Wen is with the College of Information Science and Technology, Jinan University, Guangzhou 510632, China, and also with the College of Cyber Security, Jinan University, Guangzhou 510632, China (e-mail: jinming.wen@mail.mcgill.ca).

J. Chen is with the State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710071, China (e-mail: jianchen@mail.xidian.edu.cn).

C. Tellambura is with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2V4, Canada (e-mail: chintha@ece.ualberta.ca).

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/TWC.2018.2875982

I. Introduction

THE tremendous growth in global mobile traffic calls for speedy development of fifth generation (5G) wireless networks. Thus, 5G performance targets include data rate increase by 1000+ times, end-to-end latency below 100 ms, network cost reductions by 50%, and availability to increase to 95% in locations with bad coverage [1], [2]. Among many emerging technologies for 5G, full-duplex (FD) radio and two-way relaying (TWR) promise to boost the spectral efficiency, enable bi-directional communication, extend coverage, and mitigate congestion problems. Combining these two creates a full-duplex two-way relaying (FD-TWR) system which consists of two FD sources and an FD relay. This system allows the two sources to communicate bi-directionally using just two time slots. This is because the relay can receive two source messages simultaneously and forward the superimposed signal to both sources. Clearly, this will compress the four distinct phase required by half-duplex (HD) relay [3]. FD-TWR thus requires fewer channel uses, achieves the highest spectral efficiency among existing cooperative relaying protocols and can outperform fourth generation (4G) HD unidirectional relaying [4].

FD-TWR has been preliminarily studied without the consideration of *self-interference* (SI) [5], [6]. However, FD-TWR system realizations depend on the development of interference cancellation and the reduction of hardware/software complexity [7]. Implementation complexity of FD-TWR can be high due to the need for interference cancellation [4] because the sources and relay need to suppress SI created by the co-channel FD signals and the back-propagating interference (BPI) due to the TWR protocol. In fact, even in plain FD radio, the SI leakage from the local transmitter to the local receiver is as high as 100 dB, which cannot be eliminated completely for the current FD radios and residual self-interference (RSI) mitigation must also be used for FD-TWR [8], [9]. Thus, FD-TWR faces stringent technical challenges in practice.

A. Practical Issues in FD SI Cancellation

The input signal to the analog-to-digital converter (ADC) consists of multipath RSI echoes and SI leakages due

to imperfect hardware. They prevent passive and analog approaches from completely removing SI [10], [11]. With the ADC output signal, the receiver typically proceeds with digital or spatial cancellation [12], [13] to take advantage of digital signal processing to overcome key bottlenecks, including limited ADC dynamic range [14], inaccurate channel state information (CSI) of SI channel [7] and computational complexity[1] [7] and others [14].

Although judicious application of these techniques can yield approximate 90 dB SI suppression, RSI is still a major concern [14]. To further improve the FD performance, we will look for efficient methods completely eliminating the RSI leftover after joint passive-analog-digital/spatial cancellation, which is the removal of 'last' RSI. Future wireless will benefit if 'last' RSI mitigation is feasible with low-complexity algorithms, which reduces additional burden on existing FD software/hardware, i.e., dispensing with SI/RSI CSI, additional hardware and complex software in FD signal processing [8], [11], [16].

B. Related Works

1) FD-TWR Systems: The RSI is modeled as a fraction of transmitted signal propagating over a linear RSI channel, and FD-TWR performance is shown tremendously dependent on the power of RSI or the accuracy of RSI CSI. Reference [9] is an early attempt to incorporate RSI effect in basic FD-TWR system, while [4] and [17]–[20] consider RSI in more complex scenarios (multi-relay, massive MIMO and sum secrecy rate). However, these works do not consider RSI mitigation methods.

To mitigate RSI, a FD-TWR scheme based on Dempster-Shafter evidence theory is explored [21], while their assumption of perfect RSI CSI may not be practical. Although the CSI in HD TWR network has been well estimated in pioneering works [22]–[24], the characterization of the CSI of RSI channel in FD-TWR networks is sparse [25].[2] A single-block training scheme has been proposed to increase the accuracy of RSI CSI estimation in FD-TWR system [25]. However, CSI estimation error is still inevitable due to limited ADC dynamic range, hardware imperfectness and practical issues. Because the complete removal of RSI is impossible, some works have tried to turn the disadvantage into an advantage. Thus, RSI may be exploited to achieve cooperative diversity in FD-TWR by a self-coded distributed space-time coding scheme [26]. Similarly, for conventional FD one-way relaying, reference [27] uses imperfect RSI CSI to precode for RSI cancellation, but approximately 10% performance loss occurred compared with perfect RSI CSI scenarios.

To the best of our knowledge, finding both practical and efficient RSI mitigation algorithms for FD-TWR remains a challenging open problem.

[1]Due to complexity constraint or due to space limitation, spatial cancellation may not eliminate all of the SI [11], [15].

[2]According to [25], the estimation error of diffuse part of SI channel typically results in RSI. Due to complex features in propagation-domain and analog-circuit domain and all linear and non-linear distortions as well as transmitter noise, it is challenging to perfectly estimate the RSI channel after the pre-stage of SI channel estimation [11].

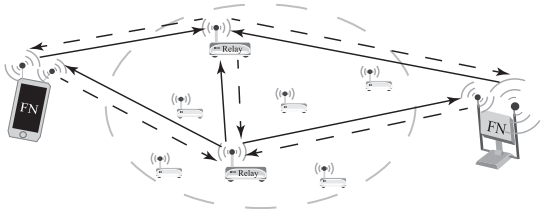


Fig. 1. Successive two-way relaying: two full-duplex source nodes and two half-duplex (HD) relays. Solid and dashed lines depict wireless links in even and odd time slots, respectively.

2) Legacy Option: Instead of hardware based FDR, near FD spectral efficiency may be achieved by virtual full-duplex (VFD) signaling [28], [29], which may circumvent the key bottlenecks (SI channel) in true FDR [7], [11], [14].

The VFD relaying system involves two HD relays. While one of them receives new data from the source, the other relay transmits the received data in the previous phase to the destination. In the next phase, the relays swap their roles. In this way, the source successively sends new messages as if a true FDR is employed [11]. Our previous work [30] showed that VFD achieves both near FD rate and full-diversity. The VFD concept has also been applied to two HD base stations to mimic one FD base station [31], [32]. In this system, multiple base stations can cooperate over wired backhaul links and can thus perfectly cancel the SI. However, this paper treats the case of wireless connectivity only. Efficient mitigation of the inter-relay interference (IRI) without additional links is critical in VFD relaying design.[3]

Notably, the deployment of HD relaying devices has been scheduled to extend 4G cellular coverage [30] and directly improve the performance by bringing antennas close to end users. Until complexity of SI mitigation and hardware cost become acceptable, this legacy option may also help ultra-dense FD-based wireless networks [33].

C. Contributions

The main motivation for successive two-way relaying (STWR) (Fig. 1) is that since the RSI may not be completely canceled in FD-TWR, a VFD TWR system can thus be formed to provide exactly similar services. STWR, which uses two distributed HD relays to help FD sources successively exchange data, can be defined as an updated version of VFD supporting two-way transmission or a virtual alternative using '4G legacy' for true FD-TWR. Consequently, the STWR concept relieves the high burden of estimating RSI channel.

The distinctive features and contributions of this paper are as follows:

1) **Generalized self-interference (GSI) formulation for interference mitigation.** We manage to treat BPI, IRI and RSI[4] as the output of a GSI channel, which simply comprises an ISI channel matrix and additive

[3]In contrast to other VFD relaying/BS systems, our proposed model enables two FD users exchange information bidirectionally, which needs to mitigate IRI, BPI and RSI.

[4]When we refer to RSI, it always means RSI occurring at the FD sources.

TABLE I
IMPORTANT NOTATIONS

Notation	Explanation
Bold upper case letters, e.g., $\mathbf{H}$; $H(i,j)$; $\mathbf{H}^{\mathrm{T}}$/$\mathbf{H}^{\mathrm{H}}$; bold lower case letters, e.g., $\mathbf{h}$; $h(i)$; diag$\{\mathbf{h}\}$; $\mathbf{I}_m$; $\mathbf{0}_{i\times j}$; tr($\mathbf{H}$); $\lambda_i[\mathbf{H}]$; $\mathbb{E}[\mathbf{n}]$	Matrices; entry of matrix $\mathbf{H}$ in the i-th row and j-th column; transpose and conjugate transpose of the matrix $\mathbf{H}$; vectors; i-th elements of vector $\mathbf{h}$; Diagonal matrix defined by the vector $\mathbf{h}$; $m\times m$ unity matrix; $i\times j$ all zero matrix; trace of a matrix $\mathbf{H}$; i-th eigenvalue of matrix $\mathbf{H}$; statistical expectation of $\mathbf{n}$.
a_k, b_k, r_k; $\bar{a}_k = 9-a_k$, $\bar{b}_k = 9-b_k$, $\bar{r}_k = 3-r_k, r_k \in \{1,2\}$	At time sllot k: the indexes of the transmitting antennas of FN_A, FN_B and $R_1(R_2)$, where $a_k \in \{3,6\}, b_k \in \{4,5\}$; the indexes of the receiving antennas of these nodes.
x_A^k, x_B^k, $x_{r_k}^k$ (y_A^k, y_B^k, $y_{\bar{r}_k}$); n_A^k, n_B^k, $n_{\bar{r}_k}$	The signals transmitted (received) by FN_A, FN_B and $R_2(R_1)$; the additive Gaussian white noise (AWGN) at these nodes.
$h_{a_k\bar{r}_k}$ and $h_{b_k\bar{r}_k}$; $h_{\bar{a}_k r_k}$ and $h_{\bar{b}_k r_k}$	The channels from sources' transmitting antennas to relays' receiving antennas; the channels from relays' transmitting antennas to sources' receiving antennas.

interference plus noise terms. The GSI channel has a beneficial special structure, which benefits interference mitigation.

2) **Complete and simple self-interference mitigation without RSI CSI.** A simple GSI mitigation method is proposed by exploiting the structure of GSI channel, which does not require RSI CSI and is a conversion with only a matrix multiplication and a matrix addition. However, this method requires the condition of full-rank matrix.
3) **Low-complexity GSI mitigation algorithms.** We propose two algorithms, namely GSIM1 and GSIM2, that dispense the need for the full-rank condition. These algorithms avoid solving for eigenvalues of a matrix or matrix-variable convex optimization problems. Hence, they offer the benefits of low computational complexity.
4) **Multi-sided performance enhancements.** We use performance analysis and simulation results to verify that STWR has the following threefold performance improvements: a) efficiency - both the spectral-efficiency and multiplexing gain of STWR approaches that of true FD relay in each one-way relaying communication; b) deployment - the distance between two distributed relays can be optimally designed to achieve the highest spectral-efficiency; and c) reliability - STWR with GSIM1 can achieve a diversity order of one while STWR with GSIM2 can recover the diversity order to two.

II. SYSTEM MODEL

A. System Model of STWR

The STWR system model (Fig. 1) employs two HD relays (R_1 and R_2) to connect two FD sources (FN_A and FN_B). Each source node has one antenna for transmission and one antenna for reception, while the relay node has only one antenna. The antennas of two relays are spatially distributed. The STWR protocol is described in Fig. 1 and Table II (where T, R and T_k denote transmitting, receiving and time slot k respectively) with the antennas indexes rule (1).

$$j_k = j_{k+2}, \quad j \in \{a, b, r, \bar{a}, \bar{b}, \bar{r}\}. \tag{1}$$

Each of the two FD sources broadcasts and receives L (assumed to be even) code words within $L+1$ time slots.

TABLE II
THE PROTOCOL OF STWR

Nodes	T_1	T_{even}	T_{odd}	T_L	T_{L+1}
Sources	T	T & R	T & R	T & R	R
R_1	-	R	T	R	T
R_2	R	T	R	T	-

At each odd/even numbered time slot, R_1/R_2 processes and broadcasts the signals to FD sources while R_2/R_1 keeps receiving. To this end, from slots 2 to L, the simultaneous virtual FD-TWR transmission between two sources on the same band is achieved. To implement the aforementioned protocol, we assume cell-level synchronization [30], [34].

We focus on mitigating the RSI after possible SI cancellation (e.g., joint passive-analog-digital/spatial cancellation). The channel coefficients of RSI channels (h_A and h_B) are assumed unknown, while the CSI of conventional wireless channels can be known by connected nodes [16]. IRI channel is denoted by h_R. We assume a block-fading model where the CSI is constant during $L+1$ time slots, but varies independently over different blocks [3]. No direct link between two sources is assumed, due to transmit power limitation or severe shadowing effect [19], [35].

B. Unified Signal Model and Generalized Self-Interference

At the k-th time slot, the transmitted symbols with power P from FD sources are $\sqrt{P}x_A^k$ and $\sqrt{P}x_B^k$, where x_A^k and x_B^k satisfy $|x_A^k|^2 = |x_B^k|^2 = 1$ and $x_A^{L+1} = x_B^{L+1} = 0$. The transmitted symbol from HD relay is

$$x_{r_k}^k = \alpha y_{\bar{r}_{k-1}}^{k-1}, \tag{2}$$

where $x_{r_1}^1 = 0$, $2 \le k \le L+1$, and α denotes the amplification factor of the relays [36], [37].

At the k-th time slot, the signals received by FN_A, FN_B and R_2 are

$$y_A^k = h_{\bar{a}_k r_k} x_{r_k}^k + \underline{\sqrt{P} h_A x_A^k} + n_A^k, \tag{3}$$

$$y_B^k = h_{\bar{b}_k r_k} x_{r_k}^k + \underline{\sqrt{P} h_B x_B^k} + n_B^k, \tag{4}$$

$$y_{\bar{r}_k}^k = \sqrt{P}(h_{a_k\bar{r}_k} x_A^k + h_{b_k\bar{r}_k} x_B^k) + h_R x_{r_k}^k + n_{\bar{r}_k}^k. \tag{5}$$

Definition 1 (Unified Signal Model and GSI Channel): All the receptions at FN_A can be formulated as a unified signal model in (6), which also is the output of GSI channel (11) with input $\boldsymbol{x}_B$.

