

Copyright © 2014 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Achievable Rates of Full-Duplex MIMO Radios in Fast Fading Channels with Imperfect Channel Estimation

Ali Cagatay Cirik, Yue Rong, *Senior Member, IEEE*, and Yingbo Hua, *Fellow, IEEE*

Abstract—We study the theoretical performance of two full-duplex multiple-input multiple-output (MIMO) radio systems: a full-duplex bi-directional communication system and a full-duplex relay system. We focus on the effect of a (digitally manageable) residual self-interference due to imperfect channel estimation (with independent and identically distributed (i.i.d.) Gaussian channel estimation error) and transmitter noise. We assume that the instantaneous channel state information (CSI) is not available to the transmitters. To maximize the system ergodic mutual information, which is a non-convex function of power allocation vectors at the nodes, a gradient projection algorithm is developed to optimize the power allocation vectors. This algorithm exploits both spatial and temporal freedoms of the source covariance matrices of the MIMO links between transmitters and receivers to achieve higher sum ergodic mutual information. It is observed through simulations that the full-duplex mode is optimal when the nominal self-interference is low, and the half-duplex mode is optimal when the nominal self-interference is high. In addition to an exact closed-form ergodic mutual information expression, we introduce a much simpler asymptotic closed-form ergodic mutual information expression, which in turn simplifies the computation of the power allocation vectors.

Index Terms—Full-duplex MIMO radio, bi-directional communication, full-duplex relays, fast fading channels.

I. INTRODUCTION

This paper concerns radio frequency (RF) wireless communication systems or simply called radios. A radio can be used as a wireless relay between two other radios, which we call a relay system. Two radios can be used to communicate directly with each other, which we call a bi-directional system.

Wireless relays have attracted a great deal of attention for next generations of wireless communication systems as relays can reduce the overall path loss and transmission power consumption and they also can increase cell coverage and capacity. A conventional wireless relay is half-duplex, which

transmits and receives using two different channels (in time or frequency). A full-duplex relay can transmit and receive using a single frequency at the same time and is more spectrally efficient [2], [3].

Bi-directional communication is commonly required in virtually all modern communication systems, where two terminals exchange information with each other. Currently, all bi-directional systems are half-duplex, which requires two different channels for two opposite directions. A full-duplex bi-directional system uses a single frequency at the same time for both directions and is twice as spectrally efficient [4]-[6].

Among the earliest works on full-duplex radio is [7] where a narrowband (200kHz) full-duplex radio testbed was reported. This research effort stayed almost dormant until the work [8] published ten years later. It was then followed by the hardware-based research activities in [9]-[17] as well as the theoretical research activities in [18]-[35].

A fundamental enabler for full-duplex radios is known as the self-interference cancellation. When a full-duplex radio transmits, it causes self-interference which must be canceled satisfactorily. The cancellation can be done by different methods, to different degrees, and at different stages along the receiving chain of a full-duplex radio. Cancellation of interference before the interference-corrupted signal is digitized is called analog cancellation. One important advantage of analog cancellation is that the desired (weak) signal from a remote radio will be less saturated with the receiver noise (including the receiver quantization noise)¹.

A simple testbed for analog cancellation was reported in [9] where two transmit antennas were used to create a null at a receive antenna. A demonstration of analog cancellation using an analog circuit was shown in [10]. Analog cancellation using real-time channel estimation was reported in [11]. Analog cancellation using a single antenna and multiple antennas capable of simultaneous reception and transmission was demonstrated in [12]-[15] and [16], respectively. The works shown in [9]-[13] assume that the interference channel is allpass. Broadband analog cancellation for frequency-selective interference channels was demonstrated in [14]-[17]. The amount of cancellation

Copyright (c) 2014 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

Part of this work was presented in the Asilomar Conference on Signals, Systems and Computers, Monterey, California, November 2013 [1]. This work was in part supported by Army Research Office. The work of Y. Rong was supported in part by the Australian Research Council's Discovery Projects funding scheme (project number DP140102131).

A. C. Cirik was with the Department of Electrical Engineering, University of California, Riverside, CA 92521, USA. He is now with Centre for Wireless Communications, University of Oulu, Oulu, Finland (email: acirik@ee.ucr.edu).

Y. Hua is with the Department of Electrical Engineering, University of California, Riverside, CA 92521, USA (email: yhua@ee.ucr.edu).

Y. Rong is with the Department of Electrical and Computer Engineering, Curtin University, Bentley, WA 6102, Australia (email: y.rong@curtin.edu.au).