Using (2), (3) and (5) recursively, we have

$$\mathbf{y}_A = \sqrt{P}(\mathbf{H}_{\bar{a}b}\mathbf{x}_B + \mathbf{H}_{\bar{a}a}\mathbf{x}_A + \underline{h_A\mathbf{x}_A}) + \mathbf{H}_{\bar{r}}\mathbf{n}_{\bar{r}} + \mathbf{n}_A, \quad (6)$$

where $\mathbf{y}_A = [y_A^1\ y_A^2 \cdots y_A^{L+1}]^{\mathrm{T}}$, $\mathbf{x}_B = [x_B^1\ x_B^2 \cdots x_B^L]^{\mathrm{T}}$, $\mathbf{x}_A = [x_A^1\ x_A^2 \cdots x_A^{L+1}]^{\mathrm{T}}$, $\mathbf{n}_{\bar{r}} = [n_{\bar{r}}^1\ n_{\bar{r}}^2 \cdots n_{\bar{r}}^L]^{\mathrm{T}}$, $\mathbf{n}_A = [n_A^1\ n_A^2 \cdots n_A^{L+1}]^{\mathrm{T}}$ are vectors, and $\mathbf{H}_{\bar{a}b}$, $\mathbf{H}_{\bar{a}a}$ and $\mathbf{H}_{\bar{r}}$ are $(L+1)\times L$ matrices with entries

$$H_{\bar{a}b}(I,J) = \mathcal{I}(I-J)\alpha^{I-J}h_R^{I-J-1}h_{\bar{a}_I r_I}h_{b_J\bar{r}_J}, \quad (7)$$

$$H_{\bar{a}a}(I,J) = \mathcal{I}(I-J)\alpha^{I-J}h_R^{I-J-1}h_{\bar{a}_I r_I}h_{a_J\bar{r}_J}, \quad (8)$$

$$H_{\bar{r}}(I,J) = \mathcal{I}(I-J)\alpha^{I-J}h_R^{I-J-1}h_{\bar{a}_I r_I}. \quad (9)$$

Note that $\mathcal{I}(\diamond)$ is an indicator determined by $\mathcal{I}(\diamond) = \begin{cases}1, \diamond > 0\\ 0, \diamond \le 0\end{cases}$. With generalized self-interference

$$\mathbf{I}_{GSI} = \sqrt{P}\mathbf{H}_{\bar{a}a}\mathbf{x}_A + \underline{\sqrt{P}h_A\mathbf{x}_A} = \sqrt{P}(\mathbf{H}_{\bar{a}a} + h_A)\mathbf{x}_A \quad (10)$$

and equivalent noise $\mathbf{n}_{GSI} = \mathbf{H}_{\bar{r}}\mathbf{n}_{\bar{r}} + \mathbf{n}_A$, the GSI interference channel can be found from

$$\mathbf{y}_A = \sqrt{P}\mathbf{H}_{\bar{a}b}\mathbf{x}_B + \mathbf{I}_{GSI} + \mathbf{n}_{GSI}. \quad (11)$$

Lemma 1 (Special Structure of GSI Channel): Considering GSI channel (11) with ISI channel matrix $\boldsymbol{H}_{\bar{a}b}$, parameters $J_1, M \in R^+$ and $\star \in \{\bar{a}b, \bar{r}\}$, the I_1-th and (I_1+M)-th entries in column J_1 for $\boldsymbol{H}_\star$ satisfies a simple rule

$$f_{I_1}H_\star(I_1+M, J_1) - g_{I_1+M}H_\star(I_1, J_1) = \begin{cases}0, & otherwise,\\ f_{I_1}H_\star(I_1+M, J_1), & I_1 \le J_1 < I_1+M,\end{cases} \quad (12)$$

where $f_{I_1} = h_{\bar{a}_{I_1}r_{I_1}}$ and $g_{1+M} = \alpha^M h_R^M h_{\bar{a}_{I_1+M}r_{I_1+M}}$ are elements in $(L-M+1)\times 1$ vectors $\boldsymbol{f} = [f_1\ f_2\ \cdots\ f_{L-M+1}]^T$ and $\boldsymbol{g} = [g_{1+M}\ g_{2+M}\ \cdots\ g_{L+1}]^T$ respectively.

Proof: It is straightforward to derive (12) from (7) and (9). □

Remark 1 (Unified Signal Model and GSI Mitigation): We further discuss the unified signal model formulation to better understand GSI.

- ***Before unified signal model formulation**, we observe three types of interference in (3), (4) and (5): 1) $\sqrt{P}h_Ax_A^k$ and $\sqrt{P}h_Bx_B^k$, the RSI of FD sources (underlined in (3), (6) and (10)); 2) $\sqrt{P}h_{a_k\bar{r}_k}x_A^k$, the BPI created by TWR protocol; and 3) $h_Rx_{r_k}^k$, the IRI induced by the other relay's transmission. Substantially, the recursion in (2) worsens the interference mitigation problem by introducing $\sum_{m=2}^{k-1}\sqrt{P}(h_{a_m\bar{r}_m}x_A^m + h_{b_m\bar{r}_m}x_B^m)$. The unified signal model is expected to convert this time-domain recursion to a condensed matrix form.*
- ***After unified signal model formulation**, recursive IRI terms are converted to ISI terms, e.g. $H_{\bar{a}b}(I,J)x_B^J$ $(I > J+1)$, which constitutes an ISI channel matrix $\boldsymbol{H}_{\bar{a}b}$. Matrix $\boldsymbol{H}_{\bar{a}b}$ should have a small number of non-zero entries to control the distortion from unintended symbols. SI of FD sources and BPI in (6) are defined as an additive interference term $\boldsymbol{I}_{GSI}$ in (11).*
- ***The target of GSI mitigation** is to reduce the number of non-zero entires in ISI channel matrix as well as mitigating additive interference $\boldsymbol{I}_{GSI}$, i.e., to find a specific conversion $\boldsymbol{y}_{SI_{AM}} = \boldsymbol{Q}_M(\boldsymbol{y}_A - \boldsymbol{T}_M)$ that yields*

$$\boldsymbol{y}_{SI_{AM}} = \boldsymbol{H}_{Q_M^{\bar{a}b}}\boldsymbol{x}_B + \boldsymbol{n}_{SI_{AM}}, \quad (13)$$

where $\boldsymbol{H}_{Q_M^{\bar{a}r}}$ is a new ISI matrix that has less non-zero entries than $\boldsymbol{H}_{\bar{a}b}$.
- ***Available knowledge to mitigate GSI** can be own data, local CSI except RSI channel and the CSI between two relays.*[5]
- *Lemma 1 reveals the **key information contained M**. If J_1 varies from 1 to L, the output of (12) constructs a vector, of which the number of non-zero entries is no larger than $M+1$.*

III. Generalized Self-Interference Mitigation

We next solve the problem of mitigating GSI by finding a specific conversion $\mathbf{y}_{SI_{AM}} = \mathbf{Q}_M(\mathbf{y}_A - \mathbf{T}_M)$ that yields (13), where $\mathbf{T}_M = \sqrt{P}\mathbf{H}_{\bar{a}a}\mathbf{x}_A$ and $\mathbf{Q}_M$ are given by (14) and (15) from Theorem 1 and Lemma 2. In this conversion: 1) the CSI of RSI channel is not needed; and 2) only local CSI and the CSI between two relays are utilized [30], [34], [37], [38].

Theorem 1 (Generalized Self-Interference Mitigation): Consider: 1) the target of GSI mitigation in Remark 1; 2) $L\times(L+1)$ matrices $\boldsymbol{Q}_M$ and $\boldsymbol{G}$; 3) full-rank diagonal matrix $\mathrm{diag}\{\boldsymbol{f}\}$ based on vector $\boldsymbol{f}$; and 4) vectors $\boldsymbol{v}^{gf} = [v_1^{gf}\ v_2^{gf}\ \cdots\ v_{L-M+1}^{gf}] = \mathrm{diag}\{\boldsymbol{f}\}^{-1}\boldsymbol{g}$ and $\boldsymbol{v}^{x_A} = \boldsymbol{x}_A - \left[\boldsymbol{0}_{1\times M}\ \ \mathrm{diag}\{\boldsymbol{v}^{gf}\}[x_A^1\ x_A^2 \cdots x_A^{L-M+1}]\right]^T$.

If $\boldsymbol{G}$ and $\boldsymbol{Q}_M$ satisfy

$$\boldsymbol{G}\boldsymbol{v}^{x_A} = \boldsymbol{0}_{L\times 1},\ and\ \boldsymbol{Q}_M = \boldsymbol{G}\left(\boldsymbol{I}_{L+1} - \begin{bmatrix}\boldsymbol{0}_{M\times(L-M+1)} & \boldsymbol{0}_{M\times M}\\ \mathrm{diag}\{\boldsymbol{v}^{gf}\} & \boldsymbol{0}_{(L-M+1)\times M}\end{bmatrix}\right), \quad (14)$$

then conversion $\boldsymbol{y}_{SI_{AM}} = \boldsymbol{Q}_M(\boldsymbol{y}_A - \sqrt{P}\boldsymbol{H}_{\bar{a}a}\boldsymbol{x}_A)$ mitigates the GSI and results in $\boldsymbol{y}_{SI_{AM}} = \boldsymbol{H}_{Q_M^{\bar{a}b}}\boldsymbol{x}_B + \boldsymbol{H}_{Q_M^{\bar{r}}}\boldsymbol{n}_{\bar{r}} + \boldsymbol{Q}_M\boldsymbol{n}_A$, where $\boldsymbol{H}_{Q_M^\star} = \boldsymbol{G}\boldsymbol{H}'_\star$, $\star \in \{\bar{a}b, \bar{r}\}$ and $\boldsymbol{H}'_\star$ has the structure in (42).

Proof: The structure of GSI channel can be used to finish the proof. Please see Appendix I. □

Remark 2: The target of GSI mitigation is achieved by Theorem 1. However, it is essential to determine $\boldsymbol{G}$ prior to applying Theorem 1. If $\boldsymbol{v}^{x_A}$ is given, multiple solutions of $\boldsymbol{G}$ satisfying $\boldsymbol{G}\boldsymbol{v}^{x_A} = \boldsymbol{0}_{N\times 1}$ can be found. Based on Remark 1 and $\boldsymbol{H}_{Q_M^\star} = \boldsymbol{G}\boldsymbol{H}'_\star$, $\boldsymbol{G}$ should be chosen by this rule: $\boldsymbol{G}$ converts $\boldsymbol{H}'_\star$ to $\boldsymbol{H}_{Q_M^\star}$ without significant increase in the number of non-zero entries. Based on Remark 1, the non-zero entries in each row of $\boldsymbol{H}_{Q_M^\star}$ is no larger than $M+1$.

[5]Due to cooperation, we assume relays provided the information about h_R to sources in hand-shake phase. This is possible if relay selection and deployment issues are needed to be solved before data exchange. The major advantage of STWR over FD-TWR is inter-relay channel h_R can be precisely estimated as conventional wireless channel, while the RSI channel estimation for true FD hardware requires considerable radio resources and the accuracy is severely affected by various factors [16].

Lemma 2: One example of $\boldsymbol{G}$ *for Theorem 1 can be*

$$\boldsymbol{G} = \boldsymbol{G}_{left} - \boldsymbol{G}_{right}, \tag{15}$$

where $\boldsymbol{G}_{left} = \left[\text{diag}\{[v_2^{x_A} \ \cdots \ v_{L+1}^{x_A}]\} \ \boldsymbol{0}_{L\times 1}\right]$ *and* $\boldsymbol{G}_{right} = \left[\boldsymbol{0}_{L\times 1} \ \text{diag}\{[v_1^{x_A} \cdots v_L^{x_A}]\}\right]$.

Proof: Based on Lemma 3 (Appendix II), we have

$$\mathbf{G}\mathbf{v}^{x_A} = \mathbf{G}_{\text{left}}\begin{bmatrix}[v_1^{x_A} \cdots v_L^{x_A}]^{\text{T}} \\ v_{L+1}^{x_A}\end{bmatrix} - \mathbf{G}_{\text{right}}\begin{bmatrix}v_1^{x_A} \\ [v_2^{x_A} \cdots v_{L+1}^{x_A}]^{\text{T}}\end{bmatrix} = \mathbf{0}_{L\times 1}.$$

Thus, $\mathbf{G}$ in (15) satisfies condition $\mathbf{G}\mathbf{v}^{x_A} = \mathbf{0}_{L\times 1}$ in Theorem 1. □

Remark 3: With $\boldsymbol{G}$ *from Lemma 2 and* $\boldsymbol{H}'_\star$ *from Theorem 1, the matrix form of* $\boldsymbol{H}_{Q_M^\star}$ *can be easily derived by applying* $\boldsymbol{H}_{Q_M^\star} = \boldsymbol{G}\boldsymbol{H}'_\star$. *1) It is straightforward to find that compared with* $\boldsymbol{H}'_\star$, $\boldsymbol{H}_{Q_M^\star}$ *only adds one non-zero entry in each row, and* $\boldsymbol{G}$ *is thus eligible for GSI mitigation. 2) In (42), the maximal number of non-zero entries of the row in ISI channel* $\boldsymbol{H}_{Q_M^\star}$ *is* $M+1$. *Thus, considering the structure and affection of* $\boldsymbol{H}_{Q_M^\star}$, *parameter* M *is the maximal number of residual ISI terms in each symbol after GSI mitigation.*

IV. Case I: $M = 1$

We now choose a minimal parameter $M = 1$ for GSI mitigation. We give the model of GSI mitigation, design a decomposed algorithm for $M = 1$ case, and analyze the performance of STWR in terms of spectral efficiency, multiplexing gain and diversity order.