¹The receiver noise includes quantization noise and nonlinearity noise. For a fixed number of quantization bits used by a receiver, the amount of quantization noise increases with the dynamic range of the signal (including the interference) received. By reducing the interference at the RF front end, the dynamic range of the signal received at the quantizer is also reduced. The nonlinearity noise also increases with the dynamic range of the signal received.

demonstrated on hardware varies and depends on many possible factors in the hardware systems.

The theoretical works shown in [18]-[27] all exploit multiple antennas for analog interference cancelation. The key idea among all these theoretical works for analog interference cancelation is based on a well-known concept of array processing, which is often referred to as transmit beamforming. The basic idea of this approach is that the self-interference can be cancelled at the front-end of the receiver by generating a cancellation signal based on the transmit signal in the baseband. For example, the authors in [20], [22] propose null-space projection and minimum mean-squared error (MMSE) filters for spatial self-interference suppression. In [23], an interference nulling algorithm is proposed through the optimization of the relay processing vectors over the continuous domain, which was shown to have better performance than the methods in [22]. In [24], an overview of beamforming and power allocation for both full-duplex and half-duplex MIMO relays operating in decode-and-forward or amplify-and-forward mode are provided.

In this paper, we assume that an (imperfect) analog interference cancelation or passive suppression has been implemented in the full-duplex radios and the residual self-interference can be handled digitally in the baseband. We focus on a theoretical performance of the full-duplex radios under the effect of the residual self-interference. The contributions shown in this paper are closely related to [26]-[27]. One of the differences between this paper and those two is that we consider fast fading channels and they considered slow fading channels. Fast fading channel results from such a fast varying environment where the channel coherence time is much less than a coding and channel estimation delay requirement. For each residual self-interference channel, we also apply the fast fading channel model. This is because the self-interference channel (even if through an RF circulator for a single antenna) still depends on the positions of the nearby moving reflectors. Consequently, we use an ergodic mutual information to measure the system performance. Note that unlike slow fading channels assumed in [26]-[27] where instantaneous CSI can be estimated with reasonable accuracy, here we do not assume any *instantaneous* CSI feedback from the receiver. Instead, we assume that the receiver feeds the transmitter with *statistical* CSI (the mean and variance of the CSI) and the knowledge of the statistics of the CSI is used at the transmitter to design optimal power schedules.

Since computing the closed form expression of the ergodic mutual information for fast fading channels is intractable, unlike [26]-[27], we assume that the variances of the transmission noise and the receiver noise do not depend on the variance of the transmitted signal and the received signal, respectively. Such an assumption is reasonable, since recent experimental results presented in [11] suggest that the residual self-interference of a point-to-point full-duplex system is additive, noise-like and its variance does not depend on the variance of the transmitted signal, which is also pointed out in [28]. In addition, the approximation of the effects of nonlinearities in [26]-[27] is valid only if higher order nonlinearities are contributing significantly [29], which is not the model we are

considering in this paper. This invariant transmission noise model has been commonly used in other papers [20], [22], [29], [30], [31], [32].²

By exploiting both spatial and temporal freedoms of the source covariance matrices of the MIMO links, the authors of [26] and [27] maximize the lower bound of the achievable rates for full-duplex MIMO relay channels and full-duplex bi-directional MIMO channels for slow fading channels using gradient projection (GP) method under transmitter and receiver distortions, respectively. Using the same transmit/receive distortion model in [26], [27], the authors in [33] consider the weighted sum-rate (WSR) maximization problem subject to total power constraint of the full-duplex bi-directional MIMO system. Based on the relationship between WSR and weighted minimum mean-squared-error problem, a low complexity iterative alternating algorithm is proposed. Sum-rate maximization problem subject to multiple generalized linear constraints is considered in [34], and is solved using two sub-optimal techniques. In this paper, we develop algorithms useful to reveal a lower bound on the ergodic mutual information of a full-duplex bi-directional MIMO system and a full-duplex MIMO relay system under a simpler transmitter distortion model for fast fading channels where the instantaneous CSI is not known at the transmitters and imperfectly known at the receivers. In particular, using statistical CSI at the transmitters, we optimize the power allocation vectors at the nodes to maximize the ergodic mutual information of the full-duplex systems subject to power constraints at the nodes under transmitter impairments. We develop a GP method to solve these non-convex optimization problems.

Moreover, based on [36], we introduce a simpler asymptotic closed-form expression for the ergodic mutual information of these full-duplex systems, which is shown to be an accurate approximation even for systems with a small number of antennas. This expression simplifies the computation of the non-convex power allocation problem. It is shown through numerical simulations that at a high self-interference power level (when the INR is above the transmission SNR), the optimal power schedule is the half-duplex mode and at a low self-interference power level (when the INR is below the transmission SNR), the optimal power schedule is the full-duplex mode.

This paper is organized as follows. In Section II, the system model of full-duplex bi-directional MIMO system is discussed. In Section III, we formulate the exact closed form of the lower bound ergodic mutual information expression for the full-duplex bi-directional MIMO system. In Section IV, we maximize the sum ergodic mutual information subject to per node average power constraints using the GP method, and

²Note that the baseband cancellation is only possible when the residual self-interference is small. Subject to a small self-interference, it is appropriate to model the transmission noise variance and receiver noise variance as independent of the variance of the transmitted and received signal, respectively. This is because that the impact of these noises is much smaller than the self-interfering “signal”. Note that the power of the transmission noise is typically 30 – 40dB below that of the transmitted signal. The model we use is completely reasonable for a small dynamic range commonly encountered in baseband processing, and this paper only claims the applicability in this situation.

a simple asymptotic closed-form ergodic mutual information expression is introduced as well. In Section V, the system model of full-duplex MIMO relay system is discussed. In Section VI, simulation results are provided to validate the performance of the algorithms. The main results of this paper are concluded in Section VII.

The following notations are used in this paper. Matrices and vectors are denoted by bold capital and lowercase letters, respectively. For matrices and vectors, $(\cdot)^T$ and $(\cdot)^H$ denote transpose and conjugate transpose, respectively. $\mathbb{E}_{\mathbf{H}}\{\cdot\}$ stands for the statistical expectation with respect to the channel matrix \mathbf{H} ; \mathbf{I}_N denotes an $N \times N$ identity matrix; $\text{tr}\{\cdot\}$ stands for matrix trace; $|\cdot|$ is the determinant; $\|\cdot\|$ is the Euclidean norm of a vector and the Frobenius-norm of a matrix; $(\cdot)'$ denotes the first order derivative; $\text{diag}\{a_1, \dots, a_n\}$ denotes a diagonal matrix with the diagonal elements given by a_1, \dots, a_n . $\mathcal{CN}(\mu, \sigma^2)$ denotes complex Gaussian distribution with mean μ and variance σ^2 . We will also refer to full-duplex as FD and half-duplex as HD.

II. SYSTEM MODEL FOR A FD BI-DIRECTIONAL LINK

In this section, we describe the system model of a FD bi-directional MIMO system. (A FD MIMO relay system is discussed in Section V.) We assume that each node has N physical antennas that can be used for simultaneous receiving and transmitting at the same carrier frequency [16].³ Also note that even for a single physical antenna, there is still a self-interference channel between the virtual transmit antenna and the virtual receive antenna, and the response of this (circuit) channel is still affected by the reflectors around the physical antenna. The number of virtual antennas may correspond to the number of front-ends. A two front-end relay case was studied in [35].

Similar to [26] and [27], we partition the data transmission period under consideration or control into two time slots, since the benefit when the number of time slots is larger than the number of links is not significant [37]. The partition of the data transmission follows the concept of space-time power scheduling for multiple concurrent co-channel links shown in [37]. Particularly, the use of two distinct time slots gives the freedom to switch between FD and HD signaling depending on the power of the self-interference channel, while one time slot forces FD signaling, regardless of the power of the self-interference channel. This is similar to the MIMO interference channel in [37] and FD systems in [26], [27]. Particularly, the data transmission period is partitioned into two non-equal-length slots normalized to $\tau \in [0, 1]$ and $1-\tau$, respectively, and τ can be optimized using a grid search [26]. For convenience, we define $\tau(1) \triangleq \tau$ and $\tau(2) \triangleq 1 - \tau$.

As illustrated in Fig. 1, the receiver $i \in \{1, 2\}$ receives signals from both transmitters via MIMO channels $\mathbf{H}_{ij} \in \mathbb{C}^{N \times N}$. Here, \mathbf{H}_{ii} is the channel for i th transmitter-receiver pair between the two nodes, and $\mathbf{H}_{ij}, j \in \{1, 2\}$ and $j \neq i$ denotes the self-interference channel from transmitter j to receiver i . All the channel matrices are assumed to be mutually

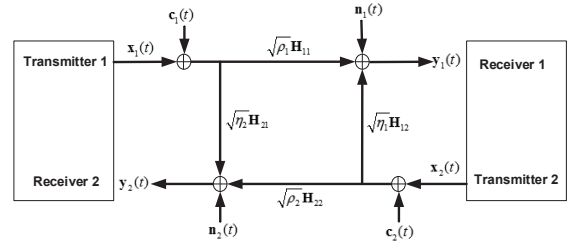


Fig. 1. The signal flow diagram of a bi-directional full-duplex MIMO system.