A. Model for Generalized Self-Interference Mitigation

Based on Theorem 1 and Lemma 2, the result of GSIM1, is

$$\mathbf{y}_{SI_{A1}} = \sqrt{P}\mathbf{H}_{SI_{A1}}\mathbf{x}_B + \mathbf{n}_{SI_{A1}}, \tag{16}$$

where $\mathbf{H}_{SI_{A1}} = \mathbf{H}_{Q_M^{\bar{a}b}}|_{M=1}$ (see (17)), $\mathbf{n}_{SI_A} = \left(\mathbf{H}_{Q_M^{\bar{r}}}\mathbf{n}_{\bar{r}} + \mathbf{Q}_M\mathbf{n}_A\right)|_{M=1} = \left[N_{SI_A}^1 \ \cdots \ N_{SI_A}^L\right]^{\text{T}}$, $N_{SI_{A1}}^k = X_{A_1}^{k+1}N_{A1}^k - X_{A_1}^k N_{A1}^{k+1}$, $X_{A_1}^k = v_{x_A}^k|_{M=1} = x_A^k - \frac{\alpha h_R h_{\bar{a}_k r_k}}{h_{\bar{a}_{k-1}r_{k-1}}}x_A^{k-1}$, $H_{R1}^k = \alpha h_{\bar{a}_k r_k}h_{b_{k-1}\bar{r}_{k-1}}$, $C_1^k = X_{A_1}^{k+1}H_{R1}^k$, $D_1^k = -X_{A_1}^k H_{R1}^{k+1}$, $N_{A1}^k = -\alpha\frac{h_{\bar{a}_k r_k}h_R}{h_{\bar{a}_{k-1}r_{k-1}}}n_A^{k-1} + \alpha h_{\bar{a}_k r_k}n_{\bar{r}_{k-1}}^{k-1} + n_A^k$.

$$\mathbf{H}_{SI_{A1}} = \begin{bmatrix} D_1^1 & 0 & \cdots & 0 & 0 & 0 \\ C_1^2 & D_1^2 & \cdots & 0 & 0 & 0 \\ 0 & C_1^3 & \ddots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & C_1^{L-1} & D_1^{L-1} & 0 \\ 0 & 0 & \cdots & 0 & C_1^L & D_1^L \end{bmatrix}, \tag{17}$$

Equ. (16) clearly indicates that the GSI is converted to ISI.

Remark 4: From the model (16), we draw the following two conclusions.

1) *Based on (1), the equivalent channel coefficients corresponding to* x_B^{k-2} *and* x_B^k *satisfy*

$$C_1^{k-1} = D_1^k = -\frac{X_{A_1}^k}{X_{A_1}^{k+2}}C_1^{k+1}, \tag{18}$$

which helps us simplify the performance analysis.

Algorithm 1 GSIM1 for $(k+1)$-th Time Slot

Input: $\mathbf{s}_{k-1} = \begin{bmatrix}x_A^{k-1} \\ y_A^{k-1}\end{bmatrix}$, $\mathbf{s}_k = \begin{bmatrix}x_A^k \\ y_A^k\end{bmatrix}$, $\mathbf{s}_{k+1} = \begin{bmatrix}x_A^{k+1} \\ y_A^{k+1}\end{bmatrix}$, $\mathbf{V}_k = \begin{bmatrix}\frac{h_R}{h_{\bar{a}_{k-1}r_{k-1}}} & 0 \\ \sqrt{P}h_{a_{k-1}\bar{r}_{k-1}}\frac{h_R}{h_{\bar{a}_{k-1}r_{k-1}}}\end{bmatrix}$, $\mathbf{V}_{k+1} = \begin{bmatrix}\frac{h_R}{h_{\bar{a}_k r_k}} & 0 \\ \sqrt{P}h_{a_k\bar{r}_k}\frac{h_R}{h_{\bar{a}_k r_k}}\end{bmatrix}$

Output: $Y_{SI_{A1}}^k$

1: **for** $i = k$ to $k+1$ **do**
2: $\quad \mathbf{u}_i = \mathbf{s}_i - \alpha h_{\bar{a}_i r_i}\mathbf{V}_i\mathbf{s}_{i-1}$
3: **end for**
4: $y_{SI_{A1}}^k = \mathbf{u}_k^T\left(\begin{bmatrix}0 & -1 \\ 1 & 0\end{bmatrix}\mathbf{u}_{k+1}\right)$

2) *If* $M = 1$, *the two equivalent channel coefficients corresponding to* x_B^k, $D_1^k = -X_{A_1}^k H_{R1}^{k+1}$ *and* $C_1^{k+1} = X_{A_1}^{k+2}H_{R1}^{k+1}$, *are linearly dependent. Thus, only one estimation of* x_A^k *can be estimated from* $Y_{SI_A}^k$ *and* $Y_{SI_A}^{k+1}$. *Accordingly, we make a conjecture that after GSIM1 the STWR system only achieves a diversity order of one, which will be proved in section IV-E.*

B. Decomposed Algorithm

Although Theorem 1 (with Lemma 2) facilitates a GSIM approach, matrix $\text{diag}\{\mathbf{f}\}$ must have full rank. Specifically, if $\exists k \in [1, L+1] \Rightarrow f_k = h_{\bar{a}_k r_k} = 0$, GSIM fails. Fortunately, for each time slot, we have managed to decompose this approach into individual segments (Algorithm 1), which has the following advantages:

1) Algorithm 1 works for any $(k+1)$-th time slot provided that $h_{\bar{a}_k r_k}, h_{\bar{a}_{k-1}r_{k-1}} \neq 0$;
2) Algorithm 1 does not need the eigenvalues of a matrix or matrix-variable convex optimization problems, and thus has a low computational complexity.

GSIM1 is transparent to the relays and can be used with any existing AF-based HD relays without using genuine FD hardware. Moreover, STWR with GSIM1 offers the potential of large-scale relay deployment with simple upgrade and maintenance.

C. Detection Method Against ISI

After GSIM1, the ISI term in (16) must be addressed in order to detect the original signals across two time slots. To this end, we consider efficient data detection method based on the structure of ISI channel $\mathbf{H}_{SI_{A1}}$. Assume QPSK modulation is adopted for $\mathbf{x}_B$. Equ. (16) is first transformed to a real linear model by letting $\bar{\mathbf{y}}_{SI_{A1}} = \begin{bmatrix}\Re\{\mathbf{y}_{SI_{A1}}\} \\ \Im\{\mathbf{y}_{SI_{A1}}\}\end{bmatrix}$, $\bar{\mathbf{H}}_{SI_{A1}} = \begin{bmatrix}\Re\{\mathbf{H}_{SI_{A1}}\} & -\Im\{\mathbf{H}_{SI_{A1}}\} \\ \Im\{\mathbf{H}_{SI_{A1}}\} & \Re\{\mathbf{H}_{SI_{A1}}\}\end{bmatrix}$, $\bar{\mathbf{x}}_B = \begin{bmatrix}\Re\{\mathbf{x}_B\} \\ \Im\{\mathbf{x}_B\}\end{bmatrix}$, $\bar{\mathbf{n}}_{SI_{A1}} = \begin{bmatrix}\Re\{\mathbf{n}_{SI_{A1}}\} \\ \Im\{\mathbf{n}_{SI_{A1}}\}\end{bmatrix}$. Then (16) can be converted to the following model: $\bar{\mathbf{y}}_{SI_{A1}} = \sqrt{P}\bar{\mathbf{H}}_{SI_{A1}}\bar{\mathbf{x}}_B + \bar{\mathbf{n}}_{SI_{A1}}$. Suppose that $\bar{\mathbf{H}}_{SI_{A1}}$ has QR factorization $\mathbf{H}_{SI_{A1}} = [\mathbf{Q}_1 \ \underset{n}{\mathbf{Q}_2}][\mathbf{R} \ \mathbf{0}]^{\text{T}}$, where $[\underset{n}{\mathbf{Q}_1} \ \underset{m-n}{\mathbf{Q}_2}] \in \mathbb{R}^{m\times m}$ is an orthogonal matrix and $\mathbf{R} \in \mathbb{R}^{n\times n}$ is an upper triangular matrix.

Denote $\tilde{\mathbf{y}}_{SI_{A1}} = \mathbf{Q}_1^T \bar{\mathbf{y}}_{SI_{A1}}$ and Babai point detector can thus be simply applied [39], [40]. We could compute

$$c^{SIC}(k) = \frac{1}{\mathrm{R}(k,k)}\left(\tilde{y}_{SI_{A1}}(k) - \sum_{p=k+1}^{2L} R(k,p) x_B^{SIC}(k)\right) \text{ and}$$

$$x_B^{SIC}(k) = \begin{cases} -1, \text{ if } c^{SIC}(k) < 0 \\ 1, \text{ if } c^{SIC}(k) \geq 0 \end{cases},$$

where $k = 2L, \ldots, 1$ and $\sum_{2L+1}^{2L} \cdot = 0$. Finally, the decoding result for QPSK symbol x_B^k is

$$\hat{x}_B^k = x_B^{SIC}(k) + j x_B^{SIC}(k+L).$$

D. Spectral Efficiency Analysis

To analyze the spectral efficiency, we consider the standard propagation model with small-scale Rayleigh fading and distance-dependent path loss given by

$$h_* = r_* \sqrt{d_*^{-\gamma}}, \tag{19}$$

where $* \in \{\bar{a}_k r_k, \bar{b}_k r_k, R\}$ $(k = 1, 2, \cdots, L+1)$, r_* is a zero-mean, unit-variance complex Gaussian variable, d_{ab} is the distance between nodes a and b, and γ is the path loss factor. We use σ_*^2 to denote the variance for channel h_*, and use σ_{**}^2 $(** \in \{A, B\})$ and σ_{***}^2 $(*** \in \{1, 2\})$ to denote the power of AWGN at sources (FN_A and FN_B) and relays (R_1 and R_2) respectively.

1) Spectral Efficiency: Define for the links "$FN_B \rightarrow \{R_1, R_2\} \rightarrow FN_A$" and "$FN_A \rightarrow \{R_1, R_2\} \rightarrow FN_B$" as R_A and R_B respectively. From (16) and Appendix III, we have

$$\begin{aligned} R_A &= \frac{1}{L+1}\log_2\left[\det\left(\mathbf{I}_L + P\mathbf{H}_{SI_{A1}}^{\mathrm{H}}\left(\mathbf{Q}^{-\frac{1}{2}}\right)^{\mathrm{H}}\mathbf{Q}^{-\frac{1}{2}}\mathbf{H}_{SI_{A1}}\right)\right] \\ &= \frac{1}{L+1}\log_2\left(\det\left(\mathbf{I}_L + P\widetilde{\mathbf{H}}_{SI_{A1}}^{\mathrm{H}}\widetilde{\mathbf{H}}_{SI_{A1}}\right)\right) \\ &\approx \frac{\log_2(M_{L-1})}{L+1} + \frac{\sum_{m\in\Omega}\log_2(M_m)}{L+1}, \end{aligned}$$

where $\mathbf{Q} = \mathbb{E}\left[\mathbf{n}_{SI_{A1}}\mathbf{n}_{SI_{A1}}^{\mathrm{H}}\right] = \mathrm{diag}\{e_1, e_2, \cdots, e_k\}$, $M_{L-1} = 1 + \frac{P}{e_A}\sum_{k=L-1}^{L}|D_1^k|^2 + \frac{P^2}{e_A^2}|D_1^{L-1}|^2|D_1^L|^2$, $M_m = 1 + \frac{P}{e_A}\sum_{k=m}^{m+3}|D_1^k|^2 + \frac{P^2}{e_A^2}\left(|D_1^m|^2|D_1^{m+1}|^2 + |D_1^m|^2|D_1^{m+3}|^2 + |D_1^{m+2}|^2|D_1^{m+3}|^2\right)$ and $e_A = e|_{*=A}$ (e_k and e_A are given in (46) and (43).

Similarly, we can derive R_B, the spectral efficiency for the link "$FN_A \rightarrow \{R_1, R_2\} \rightarrow FN_B$":

$$R_B \approx \frac{\log_2(M_{B,L-1})}{L+1} + \frac{\sum_{m\in\Omega}\log_2(M_{B,m})}{L+1}, \tag{20}$$

where $D_{1B}^k = \alpha(x_B^k h_{\bar{b}_{k+1}r_{k+1}} - x_B^{k-1}\alpha h_R h_{\bar{b}_k r_k})h_{a_k\bar{r}_k}$, $e_B = e|_{**=B}$ (see (43)), $M_{B,m} = \frac{P^2}{e_B^2}\left(|D_{1B}^m|^2|D_{1B}^{m+1}|^2 + |D_{1B}^m|^2|D_{1B}^{m+3}|^2 + |D_{1B}^{m+2}|^2|D_{1B}^{m+3}|^2\right) + \frac{P}{e_B}\sum_{k=m}^{m+3}|D_{1B}^k|^2 + 1$ and $M_{B,L-1} = 1 + \frac{P}{e_B}\sum_{k=L-1}^{L+1}|D_{1B}^k|^2 + \frac{P^2}{e_B^2}|D_{1B}^{L-1}|^2|D_{1B}^L|^2$. Thus, the spectral efficiency of the whole system is

$$R \approx R_A + R_B. \tag{21}$$

Numerical spectral efficiency values ($M = 1$) are shown in Section VI-A.