independent and the entries of each matrix are i.i.d. circular complex Gaussian variables with zero mean and unit variance⁴. We adopt the channel error model used for the FD systems in [4], [18], [19], [20], [22] and [38], where the receiver $i \in \{1, 2\}$ is provided with some partial information of the channel, \mathbf{H}_{ij} , $j = 1, 2$, and with this imperfect CSI, the receiver i performs MMSE estimation of \mathbf{H}_{ij} . Let us denote the MMSE estimation as $\hat{\mathbf{H}}_{ij}$, and the estimation error as $\Delta\mathbf{H}_{ij} = \mathbf{H}_{ij} - \hat{\mathbf{H}}_{ij}$, where $\hat{\mathbf{H}}_{ij}$ and $\Delta\mathbf{H}_{ij}$ are uncorrelated, and the entries of $\Delta\mathbf{H}_{ij}$ are zero mean circularly symmetric complex Gaussian with variance $\sigma_{e,ij}^2$, as opposed to non-i.i.d. channel estimation errors in [26], [27]. Note that $\sigma_{e,ij}^2$ is assumed to be known to both the transmitter and receiver [39]. We will assume that the channel matrices remain constant over two consecutive time slots, but change randomly over an interval of many multiples of two time slots. We will design the power schedule to maximize an ergodic system mutual information which is averaged over the statistical distribution of the channel matrices. This mutual information is achievable (approximately) over the interval of many multiples of two time slots. Therefore, our theory is valid for “fast fading” channels, i.e., the time delay due to encoding and decoding over many multiples of two time slots is tolerable.

The quality of transmitted signals suffer from non-linear distortions in the power amplifier, phase noise, and IQ-imbalance [40]. The measurement results by [41] indicate that an i.i.d. additive Gaussian noise model accurately describes the sum of all such residual transmitter impairments. Such an assumption has also been commonly used in other FD papers [20], [22], [29], [30], [31], [32].

We consider a FD bi-directional MIMO system that suffers from self-interference. The $N \times 1$ received signal vector at the i th receiver can be written as

$$\begin{aligned} \mathbf{y}_i(t) &= \sqrt{\rho_i}\mathbf{H}_{ii}(\mathbf{x}_i(t) + \mathbf{c}_i(t)) + \sqrt{\eta_i}\mathbf{H}_{ij}(\mathbf{x}_j(t) + \mathbf{c}_j(t)) \\ &\quad + \mathbf{n}_i(t) \\ &= \sqrt{\rho_i}\hat{\mathbf{H}}_{ii}\mathbf{x}_i(t) + \sqrt{\rho_i}\Delta\mathbf{H}_{ii}\mathbf{x}_i(t) + \sqrt{\rho_i}\mathbf{H}_{ii}\mathbf{c}_i(t) \\ &\quad + \sqrt{\eta_i}\hat{\mathbf{H}}_{ij}\mathbf{x}_j(t) + \sqrt{\eta_i}\Delta\mathbf{H}_{ij}\mathbf{x}_j(t) + \sqrt{\eta_i}\mathbf{H}_{ij}\mathbf{c}_j(t) \\ &\quad + \mathbf{n}_i(t), \quad i, j \in \{1, 2\} \text{ and } j \neq i, \end{aligned} \quad (1)$$

⁴Before analog domain cancellation, the self-interference channel has a strong line-of-sight (LOS) component, so can be modeled as Ricean distribution with a large K -factor. It is shown experimentally in [11] that after applying an analog domain cancellation or antenna isolation, the strong LOS component is attenuated, resulting in a Ricean distribution with a small K -factor or in a Rayleigh distribution (if the suppression through analog cancellation is high).

³A full duplex WiFi radio that uses multiple antennas that transmit and receive simultaneously was designed and implemented in [16].

where ρ_i denotes the average power gain of the i th transmitter-receiver link, η_i denotes the average power gain of the self-interference channel, $\mathbf{x}_i(t) \sim \mathcal{CN}(\mathbf{0}, \mathbf{Q}_i(t))$ is the signal vector transmitted by node i within time slot t , $\mathbf{x}_j(t) \sim \mathcal{CN}(\mathbf{0}, \mathbf{Q}_j(t))$ is the self-interference vector from the transmitter j , $j \neq i$ within time slot t , and $\mathbf{n}_i(t) \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$ is the receiver noise which is additive white Gaussian noise (AWGN) vector. We assume that $\mathbf{n}_i(t)$ is independent of $\mathbf{x}_i(t)$ and $\mathbf{x}_j(t)$.