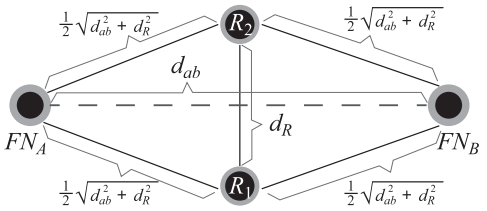


Fig. 2. Distance setting to facilitate the calculation and analysis. d_{ab} is the fixed distance of direct link between two sources.

2) Deployment of the Relays: From Corollary 1 (Appendix III), we have $\frac{\partial e_A}{\partial d_R}$, $\frac{\partial e_B}{\partial d_R} < 0$. To reduce the accumulated noise power, we can increase the distance between the two relays - d_R. However, d_R is limited by the laws of triangles and geometry. On the other hand, increasing d_R results in increasing two-hop total distance, which in turn decreases the spectral efficiency. Thus, choosing optimal d_R for STWR is crucial.

To facilitate the analysis of general deployment issue, we consider a simple symmetrical topology in Fig. 2 with log-distance path loss and thus neglect the small-scale fading. The power of AWGN is assumed to be unit. To analyze the spectral efficiency for one-way link from the TWR system, i.e. R_A, we can define the equivalent signal-to-noise ratio (SNR) term as $S_k := \frac{P|D_k|^2}{e_{**}}$. Based on the definition of D_k, S_k can be viewed as the SNR of equivalent signal propagating itself by two co-channel paths $\alpha h_{\bar{a}_{k+1}r_{k+1}}h_{b_k\bar{r}_k}$ and $-\alpha^2\ h_R h_{\bar{a}_k r_k}h_{b_k\bar{r}_k}$, and the power of equivalent AWGN is e_k (see (46)). Then, based on (43) we have

$$S_k = \begin{cases} S_1 = \dfrac{P\alpha^2\left(\frac{1}{4}\left(d_{\mathrm{ab}}^2 + d_R^2\right)\right)^{-\gamma}\left(1 + \alpha^2 d_R^{-\gamma}\right)}{2 + \alpha^2\left(\frac{1}{2}\sqrt{d_{\mathrm{ab}}^2 + d_R^2}\right)^{-\gamma}}, & k = 1, \\ S = \dfrac{P\alpha^2\left(\frac{1}{4}\left(d_{\mathrm{ab}}^2 + d_R^2\right)\right)^{-\gamma}\left(1 + \alpha^2 d_R^{-\gamma}\right)}{n_s}, & k \neq 1, \end{cases} \tag{22}$$

where $n_s = \alpha^2\ d_R^{-\gamma}\left(2\alpha^2 d_R^{-\gamma} + \alpha^2 + 2\right) + 2\alpha^2\left(1 + \alpha^2\ d_R^{-\gamma}\right)\left(\frac{1}{2}\sqrt{d_{\mathrm{ab}}^2 + d_R^2}\right)^{-\gamma} + 2$. Increasing d_R results in severer path loss but lower noise level. To find the best d_R with (22), the spectral efficiency maximization problem can be formulated as

$$\begin{aligned} \max_{d_R}\ & R_A = \frac{\log_2(\mathcal{R}_1(S))}{L+1} + \frac{\log_2(1 + 4S_1 + 3S_1^2)}{L+1}, \\ \text{s.t. } & \mathcal{R}_1(S) = \left(1 + 3S + S^2\right)\left(1 + 4S + 3S^2\right)^{\frac{L-4}{2}} \\ & d_R > 0,\ S \text{ and } S_1 \text{ are given in (22)}. \end{aligned} \tag{23}$$

If L is sufficiently large, the solution of (23) can be obtained by solving $Q(d_R) = 0$ in (44) (Appendix IV). Note that $\frac{\partial S}{\partial d_R} = 0$ can not be solved in closed-form and we provide numerical results in Section VI-C.

3) Multiplexing Gain: Define R_A^{high} as the rate R_A for high SNR regime, which is dominated by terms with highest order and given as

$$R_A^{high} = \frac{\log_2(SNR_1)}{L+1} + \frac{\sum\limits_{m\in\Omega}\log_2(SNR_m)}{L+1}, \quad (24)$$

where $SNR_1 = \frac{P^2}{e_A^2}|D_1^{L-1}|^2|D_1^L|^2$ and $SNR_m = \frac{P^2}{e_A^2}\left(|D_1^m|^2|D_1^{m+1}|^2 + |D_1^m|^2|D_1^{m+3}|^2 + |D_1^{m+2}|^2|D_1^{m+3}|^2\right)$. Then, (25) is derived from (24) as $\frac{P}{e_A}$ is sufficiently high.

$$R_A^{high} = \frac{L}{L+1}\log_2\left(\frac{P}{e_A}\right). \quad (25)$$

Similarly, we can derive $R_B^{high} = \frac{L}{L+1}\log_2\left(\frac{P}{e_B}\right)$.

Remark 5: From (25), we have the following conclusions.

1) *According to the definition of multiplexing gain, $r := \lim\limits_{SNR\to\infty}\frac{R}{SNR}$, the multiplexing gain for each separated one-way communication is $\frac{L}{L+1}$, approaching 1 (FD multiplexing gain) as L is large enough. Thus, the multiplexing gain of the whole two-way communication system approaches FD multiplexing gain if L is large enough.*
2) *The spectral efficiency at high SNR is dominated by R^{high} = $\frac{L}{L+1}\left(\log_2\left(\frac{P}{e_A}\right)+\log_2\left(\frac{P}{e_B}\right)\right)$.*

E. Diversity Analysis

Based on (16), all the $y_{SI_{A1}}^k$ can be vectorized as $\mathbf{y}_{SI_{A1}} = \sqrt{P}\mathbf{X}_{B1}\mathbf{h}_1 + \mathbf{n}_{SI_{A1}}$, where $\mathbf{h}_1 = \left[D_1^1\ C_1^2\ D_1^2\ C_1^3\ D_1^3\cdots\ D_1^{L-1}\ C_1^L\ D_1^L\right]^{\mathrm{T}}$ and

$$\mathbf{X}_{B1} = \begin{bmatrix} x_B^1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & x_B^1 & x_B^2 & \ddots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & x_B^{L-1} & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & x_B^{L-2} & x_B^{L-1} & x_B^L \end{bmatrix}. \quad (26)$$

$\mathbf{X}_{B1}$ is a $L\times(2L-1)$ matrix consisting of the transmitted symbols and $\mathbf{h}_1$ is the equivalent channel vector. Vector $\mathbf{n}_{SI_{A1}}$ is the correlated noise. To facilitate the analysis, we assume $\mathbf{n}_{SI_{A1}} \sim \mathcal{CN}(0, \mathbf{n}_{SI_{A1}}\mathbf{n}_{SI_{A1}}^{\mathrm{H}})$ [37], [38]. The non-zero elements of $\mathbf{X}_{B1}$ are given as

$$\mathbf{X}_{B1}(\beta, 2\beta-1) = \begin{cases} \mathbf{X}_{B1}(\beta+1, 2\beta) = x_B^\beta, & \text{otherwise}, \\ x_B^L, & \beta = L. \end{cases} \quad (27)$$

We can view $\mathbf{X}_{B1}$ as the equivalent space time code for x_B^k. Supposing $\mathbf{X}_{B1}$ is transmitted, the Chernoff upper bound of PEP, i.e., mistaking $\mathbf{X}_{B1}$ with $\mathbf{X}'_{B1}$, can be derived as

$$P(\mathbf{X}_{B1}\to\mathbf{X}'_{B1}) \le \mathop{\mathbb{E}}_{\mathbf{h}_1}\left[e^{-\frac{P\mathbf{h}_1^{\mathrm{H}}\mathbf{X}_{B\Delta}^{\mathrm{H}}\mathbf{X}_{B\Delta}\mathbf{h}_1}{4\mathrm{tr}(\mathbf{n}_{SI_{A1}})}}\right], \quad (28)$$

where $\mathbf{X}_{B\Delta} = \mathbf{X}_B - \mathbf{X}'_B$. To calculate (28), $\mathbf{h}_1$ can be expressed as $\mathbf{h}_1 = \mathbf{T}_1\mathbf{C}_1\mathbf{v}_1$, where $\mathbf{T}_1$ is a $(2L-1)\times(2L-1)$ matrix $\mathbf{T}_1 = \alpha\,\mathrm{diag}\{[h_{42}\,h_{42}\,h_{41}\cdots h_{41}]\}$ and $\mathbf{C}_1$ is a $(2L-1)\times 2$ matrix $\mathbf{C}_1 = \begin{bmatrix} -X_{A_1}^1 X_{A_1}^3 & 0 & \cdots -X_{A_1}^{L-1}X_{A_1}^{L+1} & 0 \\ 0 & 0 & -X_{A_1}^2\cdots & 0 & 0 & -X_{A_1}^L \end{bmatrix}^{\mathrm{T}}$ and $\mathbf{v}_1 = [h_{62}\ h_{61}]^{\mathrm{T}}$. Hence, (28) can be rewritten as

$$\begin{aligned} &P(\mathbf{X}_{B1}\to\mathbf{X}'_{B1}) \\ &\le \mathop{\mathbb{E}}_{\mathbf{T}_1}\left[\int_{+\infty}^{-\infty}\frac{1}{\pi^2}e^{-\frac{P\mathbf{v}_1^{\mathrm{H}}\mathbf{C}_1^{\mathrm{H}}\mathbf{T}_1^{\mathrm{H}}\mathbf{X}_{B1\Delta}^{\mathrm{H}}\mathbf{X}_{B1\Delta}\mathbf{T}_1\mathbf{C}_1\mathbf{v}_1}{4\mathrm{tr}(\mathbf{n}_{SI_{A1}})}}\right] \\ &= \mathop{\mathbb{E}}_{\mathbf{T}_1}\left[\det{}^{-1}\left(\mathbf{I}_{N-1} + \frac{P\mathbf{C}_1^{\mathrm{H}}\mathbf{T}_1^{\mathrm{H}}\mathbf{X}_{B1\Delta}^{\mathrm{H}}\mathbf{X}_{B1\Delta}\mathbf{T}_1\mathbf{C}_1}{4\mathrm{tr}(\mathbf{n}_{SI_{A1}})}\right)\right] \\ &\le \mathop{\mathbb{E}}_{\mathbf{T}_1}\left[\left(\prod_{k=1}^{r}\lambda_k\right)^{-1}\left(\frac{P}{4\mathrm{tr}(\mathbf{n}_{SI_{A1}})}\right)^{-r}\right], \end{aligned} \quad (29)$$

where λ_k is the k-th non-zero eigenvalue of the matrix $\mathbf{A}\mathbf{A}^{\mathrm{H}}$, r is the rank of $\mathbf{A}\mathbf{A}^{\mathrm{H}}$, and $\mathbf{A} = \mathbf{X}_{B1\Delta}\mathbf{T}_1\mathbf{C}_1$. Then, the diversity gain can be derived as $d_1 = \min_{\mathbf{X}_{B1\Delta}}\mathrm{rank}[\mathbf{X}_{B1\Delta}\mathbf{C}_1]$, as $\mathbf{T}_1$ is full rank. To obtain the minimum rank, let us consider only one codeword is different between $\mathbf{X}_{B1}$ and $\mathbf{X}'_{B1}$. The following two cases are discussed.

1) Codeword x_B^k $(k = 1, 2, \cdots, L-1)$ is different.
   In the equivalent space time code matrix (27), we have two linearly independent columns in $\mathbf{X}_{B1\Delta}$ ($(2k-1)$-th and $2k$-th). For example, if $k = L-1$, we have
   $$\mathbf{X}_{B1\Delta} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \Delta x_B^{L-1} & 0 & 0 \\ 0 & 0 & \cdots & \Delta x_B^{L-1} & 0 \end{bmatrix}, \quad (30)$$
   where $\Delta x_B^{L-1} = x_B^{L-1} - x_B^{'L-1}$, and $(2L-3)$-th and $(2L-2)$-th columns are linearly independent. However, the two corresponding rows of $\mathbf{C}_1$, i.e., the $(2L-3)$-th and $(2L-2)$-th rows, are linearly dependent. Thus, the $\mathbf{X}_{B1\Delta}\mathbf{C}_1$ has a limited rank of one.
2) Codeword x_B^L is different.
   In this case, we have $\mathrm{Rank}(\mathbf{X}_{B1\Delta}) = 1$ and $1 \le \mathrm{rank}[\mathbf{X}_{B1\Delta}\mathbf{C}_1] \le \min\{\mathrm{rank}[\mathbf{X}_{B1\Delta}], \mathrm{rank}[\mathbf{C}_1]\} = 1$. Thus, the rank of $\mathbf{X}_{B1\Delta}\mathbf{C}_1$ is one.

The STWR with GSIM1 achieves diversity order of one for all codewords $x_B^1, x_B^2, \cdots, x_B^L$. The conjecture in Remark 4 is now proved.

V. Case II: $M = 2$

In this section, we increase the value of M from 1 to 2 for GSI mitigation. We give the result of GSI mitigation, design a decomposed algorithm for $M = 2$, and analyze the performance in three aspects - SE, multiplexing gain and diversity order. In this case, STWR retrieves an additional copy of intended signal and the diversity gain can be accordingly recovered to 2.

A. Result for Generalized Self-Interference Mitigation

Based on Theorem 1 (with Lemma 2), the result of GSIM2 is

$$\mathbf{y}_{SI_{A2}} = \sqrt{P}\mathbf{H}_{SI_{A2}}\mathbf{x}_B + \mathbf{n}_{SI_{A2}}, \quad (31)$$

Algorithm 2 Generalized Self-Interference Mitigation ($M=2$) for $(k+1)$-th Time Slot

Input: $\boldsymbol{V}_k = \boldsymbol{V}_{k+2} = \begin{bmatrix} 0 & \frac{h_R}{\sqrt{P}} & 0 \\ \frac{h_{a_{k-1}\bar{r}_{k-1}}}{\alpha h_R} h_{a_k\bar{r}_k} & & \frac{h_R}{\sqrt{P}h_{\bar{a}_k r_k}} \end{bmatrix}$, $\boldsymbol{s}_{k-1} = \begin{bmatrix} x_A^{k-1} \\ x_A^{k-2} \\ y_A^{k-2} \end{bmatrix}$, $\boldsymbol{s}_k = \begin{bmatrix} x_A^k \\ y_A^k \end{bmatrix}$, $\boldsymbol{s}_{k+1} = \begin{bmatrix} x_A^{k+1} \\ x_A^k \\ y_A^k \end{bmatrix}$, $\boldsymbol{s}_{k+2} = \begin{bmatrix} x_A^{k+2} \\ y_A^{k+2} \end{bmatrix}$

Output: $Y_{SI_{A1}}^k$

1: **for** $t=0$ to 1 **do**
2: $\quad i = k+2t$
3: $\quad \boldsymbol{u}_i = \boldsymbol{s}_i - \sqrt{P}\alpha^2\ h_{\bar{a}_i r_i} h_R \boldsymbol{V}_i \boldsymbol{s}_{i-1}$
4: **end for**
5: $Y_{SI_{A1}}^k = \mathbf{u}_k^T \left(\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \mathbf{u}_{k+2} \right)$

where $\mathbf{H}_{SI_{A2}} = \mathbf{H}_{Q_M^{\bar{a}b}}|_{M=2}$ (see (32), shown at the top of the next page) $\mathbf{n}_{SI_{A2}} = \left(\mathbf{H}_{Q_M^{\bar{r}}} \mathbf{n}_{\bar{r}} + \mathbf{Q}_M \mathbf{n}_A \right)|_{M=2} = \left[N_{SI_{A2}}^1 \ \cdots \ N_{SI_{A2}}^L \right]^{\mathrm{T}}$, $N_{SI_{A2}}^k = X_{A_2}^{k+1} N_{A2}^k - X_{A_2}^k N_{A2}^{k+1}$, $X_{A2}^k = v_{x_A}^k|_{M=2} = x_A^k - \alpha^2\ h_R^2\ x_A^{k-2}$, $H_{R2}^k = \alpha^2\ h_R\ h_{\bar{a}_k r_k} h_{b_k \bar{r}_k}$, $B_2^k = X_{A_2}^{k+1} H_{R2}^k$, $C_2^k = \alpha h_{b_{k-1}\bar{r}_{k-1}} (X_{A_2}^{k+1} h_{\bar{a}_k r_k} - X_{A_2}^k \alpha^2\ h_R h_{\bar{a}_{k-1} r_{k-1}})$, $D_2^k = -X_{A_2}^k H_{R1}^{k+1}$ and $N_{SI_{A2}}^k = X_{A_2}^{k+1} N_{A2}^k - X_{A_2}^k N_{A2}^{k+1}$.

B. Decomposed Algorithm and Decoding

To relax the requirement of 'full-rank $\mathrm{diag}\{\mathbf{f}\}$', GSIM2 can be decomposed into individual segments of algorithm (see Algorithm 2), which can be implemented for $\forall(k+1)$-th time slots satisfying $h_{\bar{a}_k r_k}, h_{\bar{a}_{k-1} r_{k-1}} \neq 0$ with linear computational complexity.

We can adopt the fast decoding method provided in section IV-C because $\mathbf{H}_{SI_{A2}}$ is a lower triangular matrix.

C. SE and Multiplexing Gain

We adopt the same method in section IV-D to analyze the SE and multiplexing gain.

1) Spectral Efficiency: The SE for direction "$FN_B \rightarrow \{R_1, R_2\} \rightarrow FN_A$" is

$$R_A = \frac{1}{L+1} \log_2 \left(\det \left(\mathbf{I}_L + P \widetilde{\mathbf{H}}_{SI_{A2}}^{\mathrm{H}} \widetilde{\mathbf{H}}_{SI_{A2}} \right) \right), \quad (33)$$

where b_k and $\widetilde{\mathbf{H}}_{SI_{A2}}^{\mathrm{H}}$ is given in Appendix III and (32) respectively.

Matrix $\widetilde{\mathbf{H}}_{SI_{A2}}$ can be divided into two kinds of blocks, i.e.,

$$\widehat{H}_m = \begin{bmatrix} \frac{D_2^m}{\sqrt{b_m}} & \frac{C_2^{m+1}}{\sqrt{b_{m+1}}} & \frac{B_2^{m+2}}{\sqrt{b_{m+2}}} & 0 \\ 0 & \frac{D_2^{m+1}}{\sqrt{b_{m+1}}} & \frac{C_2^{m+2}}{\sqrt{b_{m+2}}} & \frac{B_2^{m+3}}{\sqrt{b_{m+3}}} \end{bmatrix}^{\mathrm{T}},$$

$$\widehat{H}_{L-1} = \begin{bmatrix} \frac{D_2^{L-1}}{\sqrt{b_{L-1}}} & 0 \\ \frac{C_2^L}{\sqrt{b_L}} & \frac{D_2^L}{\sqrt{b_L}} \end{bmatrix}, \quad (34)$$

where $m \in \Omega = \{1, 3, \cdots, L-3\}$. Using Corollary 1 in Appendix III, we can estimate the determinant as $DET_{SI_{A2}} \approx \widehat{M}_{**,L-1} \prod_{m\in\Omega} \widehat{M}_{**,m}$, where $** = A$, b_{**} is given in (43), $\widehat{M}_{**,m}$ is $\widehat{M}_{**,m} = \frac{P^2}{b_{**}^2}(\mathcal{M}_1 - 2\mathcal{M}_2) + \frac{P}{b_{**}}\mathcal{M}_3 + 1$, $\mathcal{M}_1 = \left|B_2^{m+3}\right|^2 \left|B_2^{m+2}\right|^2 + \left|B_2^{m+3}\right|^2 \left|C_2^{m+1}\right|^2 + \left|B_2^{m+2}\right|^2 \left|D_2^{m+1}\right|^2 + \left|B_2^{m+3}\right|^2 |D_2^m|^2 + \left|C_2^{m+1}\right|^2 \left|C_2^{m+2}\right|^2 + \left|C_2^{m+2}\right|^2 |D_2^m|^2 + |D_2^m|^2 \left|D_2^{m+1}\right|^2$, $\mathcal{M}_2 = \left|B_2^{m+2}\right|^2 \left|C_2^{m+1}\right|^2 \left|C_2^{m+2}\right|^2 \left|D_2^{m+1}\right|^2$, $\mathcal{M}_3 = \left|B_2^{m+2}\right|^2 + \left|B_2^{m+3}\right|^2 + \left|C_2^{m+1}\right|^2 + \left|C_2^{m+2}\right|^2 + |D_2^m|^2 + \left|D_2^{m+1}\right|^2$, and $\widehat{M}_{**,L-1} = 1 + \frac{P}{b_{**}} \sum_{k=L-1}^{L} |D_2^k|^2 + \frac{P^2}{b_{**}^2} |D_2^{L-1}|^2 |D_2^L|^2$. Thus, the SE R_{**} is

$$\begin{aligned} R_{**} &= \frac{1}{L+1} \log_2 (\det (DET_{SI_{A2}})) \\ &\approx \frac{\log_2 \left(\widehat{M}_{**,L-1} \right)}{L+1} + \frac{\sum_{m\in\Omega} \log_2 \left(\widehat{M}_{**,m} \right)}{L+1}. \end{aligned} \quad (35)$$

The SE for the whole system is $R \approx \sum_{**\in\{A,B\}} R_{**}$. The numerical result of SE ($M=2$) is shown in section VI-A.

2) Multiplexing Gain: Define R_{**}^{high} ($** \in \{A, B\}$) as the SE for high SNR regime, which is dominated by terms with highest order. If $\frac{P}{b_{**}}$ is sufficiently high, (35) can be derived as

$$R_{**}^{high} = \frac{L}{L+1} \log_2 \left(\frac{P}{e_{**}} \right). \quad (36)$$

Remark 6: According to the definition $r := \lim_{SNR\to\infty} \frac{R}{SNR}$, the multiplexing gain for each separated one-way communication is $\frac{L}{L+1}$, approaching 1 (FD multiplexing gain) as L is large enough. Thus, the multiplexing gain of the whole two-way communication system approaches ideal FD-TWR. Compared with GSIM1, GSIM2 imposes no multiplexing loss on STWR either.

D. Diversity Analysis

The derivation of PEP is similar to section IV-E, and we thus only focus on the diversity analysis. Based on (31), all the $Y_{SI_{A2}}^k$ can be vectorized as $\mathbf{y}_{SI_{A2}} = \sqrt{P}\mathbf{X}_{B2}\mathbf{h}_2 + \mathbf{n}_{SI_{A2}}$, where $\mathbf{X}_{B2}$ is a $L \times (3L-3)$ matrix given in (37), shown at the top of the next page, the equivalent channel vector $\mathbf{h}_2$ is

$$\mathbf{h}_2 = \left[D_2^1\ C_2^2\ D_2^2\ B_2^2\ C_2^3\ D_2^3 \cdots B_2^L\ C_2^L\ D_2^L \right]^{\mathrm{T}}, \quad (38)$$

and $\mathbf{n}_{SI_{A2}}$ is the correlated noise. Similar to [37] and [38], we assume $\mathbf{n}_{SI_{A2}} \sim \mathcal{CN}(0, \mathbf{n}_{SI_{A2}} \mathbf{n}_{SI_{A2}}^{\mathrm{H}})$.

Remark 7: The non-zero elements of $\boldsymbol{X}_{B2}$ are

- $\boldsymbol{X}_{B2}(1,1) = \boldsymbol{X}_{B2}(2,2) = \boldsymbol{X}_{3,4}$;
- $\boldsymbol{X}_{B2}(\beta{+}1, 3\beta{-}1) = \boldsymbol{X}_{B2}(\beta{+}1, 3\beta{-}1) = \boldsymbol{X}_{B2}(\beta{+}2, 3\beta{+}1) = x_B^\beta$, *if* $\beta = 2, , 3, \cdots, L-1$;
- x_B^L, *if* $\beta = L$.

We can view $\boldsymbol{X}_{B2}$ as the equivalent space time code for x_B^k.

$\mathbf{h}_2$ can be expressed as $\mathbf{h}_2 = \mathbf{T}_2 \mathbf{C}_2 \mathbf{v}_2$, where $\mathbf{T}_2 = \alpha\, \mathrm{diag}\{[h_{42}\ h_{42}\ h_{41}]\}$, $\mathbf{C}_2$ is given in (39), shown at the top of the next page, and $\mathbf{v}_2 = [h_{62}\ h_{61}]^{\mathrm{T}}$. Hence, we have $\mathbf{A} = \mathbf{X}_{B2\Delta} \mathbf{T}_2 \mathbf{C}_2$, where $\mathbf{X}_{B2\Delta} = \mathbf{X}_{B2} - \mathbf{X}'_{B2}$ is the space

$$\mathbf{H}_{SI_{A2}}=\begin{bmatrix} D_2^1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ C_2^2 & D_2^2 & 0 & \cdots & 0 & 0 & 0 \\ B_2^3 & C_2^3 & D_2^3 & \ddots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ddots & D_2^{L-2} & 0 & 0 \\ 0 & 0 & 0 & \cdots & C_2^{L-1} & D_2^{L-1} & 0 \\ 0 & 0 & 0 & \cdots & B_2^L & C_2^L & D_2^L \end{bmatrix}, \quad \widetilde{\mathbf{H}}_{SI_{A2}}=\begin{bmatrix} \frac{D_2^1}{\sqrt{b_1}} & 0 & 0 & \cdots & 0 & 0 \\ \frac{C_2^2}{\sqrt{b_2}} & \frac{D_2^2}{\sqrt{b_2}} & 0 & \cdots & 0 & 0 \\ \frac{B_2^3}{\sqrt{b_3}} & \frac{C_2^3}{\sqrt{b_3}} & \frac{D_2^3}{\sqrt{b_3}} & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{D_2^{L-1}}{\sqrt{b_2}} & 0 \\ 0 & 0 & 0 & \cdots & \frac{C_2^L}{\sqrt{b_3}} & \frac{D_2^L}{\sqrt{b_3}} \end{bmatrix} \tag{32}$$

$$\mathbf{X}_{B2}=\begin{bmatrix} x_B^1 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & x_B^1 & x_B^2 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & x_B^1 & x_B^2 & \ddots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & x_B^{L-1} & x_B^L \end{bmatrix}, \quad \mathbf{X}_{B2\Delta}=\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \Delta x_B^2 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \Delta x_B^2 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \Delta x_B^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \tag{37}$$

$$\mathbf{C}_2=\begin{bmatrix} -X_{A_2}^1 & X_{A_2}^3\alpha h_R & X_{A_2}^3 & 0 & 0 & X_{A_2}^4 & -X_{A_2}^3 & \cdots & 0 \\ 0 & 0 & -X_{A_2}^2\alpha h_R & -X_{A_2}^2 & X_{A_2}^4\alpha h_R & -X_{A_2}^3\alpha h_R & 0 & \cdots & -X_{A_2}^L \end{bmatrix}^{\mathrm{T}}, \tag{39}$$

time code difference matrix. Then, the diversity gain can be derived as [41]

$$d_2 = \min_{\mathbf{X}_{B2\Delta}} \operatorname{rank}[\mathbf{X}_{B2\Delta}\mathbf{C}_2], \tag{40}$$

as $\mathbf{T}_2$ is full rank. To obtain the minimum rank in (40), let us consider only one code word is different between $\mathbf{X}_{B2}$ and $\mathbf{X}'_{B2}$. The following three cases are discussed.