In (1), $\sqrt{\rho_i}\mathbf{c}_i(t)$ denotes the transmission noise from the i th transmitter, where $\mathbf{c}_i(t) \sim \mathcal{CN}(\mathbf{0}, \sigma_i^2\mathbf{I}_N)$, $i = 1, 2$. Note that the transmit noise in (1) is $\sqrt{\rho_i}\mathbf{c}_i(t)$, not $\mathbf{c}_i(t)$ alone. And since the signal power is $\rho_i P_i$, while the transmission noise power is $\rho_i\sigma_i^2$, the transmitter noise depends on the power level. Here P_i is the averaged transmit power from the i th transmitter. In particular, incorporating $\sqrt{\rho_i}$ into $\mathbf{c}_i(t)$, we have the same transmission noise model as [20], [22], [29], [30], [31], [32], [41].

The receiver $i \in \{1, 2\}$ knows the interfering signal $\mathbf{x}_j(t)$ from transmitter $j \in \{1, 2\}$, $j \neq i$, so the self-interference term $\sqrt{\eta_i}\tilde{\mathbf{H}}_{ij}\mathbf{x}_j(t)$ can be subtracted from $\mathbf{y}_i(t)$ [26], [27].

$$\begin{aligned}\tilde{\mathbf{y}}_i(t) &\triangleq \mathbf{y}_i(t) - \sqrt{\eta_i}\tilde{\mathbf{H}}_{ij}\mathbf{x}_j(t) \\ &= \sqrt{\rho_i}\tilde{\mathbf{H}}_{ii}\mathbf{x}_i(t) + \mathbf{v}_i(t),\end{aligned}\quad (2)$$

where

$$\begin{aligned}\mathbf{v}_i(t) &= \sqrt{\rho_i}\Delta\mathbf{H}_{ii}\mathbf{x}_i(t) + \sqrt{\rho_i}\mathbf{H}_{ii}\mathbf{c}_i(t) + \sqrt{\eta_i}\Delta\mathbf{H}_{ij}\mathbf{x}_j(t) \\ &\quad + \sqrt{\eta_i}\mathbf{H}_{ij}\mathbf{c}_j(t) + \mathbf{n}_i(t)\end{aligned}\quad (3)$$

is the total noise in $\tilde{\mathbf{y}}_i(t)$. The covariance matrix of $\mathbf{v}_i(t)$ can be written as

$$\begin{aligned}\tilde{\Sigma}_i(t) &= \mathbb{E}\left\{\mathbf{v}_i(t)\mathbf{v}_i(t)^H|\tilde{\mathbf{H}}_{ii}, \tilde{\mathbf{H}}_{ij}\right\} \\ &= \rho_i\mathbb{E}_{\Delta\mathbf{H}_{ii}}\left\{\Delta\mathbf{H}_{ii}\mathbf{Q}_i(t)\Delta\mathbf{H}_{ii}^H\right\} + \rho_i\sigma_i^2\mathbb{E}_{\Delta\mathbf{H}_{ii}}\left\{\mathbf{H}_{ii}\mathbf{H}_{ii}^H\right\} \\ &\quad + \eta_i\mathbb{E}_{\Delta\mathbf{H}_{ij}}\left\{\Delta\mathbf{H}_{ij}\mathbf{Q}_j(t)\Delta\mathbf{H}_{ij}^H\right\} + \eta_i\sigma_j^2\mathbb{E}_{\Delta\mathbf{H}_{ij}}\left\{\mathbf{H}_{ij}\mathbf{H}_{ij}^H\right\} + \mathbf{I}_N \\ &= \rho_i\sigma_{e,ii}^2\text{tr}\{\mathbf{Q}_i(t)\}\mathbf{I}_N + \rho_i\sigma_i^2\left(\tilde{\mathbf{H}}_{ii}\tilde{\mathbf{H}}_{ii}^H + \sigma_{e,ii}^2N\mathbf{I}_N\right) \\ &\quad + \eta_i\sigma_{e,ij}^2\text{tr}\{\mathbf{Q}_j(t)\}\mathbf{I}_N + \eta_i\sigma_j^2\left(\tilde{\mathbf{H}}_{ij}\tilde{\mathbf{H}}_{ij}^H + \sigma_{e,ij}^2N\mathbf{I}_N\right) \\ &\quad + \mathbf{I}_N, \quad i, j \in \{1, 2\} \text{ and } j \neq i,\end{aligned}\quad (4)$$

where the first expectation is taken with respect to $\mathbf{x}_i(t)$, $\mathbf{x}_j(t)$ and $\mathbf{n}_i(t)$, and here we have used the identity of $\mathbb{E}_{\Delta\mathbf{H}_{ij}}\left\{\Delta\mathbf{H}_{ij}\mathbf{A}\Delta\mathbf{H}_{ij}^H\right\} = \sigma_{e,ij}^2\text{tr}\{\mathbf{A}\}\mathbf{I}_N$, where the entries of $\Delta\mathbf{H}_{ij}$ are i.i.d. with $\mathcal{CN}(0, \sigma_{e,ij}^2)$ and $\mathbf{A} \in \mathbb{C}^{N \times N}$ is a known matrix.