1) Codeword x_B^1 is different.
   Two linearly independent rows can be found in the 1-th, 2-th and 4-th rows of $\mathbf{C}_2$. Thus, the $\mathbf{X}_{B2\Delta}\mathbf{C}_2$ has a rank of two.
2) Codeword x_B^k $(k = 2, 3, \cdots, L-1)$ is different.
   Two linearly independent rows can be found in the three corresponding rows of $\mathbf{C}_2$, i.e., the 3-th, 5-th and 7-th rows. Thus, the $\mathbf{X}_{B2\Delta}\mathbf{C}_2$ has a rank of two.
3) Codeword x_B^L is different.
   In this case, we have $\text{Rank}(\mathbf{X}_{B2\Delta}) = 1$ and $1 \leq \operatorname{rank}[\mathbf{X}_{B2\Delta}\mathbf{C}_2] \leq \min\{\operatorname{rank}[\mathbf{X}_{B2\Delta}], \operatorname{rank}[\mathbf{C}_2]\} = 1$. Thus, the rank of $\mathbf{X}_{B2\Delta}\mathbf{C}_2$ is one.

Thus, with GSIM2, the STWR achieves diversity order of two for codewords $x_B^1, \cdots, x_B^{L-1}$, and diversity order of one for x_B^L.

VI. Numerical Results

In the following simulations, we set both the AWGN terms and r_* in (19) with same variance $\sigma^2 = 1$. Frame length is $L = 32$. BPSK modulation is adopted.

A. Spectral Efficiency With Fixed Relay Distance

The distance between source and relay is fixed $d_{ab} = 3$, while the relay distance is $d_R = 2.5$. Path-loss factor is $\gamma = 3$. E_b/N_0 is defined as P/σ^2.

In Fig. 3(a), we compare the spectral efficiencies of STWR (after GSMI), conventional SI cancellation achieving a 3 dB interference-to-noise ratio (INR) [16], the ideal FD-TWR without SI, and conventional half-duplex two-way relaying (HD-TWR) with perfect interference cancellation. Importantly, note that the spectral efficiency of STWR ($d_R = 5$) is very close (the gap is less than 3 bit/s/Hz) to that of the ideal FD-TWR (free of SI), which validates Remark 5. The spectral efficiency gap between STWR and ideal FD-TWR is expected because the frame length ($L = 32$) is not sufficiently large (Remark 5). On the other hand, the spectral efficiency approximation, (21), is accurate at high SNR. Moreover, STWR achieves twice the spectral efficiency of HD-TWR, and it dramatically outperforms conventional FD-TWR (with 3 dB SNR performance for SI cancellation).

Fig. 3(b) shows that after GSIM1, the spectral efficiency of STWR approaches ideal FD-TWR (the gap is less than 4 bit/s/Hz). Although we showed that its multiplexing gain is $\frac{L}{L+1}$, GSIM2 loses spectral efficiency compared to GSIM1 (less than 2 bit/s/Hz). This result is unsurprising because GSIM2 is designed to cancel part of the interference where more accumulated noise term is retained (Fig. 5(b)). However, GSIM2 achieves higher diversity order than GSIM1.

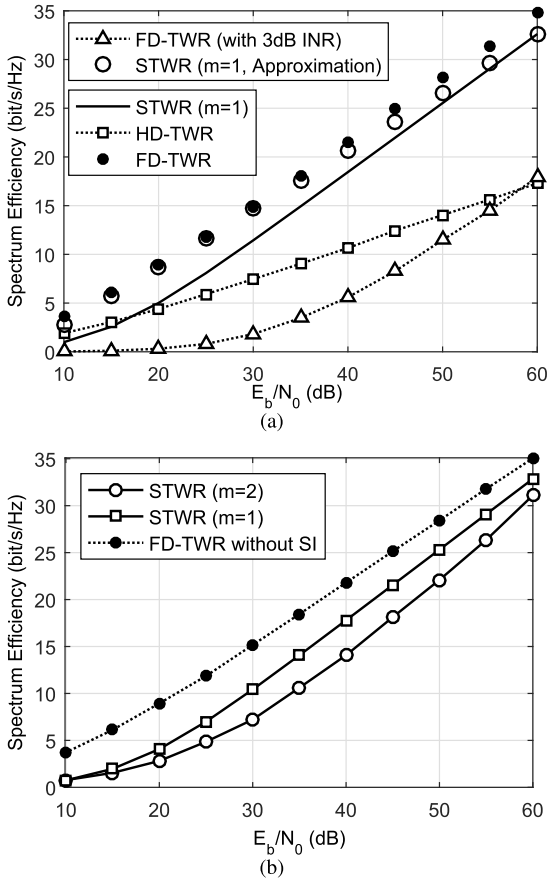


Fig. 3. (a) Spectral efficiencies of STWR, conventional HD-TWR, FD-TWR (3 dB INR) and ideal FD-TWR without SI. (b) Spectral efficiency of STWR with GSIM1 and GSIM2, and ideal FD-TWR without SI.

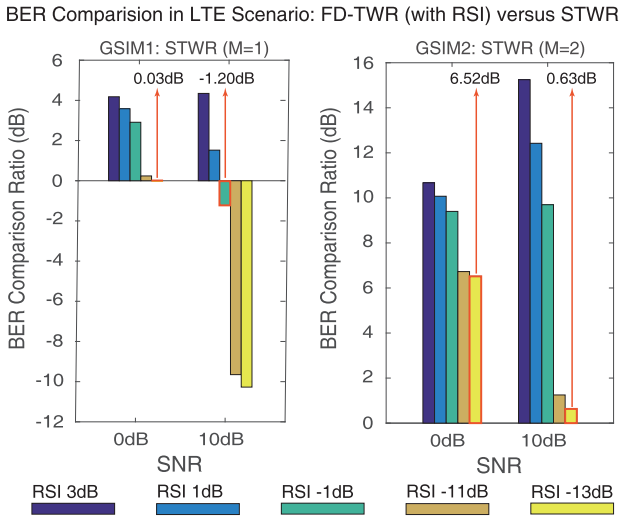


Fig. 4. Bit error ratio (BER) comparison of STWR and FD-TWR (with RSI from [3,1,-1,-11,-13]dB) in a LTE scenario with SNR from [0,10] dB. BER comparison ratio is measured by BER-RSI/BER-m (dB) where BER-m is the BER of STWR (m=1 or 2) and BER-RSI is the BER of FD-TWR with different residual self-interference (bars with different colors).

B. Bit Error Ratio Comparison in LTE Scenario With Fixed Relay Distance

We now further evaluate the bit error ratio (BER) performance in Long Term Evolution (LTE) scenario with key parameters from [42] and [43]. Note that FD radio has not been applied in realistic LTE systems, and thus we use simulation method to evaluate the FD-TWR by combining conventional Rician fading RSI channel [44] with simultaneous transmission and reception functionality of two antennas. The Rician parameter K is chosen as 5dB [44]. In the simulations, the fading channel is generated from 3GPP LTE Release 10 as EPA 5Hz 2×2 MIMO model. Path loss (dB) follows Urban Macro (UMa) LOS as PL $= 22.0 \log_{10}(d) + 28.0 + 20 \times \log_{10}(fc)$, where the carrier frequency fc for evaluation (representative of IMT bands) is 2 GHz. The distances between two HD relay nodes (for STWR) and antennas of each FD transceiver are 10m and 10cm respectively, and the distance between relay and source is 50m. The peak date rate over each link is 38.7 Mb/s per antenna using QPSK modulation. Before transmission, symbols are encoded by orthogonal space time block coding (OSTBC) for a 2×2 MIMO scenario.

In Fig. 4, we use BER comparison ratio $(10 \log_{10} \frac{\text{BER-RSI}}{\text{BER-m}})$ to show the performance gaps between STWR and conventional SI cancellation achieving different RSI INR ratio (dB). At the SNR of 0 dB, both GSIM1 and GSIM2 perform better than -13 dB RSI scheme, i.e., the proposed schemes could achieve a RSI INR ratio lower than -13 dB in terms of BER. At the SNR of 10 dB, GSIM2 could achieve better BER performance than conventional SI cancellation with -13 dB INR ratio, while the BER of GSIM1 is close to that of conventional SI cancellation with -1 dB INR ratio.

C. Deployment Issue and Accumulated Noise With Varying Relay Distance

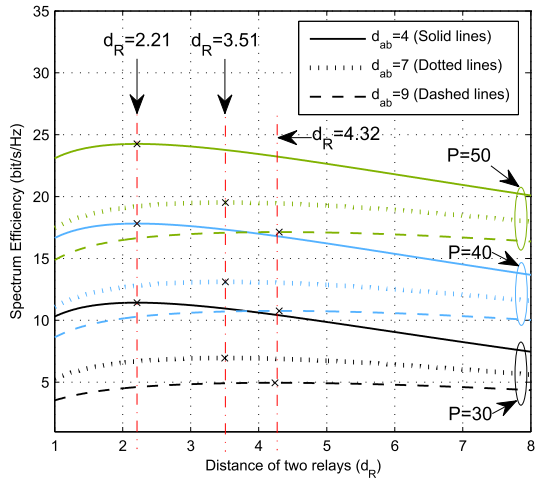
We now vary the relay distance. Path-loss factor is set as $\gamma = 2$. The distances d_R and d_{ab} are described in Fig. 2.

1) Relay Deployment Issue: Based on the analysis in section IV-D and the numerical results in Fig. 5(a) ($d_{ab} \in \{4, 7, 9\}$), we find the following observations and conclusions.

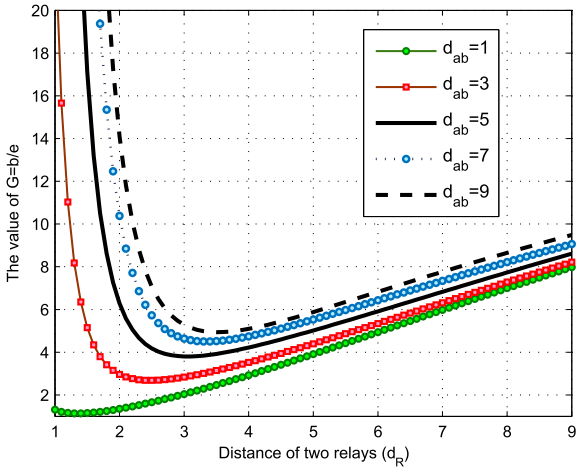
- The value of P does not impact the solution of $\frac{\partial S}{\partial d_R} = 0$. Thus, importantly, transmit power is not the key factor to be considered in the optimal deployment (of relative distance between two relays) for STWR. Fig. 5(a) verifies that if d_{ab} is fixed, optimal d_R does not change with P.
- If P is fixed, Fig. 5(a) shows that the spectral efficiency is not monotone decreasing with d_R and larger d_{ab} results in a larger value of optimal d_R. Thus, optimal deployment depends on the the distance between the two source nodes.

2) Accumulated Noise: To compare the accumulated noise for GSIM2 and GSIM1 (e_k and b_k), we define $\mathcal{G} = \frac{b}{e}$ (Corollary 1, Appendix III) to measure the ratio of b to e. The values of $\mathcal{G}$ is shown in Fig. 5(b), where the distance between two sources is $d_{ab} \in \{1, 3, 5, 7, 9\}$ and $\frac{E_b}{N_0}$ is fixed at 10 dB.

If $\mathcal{G}$ is larger than one, GSIM2 accumulates more noise than GSIM1 and vice versa. When transmit power P is fixed, Fig. 5(b) shows that the minimum value of $\mathcal{G}$ is higher than 1. However, we may choose the distances between two relays



(a)



(b)

Fig. 5. (a) Spectral efficiency for different P, d_R and d_{ab}. (b) The value of $\frac{b}{e}$, with different d_R and d_{ab}.

(d_R) and between two sources (d_{ab}) to reduce the negative effect to some extent ($\mathcal{G}$ approaching 1 is possible).

VII. Conclusion

This paper has proposed a novel STWR system and RSI mitigation method. This new system circumvents the challenges of complex full-duplex hardware and instead employs two half-duplex relays to mimic a FD-TWR. Efficient, low-complexity RSI mitigation algorithms are also designed. The theoretical and numerical results show that STWR achieves:1) the same spectral efficiency as that of ideal FD-TWR (free of SI) when the distance between the two relays is optimally chosen, 2) a diversity order of two is achieved with the second self-interference mitigation algorithm, and 3) higher data rates are achieved with the first self-interference mitigation algorithm. Overall, these benefits may be translated into cost savings and spectral efficiency gains. The STWR concept may thus help the large-scale relay deployment in future networks.

Appendix I: Proof of Theorem 1

Apply the GSI mitigation: $\mathbf{y}_{SI_{AM}} = \mathbf{Q}_M(\mathbf{y}_A - \sqrt{P}\mathbf{H}_{\bar{a}a}\mathbf{x}_A) = \sqrt{P}\mathbf{Q}_M(\mathbf{H}_{\bar{a}b}\mathbf{x}_B + h_A\mathbf{x}_A) + \mathbf{Q}_M\mathbf{H}_{\bar{r}}\mathbf{n}_{\bar{r}} + \mathbf{Q}_M\mathbf{n}_A$. Based on (14), $\mathbf{H}_{Q^\star_M}$ is $\mathbf{H}_{Q^\star_M} = \mathbf{Q}_M\mathbf{H}_\star = \mathbf{G}\mathbf{H}'_\star$, where the entries of $\mathbf{H}'_\star$ are derived as

$$H'_\star(I,J) = \begin{cases} H_\star(I,J), & I \le M, \\ \dfrac{f_{I_1}H_\star(I_1+M,J_1)}{f_I} - \dfrac{g_{I_1+M}H_\star(I_1,J_1)}{f_I}, & I > M. \end{cases} \tag{41}$$

Then, (42), shown at the top of the next page, can be derived from (41) and Lemma 1. Due to $\mathbf{G}\mathbf{v}^{x_A} = \mathbf{0}_{L\times 1}$ in (14), we obtain

$$\begin{aligned} &\sqrt{P}h_A\mathbf{Q}_M\mathbf{x}_A \\ &= \sqrt{P}h_A\mathbf{G}\mathbf{x}_A - \sqrt{P}h_A\mathbf{G}\begin{bmatrix}\mathbf{0}_{M\times(L-M+1)} & \mathbf{0}_{M\times M} \\ \mathrm{diag}\{\mathbf{v}^{gf}\} & \mathbf{0}_{(L-M+1)\times M}\end{bmatrix}\mathbf{x}_A \\ &= \sqrt{P}h_A\mathbf{G}\mathbf{v}^{x_A} = \mathbf{0}_{L\times 1}. \end{aligned}$$

Then, we obtain $\mathbf{y}_{SI_{AM}} = \mathbf{H}_{Q^{\bar{a}r}_M}\mathbf{x}_B + \mathbf{H}_{Q^{\bar{r}}_M}\mathbf{n}_{\bar{r}} + \mathbf{Q}_M\mathbf{n}_A$.

Appendix II: Lemma 3

Lemma 3: If $\mathrm{diag}\{\boldsymbol{a}\} = [a_1\ a_2 \cdots\ a_N]$ *and* $\mathrm{diag}\{\boldsymbol{b}\} = [b_1\ b_2\ \cdots\ b_N]$ *are two vectors of the same length* N*, we have* $\mathrm{diag}\{\boldsymbol{a}\}\boldsymbol{b} = \mathrm{diag}\{\boldsymbol{b}\}\boldsymbol{a}$.

Proof: $\mathrm{diag}\{\mathbf{a}\}\mathbf{b} = \mathrm{diag}\{\mathbf{b}\}\mathbf{a}$ can be easily derived by $\mathrm{diag}\{\mathbf{a}\}\mathbf{b} = [a_1\ b_1\ a_2\ b_3\ \cdots\ a_N b_N]$ and $\mathrm{diag}\{\mathbf{b}\}\mathbf{a} = [a_1\ b_1\ a_2\ b_3\ \cdots\ a_N b_N]$. ■

Appendix III: Derivation of (20)

In (20), $\mathbf{Q} = \mathbb{E}\left[\mathbf{n}_{SI_{A1}}\mathbf{n}^{\mathrm{H}}_{SI_{A1}}\right]$ is a Hermitian symmetric matrix given by

$$\begin{bmatrix} e_1 & f_1 & g_1 & 0 & \cdots & 0 \\ f_1^* & e_2 & f_2 & g_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & f^*_{L-1} & e_L \end{bmatrix}, \tag{45}$$

where $e_k = \mathbb{E}\left[N^k_{SI_A}N^{k*}_{SI_A}\right]$, $f_k = \mathbb{E}\left[N^k_{SI_A}N^{*k+1}_{SI_A}\right]$ and $g_k = \mathbb{E}\left[N^k_{SI_A}N^{*k+2}_{SI_A}\right]$.

Expanding e_k and using Lemma 4 and Lemma 5, e_k can be simplified as

$$e_k = \begin{cases} 2\sigma_A^2 + \alpha^2\sigma^2_{\bar{r}_k}\sigma^2_{\bar{a}_{k+1}r_{k+1}}, & k = 1, \\ \mathbb{E}\left[N^k_{SI_A}N^{k*}_{SI_A}\right], & k \ne 1. \end{cases} \tag{46}$$

Similarly, we have $f_k = g_k = 0$. Following a same procedure, we can derive b_k in (32). Considering the normalized channel matrix $\widetilde{\mathbf{H}}_{SI_{A1}} = \mathbf{Q}^{-1/2}\mathbf{H}_{SI_{A1}}$, the determinant can be approximated [34, eq. (44)] as

$$\begin{aligned} DET_{SI_{A1}} &:= \det\left(\mathbf{I}_L + P\widetilde{\mathbf{H}}^{\mathrm{H}}_{SI_{A1}}\widetilde{\mathbf{H}}_{SI_{A1}}\right) \\ &\approx \det(\mathbf{I}_2 + P\widetilde{H}_{L-1}\widetilde{H}^{\mathrm{H}}_{L-1})\prod_{m\in\Omega}[\det(\mathbf{I}_3 + P\widetilde{H}_m\widetilde{H}^{\mathrm{H}}_m)], \end{aligned} \tag{47}$$

$$\mathbf{H}'_{\star}=\begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ H_{\star}(2,1) & 0 & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ H_{\star}(M+1,1) & H_{\star}(M+1,2) & \cdots & H_{\star}(M+1,M) & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & H_{\star}(L+1,L-M+1) & \cdots & H_{\star}(L+1,L) \end{bmatrix} \tag{42}$$

$$e_{**}=\begin{cases} 2\sigma_{**}^2+\alpha^2\sigma_2^2 d_{\bar{s}r}^{-\gamma}, & k=1, \\ \alpha^2(\sigma_1^2+\sigma_2^2)\left(1+\alpha^2 d_R^{-\gamma}\right)d_{\bar{s}r}^{-\gamma}+\sigma_*^2\left(\alpha^2 d_R^{-\gamma}\left(2+\alpha^2\right)+2(1+\alpha^4 d_R^{-2\gamma})\right), & k\neq 1. \end{cases} \tag{43}$$

$$Q(d_R)=\frac{\alpha^2 2^{\gamma+1}\left(d_{\text{ab}}^2+d_R^2\right)^{-\frac{\gamma}{2}}\left(\alpha^2+\frac{\alpha^2 d_R^2+d_R^{\gamma+2}}{d_{\text{ab}}^2+d_R^2}\right)+\alpha^2(4\alpha^2 d_R^{-\gamma}+\alpha^2+2)}{\alpha^2 2^{\gamma+1}\left(d_{\text{ab}}^2+d_R^2\right)^{-\frac{\gamma}{2}}\left(\alpha^2 d_R^{-\gamma}+1\right)+\alpha^2 d_R^{-\gamma}\left(2\alpha^2 d_R^{-\gamma}+\alpha^2+2\right)+2}-\frac{\alpha^2}{\alpha^2 d_R^{-\gamma}+1}-\frac{2d_R^{\gamma+2}}{d_{\text{ab}}^2+d_R^2} \tag{44}$$

where $\Omega = \{1,3,\cdots,L-3\}$ and $\widetilde{H}_m = \begin{bmatrix} \frac{D_1^m}{\sqrt{e_m}} & \frac{C_1^{m+1}}{\sqrt{e_{m+1}}} & 0 \\ 0 & \frac{D_1^{m+1}}{\sqrt{e_{m+1}}} & \frac{C_1^{m+2}}{\sqrt{e_{m+2}}} \end{bmatrix}^{\mathrm{T}}$, $\widetilde{H}_{L-1} = \begin{bmatrix} \frac{D_1^{L-1}}{\sqrt{e_{L-1}}} & 0 \\ \frac{C_1^L}{\sqrt{e_L}} & \frac{D_1^L}{\sqrt{e_L}} \end{bmatrix}$. Using (18) and Corollary 1 (to simplify the calculation, we assume $d_{\bar{a}_k r_k} = d_{\bar{a}_{k+1} r_{k+1}} = d_{\bar{a}r}$), for $\forall m \in \Omega$ we have $\det(\mathbf{I}_3 + P\widetilde{H}_{L-1}\widetilde{H}_{L-1}^{\mathrm{H}}) = M_{L-1}$ and $\det(\mathbf{I}_3 + P\widetilde{H}_m\widetilde{H}_m^{\mathrm{H}}) = M_m$. Then, $DET_{SI_{A1}}$ can be approximated as $DET_{SI_{A1}} \approx M_{L-1}\prod_{m\in\Omega} M_m$. Thus, the SE is (20).

Lemma 4: if h_M is given by (19),[6] *we can have the following statistical expectations:* $\mathbb{E}[h_M] = \mathbb{E}[h_M^2] = \mathbb{E}[h_M^{*2}] = \mathbb{E}\left[\Re[h_M^2]\right] = \mathbb{E}[h_M|h_M|^2] = \mathbb{E}[h_M^*|h_M|^2] = 0$, $\mathbb{E}[|h_M|^2] = \sigma_M^2 = d_M^{-\gamma}$, $\mathbb{E}[|h_M|^4] = \mathbb{E}[|h_M^2|^2] = 3d_M^{-2\gamma}$, $\mathbb{E}[|h_M^3|^2] = 15d_M^{-3\gamma}$ *and* $\mathbb{E}[|h_M^4|^2] = 24d_M^{-4\gamma}$.

Proof: h_M satisfies $h_M \sim \mathcal{CN}(0, d_M^{-\gamma})$, and we have $\mathbb{E}[h_M] = 0$, $\mathbb{E}[|h_M|^2] = \sigma_M^2 = d_M^{-\gamma}$. Expand h_M as $h_M = p + qi$, where p and q are independent and satisfies $p, q \sim \mathcal{N}(0, \frac{d_M^{-\gamma}}{2})$. We have $\mathbb{E}[p^2] = \mathbb{E}[q^2] = \frac{d_M^{-\gamma}}{2}$, $\mathbb{E}[p^3] = \mathbb{E}[q^3] = 0$, $\mathbb{E}[p^4] = \mathbb{E}[q^4] = \frac{3d_M^{-2\gamma}}{4}$, $\mathbb{E}[p^6] = \mathbb{E}[q^6] = \frac{15d_M^{-3\gamma}}{8}$, and $\mathbb{E}[p^8] = \mathbb{E}[q^8] = \frac{105d_M^{-4\gamma}}{16}$. Then, we derive that $\mathbb{E}[h_M^2] = \mathbb{E}[p^2 - q^2 + 2qpi] = 0$, $\mathbb{E}[h_M^{*2}] = \mathbb{E}[p^2 - q^2 - 2qpi] = 0$, $\mathbb{E}\left[\Re[h_M^2]\right] = \mathbb{E}[p^2 - q^2] = 0$, $\mathbb{E}[h_M|h_M|^2] = \mathbb{E}[p^3 + pq^2 + (p^2q + q^3)i] = 0$, $\mathbb{E}[h_M^*|h_M|^2] = \mathbb{E}[p^3 + pq^2 - (p^2q + q^3)i] = 0$, $\mathbb{E}[|h_M^2|^2] = \mathbb{E}[p^4 + q^4 + 2p^2q^2] = 2d_M^{-2\gamma}$, $\mathbb{E}[|h_M^3|^2] = \mathbb{E}[p^6 + 3\ p^4\ q^2 + 3\ p^2\ q^4 + q^6] = 6d_M^{-3\gamma}$, and $\mathbb{E}[|h_M^4|^2] = \mathbb{E}[p^8 + 4\ p^6\ q^2 + 6\ p^4\ q^4 + 4\ p^2\ q^6 + q^8] = 24d_M^{-4\gamma}$. □

Lemma 5: We have $\mathbb{E}[|x_A^k|]^2 = 1$ and $\mathbb{E}[x_A^k] = \mathbb{E}[x_A^{*k} x_A^{k+m}] = 0$, if the transmitted symbol x_A^k satisfies four conditions: 1) each x_A^k has unit power; 2) x_A^k is chosen from points in constellation Θ with equal probability; 3) x_A^k and x_A^{k+m} are considered independent if integer $|m| \geq 2$; 4) the topology of constellation Θ is central symmetry.

Proof: From condition 1), we have $\mathbb{E}[|x_A^k|]^2 = 1$. Expand x_A^k as $x_A^k = p + qi$. From conditions 2) and 4), for any x_A^k we could find a symmetric $x_A^{'k} = -p - qi$ with equal probability P_{x_A}. Thus, we have $\mathbb{E}[x_A^k] = 0$. Similarly, considering condition 3), we can also have $\mathbb{E}[x_A^{*k} x_A^{k+m}] = 0$. □

[6]Here we use M to replace $*$ in (19) in case of confusing index with conjugate.

Corollary 1: If $d_{\bar{a}_k r_k} = d_{\bar{b}_k r_k} = d_{\bar{a}_{k+1} r_{k+1}} = d_{\bar{b}_{k+1} r_{k+1}} = d_{\bar{s}r}$, $\forall * * \in \{A, B\}$ and $k \neq 1$, we can further simplify (46) as $e_k = e_{k+1} := e_{**}$ in (43), shown at the top of this page. Following a same procedure, b_k can be simplified as $b_k = b_{b+1} := b_{**}$.

Proof: According to Lemma 4 and referring to (1), (46) can be rewritten as (43) easily. □

Appendix IV: Solution of (23)

If L is large enough, (23) is equivalent to

$$\max_{d_R}\ \tilde{R}_A(S)=\frac{\log_2\left(\left(1+3S+S^2\right)\left(1+4S+3S^2\right)^{\frac{L-4}{2}}\right)}{L+1}, \quad \text{s.t. } d_R>0,\ S \text{ is given in (22).} \tag{48}$$

Due to the monotonicity of $\tilde{R}_A(S)$, the solution of (48) can be obtained by solving (49).

$$\max_{d_R}\ S, \quad \text{s.t. } d_R>0. \tag{49}$$

By solving $\frac{\partial S}{\partial d_R} = 0$, i.e. $Q(d_R) = 0$ in (44), shown at the top of this page, the maximal SE can be achieved with optimal d_R.

Acknowledgment

The authors thank the reviewers for their detailed reviews and constructive comments, which helped improve the quality of this paper.

References

[1] M. Fallgren, *Scenarios, Requirements and KPIs for 5G Mobile and Wireless System*, document ICT-317669-METIS/D1.1, Apr. 2013.