III. ACHIEVABLE RATES

In this section, we formulate the ergodic mutual information expression for the FD bi-directional MIMO system when the transmitters do not have instantaneous CSI and the receivers have imperfect instantaneous CSI, i.e., \mathbf{H}_{ii} is unknown at the transmitter i but partially known at the receiver i . As a result of the channel estimation errors and transmitter impairments in (3), the noise $\mathbf{v}_i(t)$ is generally non-Gaussian. To the best of our knowledge, the exact mutual information of MIMO channels with channel estimation errors is still an

open problem even for point-to-point MIMO systems [39], [42]. However, assuming $\mathbf{v}_i(t)$ as Gaussian, which is the worst noise distribution from the perspective of mutual information, we can obtain the lower bound [42], which was also used in [26], [27].

For a given time-sharing parameter τ , the lower bound of the sum mutual information of the system averaged over two time slots can be written as

$$I(\mathbf{Q}_1, \mathbf{Q}_2) = \sum_{i=1}^2 \sum_{t=1}^2 \tau(t) \log_2 \left| \mathbf{I}_N + \rho_i \tilde{\mathbf{H}}_{ii} \mathbf{Q}_i(t) \tilde{\mathbf{H}}_{ii}^H \tilde{\Sigma}_i(t)^{-1} \right|$$

where $\mathbf{Q}_i \triangleq [\mathbf{Q}_i^T(1), \mathbf{Q}_i^T(2)]^T$, $i = 1, 2$. Then, a lower bound of the ergodic sum mutual information of the system averaged over two time slots can be written as

$$\begin{aligned}\bar{I}(\mathbf{Q}_1, \mathbf{Q}_2) &= \sum_{i=1}^2 \sum_{t=1}^2 \tau(t) \mathbb{E}_{\tilde{\mathbf{H}}_{ii}, \tilde{\mathbf{H}}_{ij}} \left\{ \log_2 \left| \mathbf{I}_N + \rho_i \tilde{\mathbf{H}}_{ii} \mathbf{Q}_i(t) \tilde{\mathbf{H}}_{ii}^H \tilde{\Sigma}_i(t)^{-1} \right| \right\} \\ &\quad \underbrace{\hspace{10em}}_{\bar{I}_i(\mathbf{Q}_1, \mathbf{Q}_2)}\end{aligned}\quad (5)$$

To derive a closed-form expression for the ergodic sum mutual information (5), we use the eigendecomposition of $\mathbf{Q}_i(t)$, which can be written as $\mathbf{Q}_i(t) = \mathbf{U}_i(t)\mathbf{D}_i(t)\mathbf{U}_i(t)^H$, $i = 1, 2$, where $\mathbf{U}_i(t)$ is the unitary matrix of eigenvectors, and

$$\mathbf{D}_i(t) = \text{diag}\{d_{i1}(t), d_{i2}(t), \dots, d_{iN}(t)\}, \quad i = 1, 2$$

is a diagonal matrix of all eigenvalues. For convenience, we will use the column vectors $\mathbf{d}_1(t)$ and $\mathbf{d}_2(t)$ defined as

$$\mathbf{d}_i(t) = [d_{i1}(t), d_{i2}(t), \dots, d_{iN}(t)]^T, \quad i = 1, 2.$$

Now we can rewrite (5) as

$$\begin{aligned}\bar{I}(\mathbf{D}_1, \mathbf{D}_2) &= \sum_{i=1}^2 \sum_{t=1}^2 \tau(t) \mathbb{E}_{\hat{\mathbf{H}}_{ii}, \hat{\mathbf{H}}_{ij}} \left\{ \log_2 \left| \mathbf{I}_N + \rho_i \hat{\mathbf{H}}_{ii} \mathbf{D}_i(t) \hat{\mathbf{H}}_{ii}^H \hat{\Sigma}_i(t)^{-1} \right| \right\}\end{aligned}\quad (6)$$

where

$$\begin{aligned}\mathbf{D}_i &\triangleq [\mathbf{D}_i^T(1), \mathbf{D}_i^T(2)]^T, \quad i = 1, 2 \\ \hat{\mathbf{H}}_{ii} &\triangleq \tilde{\mathbf{H}}_{ii} \mathbf{U}_i(t), \quad i = 1, 2 \\ \hat{\mathbf{H}}_{ij} &\triangleq \tilde{\mathbf{H}}_{ij} \mathbf{U}_j(t), \quad (i, j) \in \{1, 2\} \text{ and } j \neq i \\ \hat{\Sigma}_i(t) &\triangleq \rho_i \sigma_{e,ii}^2 \text{tr}\{\mathbf{D}_i(t)\}\mathbf{I}_N + \rho_i \sigma_i^2 \left(\hat{\mathbf{H}}_{ii} \hat{\mathbf{H}}_{ii}^H + \sigma_{e,ii}^2 N \mathbf{I}_N \right) \\ &\quad + \eta_i \sigma_{e,ij}^2 \text{tr}\{\mathbf{D}_j(t)\}\mathbf{I}_N + \eta_i \sigma_j^2 \left(\hat{\mathbf{H}}_{ij} \hat{\mathbf{H}}_{ij}^H + \sigma_{e,ij}^2 N \mathbf{I}_N \right) \\ &\quad + \mathbf{I}_N, \quad i, j \in \{1, 2\} \text{ and } j \neq i.\end{aligned}$$

Since $\tilde{\mathbf{H}}_{ii}$ has i.i.d. Gaussian entries and $\mathbf{U}_i(t)$ is unitary, the statistics of $\hat{\mathbf{H}}_{ii}$ is identical to that of $\tilde{\mathbf{H}}_{ii}$ [43, Lemma 5]. Therefore, for notational simplicity, in the sequel we will drop the hats on the matrices. Thus, the ergodic sum mutual

information expression (6) can be rewritten as

$$\begin{aligned}
& \bar{I}(\mathbf{d}_1, \mathbf{d}_2) \\
&= \sum_{i=1}^2 \sum_{t=1}^2 \tau(t) \mathbb{E}_{\tilde{\mathbf{H}}_{ii}, \tilde{\mathbf{H}}_{ij}} \left\{ \log_2 \left| \mathbf{I}_N + \rho_i \tilde{\mathbf{H}}_{ii} \mathbf{D}_i(t) \tilde{\mathbf{H}}_{ii}^H \tilde{\Sigma}_i(t)^{-1} \right| \right\} \\
&= \sum_{i=1}^2 \sum_{t=1}^2 \tau(t) \left[\mathbb{E}_{\tilde{\mathbf{H}}_{ii}} \left\{ \log_2 \left| \rho_i \tilde{\mathbf{H}}_{ii} \mathbf{D}_i(t) \tilde{\mathbf{H}}_{ii}^H + \tilde{\Sigma}_i(t) \right| \right\} \right. \\
&\quad \left. - \mathbb{E}_{\tilde{\mathbf{H}}_{ii}} \left\{ \log_2 \left| \tilde{\Sigma}_i(t) \right| \right\} \right] \\
&= \sum_{i=1}^2 \sum_{t=1}^2 \tau(t) \left[\mathbb{E}_{\tilde{\mathbf{H}}_{ii}} \left\{ \log_2 \left| \rho_i \tilde{\mathbf{H}}_{ii} (\mathbf{D}_i(t) + \sigma_t^2 \mathbf{I}_N) \tilde{\mathbf{H}}_{ii}^H \right. \right. \right. \\
&\quad \left. \left. + \eta_i \sigma_t^2 \tilde{\mathbf{H}}_{ij} \tilde{\mathbf{H}}_{ij}^H + c_i(t) \mathbf{I}_N \right| \right\} \\
&\quad \left. - \mathbb{E}_{\tilde{\mathbf{H}}_{ii}} \left\{ \log_2 \left| \rho_i \sigma_t^2 \tilde{\mathbf{H}}_{ii} \tilde{\mathbf{H}}_{ii}^H + \eta_i \sigma_t^2 \tilde{\mathbf{H}}_{ij} \tilde{\mathbf{H}}_{ij}^H + c_i(t) \mathbf{I}_N \right| \right\} \right] \\
&= \sum_{i=1}^2 \sum_{t=1}^2 \tau(t) \left[\mathbb{E}_{\tilde{\mathbf{H}}_{ii}} \left\{ \log_2 \left| \tilde{\mathbf{H}}_{ii} \Lambda_i(t) \tilde{\mathbf{H}}_{ii}^H + \mathbf{I}_N \right| \right\} \right. \\
&\quad \left. - \mathbb{E}_{\tilde{\mathbf{H}}_{ii}} \left\{ \log_2 \left| \tilde{\mathbf{H}}_{ii} \bar{\Lambda}_i(t) \tilde{\mathbf{H}}_{ii}^H + \mathbf{I}_N \right| \right\} \right], \quad (7)
\end{aligned}$$