[2] X. Dai, Z. Zhang, B. Bai, S. Chen, and S. Sun, "Pattern division multiple access: A new multiple access technology for 5G," *IEEE Wireless Commun.*, vol. 25, no. 2, pp. 54–60, Apr. 2018.

[3] L. Li, C. Dong, L. Wang, and L. Hanzo, "Spectral-efficient bidirectional decode-and-forward relaying for full-duplex communication," *IEEE Trans. Veh. Technol.*, vol. 65, no. 9, pp. 7010–7020, Sep. 2016.

[4] Z. Zhang, Z. Ma, M. Xiao, G. K. Karagiannidis, Z. Ding, and P. Fan, "Two-timeslot two-way full-duplex relaying for 5G wireless communication networks," *IEEE Trans. Commun.*, vol. 64, no. 7, pp. 2873–2887, Jul. 2016.

[5] R. Vaze and R. W. Heath, Jr., "On the capacity and diversity-multiplexing tradeoff of the two-way relay channel," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4219–4234, Jul. 2011.
[6] D. Gunduz, A. Goldsmith, and H. V. Poor, "MIMO two-way relay channel: Diversity-multiplexing tradeoff analysis," in *Proc. 42nd Allerton Conf. Commun., Control, Comput.*, Oct. 2008, pp. 1474–1478.
[7] Z. Zhang, K. Long, A. V. Vasilakos, and L. Hanzo, "Full-duplex wireless communications: Challenges, solutions, and future research directions," *Proc. IEEE*, vol. 104, no. 7, pp. 1369–1409, Jul. 2016.
[8] S. Hong *et al.*, "Applications of self-interference cancellation in 5G and beyond," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 114–121, Feb. 2014.
[9] X. Cheng, B. Yu, X. Cheng, and L. Yang, "Two-way full-duplex amplify-and-forward relaying," in *Proc. IEEE Military Commun. Conf. (MILCOM)*, Nov. 2013, pp. 1–6.
[10] E. Everett, A. Sahai, and A. Sabharwal, "Passive self-interference suppression for full-duplex infrastructure nodes," *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 680–694, Jan. 2014.
[11] G. Liu, F. R. Yu, H. Ji, V. C. M. Leung, and X. Li, "In-band full-duplex relaying: A survey, research issues and challenges," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 2, pp. 500–524, 2nd Quart., 2015.
[12] T. Cui, F. Gao, T. Ho, and A. Nallanathan, "Distributed space–time coding for two-way wireless relay networks," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 658–671, Feb. 2009.
[13] C. Psomas, M. Mohammadi, I. Krikidis, and H. A. Suraweera, "Impact of directionality on interference mitigation in full-duplex cellular networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 487–502, Jan. 2017.
[14] A. Sabharwal, P. Schniter, D. Guo, D. W. Bliss, S. Rangarajan, and R. Wichman, "In-band full-duplex wireless: Challenges and opportunities," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1637–1652, Sep. 2014.
[15] S. H. Chae and K. Lee, "Degrees of freedom of full-duplex cellular networks: Effect of self-interference," *IEEE Trans. Commun.*, vol. 65, no. 10, pp. 4507–4518, Oct. 2017.
[16] M. Xu, M. Wen, Y. Feng, F. Ji, and W. Pan, "A novel self-interference cancellation scheme for full duplex with differential spatial modulation," in *Proc. IEEE Int. Symp. PIMRC*, Sep. 2015, pp. 482–486.
[17] H. Cui, M. Ma, L. Song, and B. Jiao, "Relay selection for two-way full duplex relay networks with amplify-and-forward protocol," *IEEE Trans. Wireless Commun.*, vol. 13, no. 7, pp. 3768–3777, Jul. 2014.
[18] Q. Li, W.-K. Ma, and D. Han, "Sum secrecy rate maximization for full-duplex two-way relay networks using alamouti-based rank-two beamforming," *IEEE J. Sel. Areas Commun.*, vol. 10, no. 8, pp. 1359–1374, Dec. 2016.
[19] J. Feng, S. Ma, G. Yang, and B. Xia, "Power scaling of full-duplex two-way massive MIMO relay systems with correlated antennas and MRC/MRT processing," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4738–4753, Jul. 2017.
[20] J. S. Lemos and F. A. Monteiro, "Full-duplex massive MIMO with physical layer network coding for the two-way relay channel," in *Proc. IEEE Sensor Array Multichannel Signal Process. Workshop (SAM)*, Jul. 2016, pp. 1–5.
[21] Y. Yang, Y. Sun, E. Sun, and Y. Zhang, "Evidence theory based self-interference suppression for two-way full-duplex MIMO relays," in *Proc. IEEE Int. Conf. Comput. Inf. Technol., Ubiquitous Comput. Commun., Dependable, Auton. Secure Comput., Pervasive Intell. Comput.*, Oct. 2015, pp. 52–58.
[22] F. Gao, T. Cui, and A. Nallanathan, "On channel estimation and optimal training design for amplify and forward relay networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 5, pp. 1907–1916, May 2008.
[23] F. Gao, R. Zhang, and Y. C. Liang, "Channel estimation for OFDM modulated two-way relay networks," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4443–4455, Nov. 2009.
[24] F. Gao, R. Zhang, and Y.-C. Liang, "Optimal channel estimation and training design for two-way relay networks," *IEEE Trans. Commun.*, vol. 57, no. 10, pp. 3024–3033, Oct. 2009.
[25] X. Li, C. Tepedelenlioğlu, and H. Şenol, "Channel estimation for residual self-interference in full-duplex amplify-and-forward two-way relays," *IEEE Trans. Wireless Commun.*, vol. 16, no. 8, pp. 4970–4983, Aug. 2017.
[26] Y. Liu, X.-G. Xia, Z. Zhang, and H. Zhang, "Distributed space-time coding based on the self-coding of RLI for full-duplex two-way relay cooperative networks," *IEEE Trans. Signal Process.*, vol. 65, no. 12, pp. 3036–3047, Jun. 2017.
[27] S. Huberman and T. Le-Ngoc, "Self-interference pricing-based MIMO full-duplex precoding," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 549–552, Dec. 2014.
[28] T. Oechtering and A. Sezgin, "A new cooperative transmission scheme using the space-time delay code," in *Proc. ITG Workshop Smart Antennas*, Mar. 2004, pp. 41–48.
[29] A. Ribeiro, X. Cai, and G. B. Giannakis, "Opportunistic multipath for bandwidth-efficient cooperative networking," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, vol. 4, Montreal, QC, Canada, May 2004, pp. 549–552.
[30] C. Ren, J. Chen, and C. Tellambura, "Spectrum sharing with device-to-device successive relaying and hybrid complex field network coding," *IEEE Trans. Veh. Technol.*, vol. 66, no. 9, pp. 7947–7963, Sep. 2017.
[31] H. Thomsen, P. Popovski, E. D. Carvalho, N. K. Pratas, D. M. Kim, and F. Boccardi, "CoMPflex: CoMP for in-band wireless full duplex," *IEEE Wireless Commun. Lett.*, vol. 5, no. 2, pp. 144–147, Apr. 2016.
[32] H. Thomsen, D. M. Kim, P. Popovski, N. K. Pratas, and E. de Carvalho, "Full duplex emulation via spatial separation of half duplex nodes in a planar cellular network," in *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Edinburgh, U.K., Jul. 2016, pp. 1–5.
[33] L. Li, H. V. Poor, and L. Hanzo, "Non-coherent successive relaying and cooperation: Principles, designs, and applications," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 3, pp. 1708–1737, 3rd Quart., 2015.
[34] C. Zhai, W. Zhang, and P. C. Ching, "Cooperative spectrum sharing based on two-path successive relaying," *IEEE Trans. Commun.*, vol. 61, no. 6, pp. 2260–2270, Jun. 2013.
[35] H. Q. Ngo, H. A. Suraweera, M. Matthaiou, and E. G. Larsson, "Multipair full-duplex relaying with massive arrays and linear processing," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1721–1737, Sep. 2014.
[36] C. C. Chu, H. C. Wang, and C. L. Wang, "Suboptimal power allocation for a two-path successive relay system with full interference cancellation," in *Proc. IEEE 75th Veh. Technol. Conf. (VTC Spring)*, May 2012, pp. 1–5.
[37] Y. Ji, C. Han, A. Wang, and H. Shi, "Partial inter-relay interference cancellation in two path successive relay network," *IEEE Commun. Lett.*, vol. 18, no. 3, pp. 451–454, Mar. 2014.
[38] H. Wicaksana, S. H. Ting, C. K. Ho, W. H. Chin, and Y. L. Guan, "AF two-path half duplex relaying with inter-relay self interference cancellation: Diversity analysis and its improvement," *IEEE Trans. Wireless Commun.*, vol. 8, no. 9, pp. 4720–4729, Sep. 2009.
[39] L. Babai, "On Lovász' lattice reduction and the nearest lattice point problem," *Combinatorica*, vol. 6, no. 1, pp. 1–13, 1986.
[40] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger, "Closest point search in lattices," *IEEE Trans. Inf. Theory*, vol. 48, no. 8, pp. 2201–2214, Aug. 2002.
[41] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.
[42] *Physical Layer Aspects*, document 3GPP 36.873 (V12.2.0), 3GPP, Jul. 2015.
[43] *Guidelines for Evaluation of Radio Interface Technologies for IMT-Advanced*, document ITU-R Rep M.2135-1, ITU-R, Dec. 2009.
[44] M. Duarte, C. Dick, and A. Sabharwal, "Experiment-driven characterization of full-duplex wireless systems," *IEEE Trans. Wireless Commun.*, vol. 11, no. 12, pp. 4296–4307, Dec. 2012.

Chao Ren (S'15–M'18) received the B.Eng. degree from the Ocean University of China in 2011 and the Ph.D. degree from Xidian University in 2017. From 2015 to 2017, he was with the University of Alberta, Canada, as a joint Ph.D. Student sponsored by the China Scholarship Council. He is currently a Lecturer with the University of Science and Technology Beijing, China. His current research interests include cognitive radio, cooperative networks, device-to-device communications, edge computing, and full-duplex relaying. He served as a Technical Program Committee Member for the IEEE VTC in 2017 and 2018, and ICC in 2019.

Haijun Zhang (M'13–SM'17) was a Post-Doctoral Research Fellow with the Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver Campus, Canada. He is currently a Full Professor with the University of Science and Technology Beijing, China. He was a recipient of the IEEE CSIM Technical Committee Best Journal Paper Award in 2018 and the IEEE ComSoc Young Author Best Paper Award in 2017. He serves as an Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS, the IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING, and the IEEE 5G TECH FOCUS. He serves/served as a Leading Guest Editor for *IEEE Communications Magazine* and the IEEE TRANSACTIONS ON EMERGING TOPICS IN COMPUTING. He serves/served as the General Co-Chair for GameNets 2016, as the Symposium Chair for Globecom 2019, as the TPC Co-Chair for INFOCOM 2018 Workshop IECCO, as the General Co-Chair for ICC 2018/ICC 2017/Globecom 2017 Workshop on UDN, and as the General Co-Chair for Globecom 2017 Workshop on LTE-U.

Jinming Wen received the bachelor's degree in information and computing science from the Jilin Institute of Chemical Technology, Jilin City, China, in 2008, the M.Sc. degree in pure mathematics from the Mathematics Institute, Jilin University, Jilin City, in 2010, and the Ph.D. degree in applied mathematics from McGill University, Montreal, Canada, in 2015. He was a Post-Doctoral Research Fellow with the Laboratoire LIP from 2015 to 2016, with the University of Alberta from 2016 to 2017, and also with the University of Toronto from 2017 to 2018. Since 2018, he has been a Full Professor with the College of Information Science and Technology and also with the College of Cyber Security, Jinan University, Guangzhou. He has published around 40 papers in top journals, including *Applied and Computational Harmonic Analysis*, the IEEE TRANSACTIONS ON INFORMATION THEORY, the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and conferences. His research interests are in the areas of lattice reduction and sparse recovery. He is an Associate Editor of the IEEE ACCESS.

Jian Chen (M'14) received the B.Eng. degree from Xi'an Jiaotong University, Xi'an, China, in 1989, the M.Eng. degree from the Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, in 1992, and the Ph.D. degree in telecommunications engineering from Xidian University, Xi'an, in 2005. From 2007 to 2008, he was a Visiting Scholar with The University of Manchester, U.K. He is currently a Full Professor with the School of Telecommunications Engineering, Xidian University. His research interests are cognitive radio, physical-layer security, wireless sensor networks, compress sensing, and signal processing.

Chintha Tellambura (F'11) received the B.Sc. degree (Hons.) from the University of Moratuwa, Sri Lanka, the M.Sc. degree in electronics from the King's College, University of London, U.K., and the Ph.D. degree in electrical engineering from the University of Victoria, Canada. He was with Monash University, Australia, from 1997 to 2002. He is currently a Professor with the Department of Electrical and Computer Engineering, University of Alberta. His current research interests include the design, modeling, and analysis of cognitive radio, heterogeneous cellular networks, fifth-generation wireless networks, and machine learning algorithms.

He has authored or co-authored over 500 journal and conference papers, with an h-index of 66 (Google Scholar). In 2017, he was an Elected Fellow of The Canadian Academy of Engineering. He received the Best Paper Award from the Communication Theory Symposium at the 2012 IEEE International Conference on Communications (ICC), Canada, and the 2017 ICC, France. He is the Winner of the prestigious McCalla Professorship and the Killam Annual Professorship from the University of Alberta. He served as an Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS from 1999 to 2011 and the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2001 to 2007. He was an Area Editor of *Wireless Communications Systems and Theory* from 2007 to 2012.