where

$$\begin{aligned}
& \tilde{\mathbf{H}}_i = \begin{bmatrix} \tilde{\mathbf{H}}_{ii} & \tilde{\mathbf{H}}_{ij} \end{bmatrix} \\
& c_i(t) \triangleq \rho_i \sigma_{e,ii}^2 \mathbf{1}_N^T \mathbf{d}_i(t) + \rho_i \sigma_t^2 \sigma_{e,ii}^2 N + \eta_i \sigma_{e,ij}^2 \mathbf{1}_N^T \mathbf{d}_j(t) \\
&\quad + \eta_i \sigma_t^2 \sigma_{e,ij}^2 N + 1, \quad i, j \in \{1, 2\} \text{ and } j \neq i \\
& \mathbf{d}_i \triangleq [\mathbf{d}_i(1)^T, \mathbf{d}_i(2)^T]^T, \quad i = 1, 2 \\
& \tilde{\Sigma}_i(t) \triangleq \rho_i \sigma_{e,ii}^2 \text{tr} \{ \mathbf{D}_i(t) \} \mathbf{I}_N + \rho_i \sigma_t^2 \left(\tilde{\mathbf{H}}_{ii} \tilde{\mathbf{H}}_{ii}^H + \sigma_{e,ii}^2 N \mathbf{I}_N \right) \\
&\quad + \eta_i \sigma_{e,ij}^2 \text{tr} \{ \mathbf{D}_j(t) \} \mathbf{I}_N + \eta_i \sigma_t^2 \left(\tilde{\mathbf{H}}_{ij} \tilde{\mathbf{H}}_{ij}^H + \sigma_{e,ij}^2 N \mathbf{I}_N \right) \\
&\quad + \mathbf{I}_N, \quad i, j \in \{1, 2\} \text{ and } j \neq i \\
& \Lambda_i(t) \triangleq \text{diag} \left\{ \boldsymbol{\lambda}_{1,i}^T(t), \boldsymbol{\lambda}_{2,i}^T(t) \right\} \quad i = 1, 2 \\
& \bar{\Lambda}_i(t) \triangleq \text{diag} \left\{ \bar{\boldsymbol{\lambda}}_{1,i}^T(t), \bar{\boldsymbol{\lambda}}_{2,i}^T(t) \right\} \quad i = 1, 2 \\
& \boldsymbol{\lambda}_{1,i}(t) = \rho_i \frac{\mathbf{d}_i(t) + \sigma_t^2 \mathbf{1}_N}{c_i(t)} \\
& \boldsymbol{\lambda}_{2,i}(t) = \eta_i \frac{\sigma_t^2}{c_i(t)} \mathbf{1}_N \\
& \bar{\boldsymbol{\lambda}}_{1,i}(t) = \rho_i \frac{\sigma_t^2}{c_i(t)} \mathbf{1}_N.
\end{aligned}$$

Here $\mathbf{1}_N$ is an $N \times 1$ column vector of ones. Note that $\mathbf{1}_N^T \mathbf{d}_i(t)$, $(i, t) = 1, 2$ is the power consumed at the i th node at time slot t and it is not fixed and changes with respect to self-interference power as we will see in the simulations, whereas $\sum_{t=1}^2 \mathbf{1}_N^T \mathbf{d}_i(t)$ is the total power consumed by the node i and it is fixed.

The expression $\mathbb{E}_{\tilde{\mathbf{H}}_{ii}} \left\{ \log_2 \left| \tilde{\mathbf{H}}_{ii} \Lambda_i(t) \tilde{\mathbf{H}}_{ii}^H + \mathbf{I}_N \right| \right\}$ in (7) can be viewed as the ergodic mutual information of a point-to-point MIMO channel with $2N$ transmit and N receive antennas. A closed-form expression for the ergodic mutual information of such a system has been shown in [44], where a determinant representation for the distribution of quadratic forms of a complex Gaussian matrix has been used. Using the

results in [44], (7) can be equivalently expressed as ⁵

$$\begin{aligned}
& \bar{I}(\mathbf{d}_1, \mathbf{d}_2) \\
&= \log_2(e) \sum_{i=1}^2 \sum_{t=1}^2 \tau(t) \left[\sum_{n=0}^{N-1} \sum_{k=1}^{2N} (c_{tikn}(\Lambda_i(t)) Q(n, \lambda_{tik}) \right. \\
&\quad \left. - c_{tikn}(\bar{\Lambda}_i(t)) Q(n, \bar{\lambda}_{tik}) \right], \quad (8)
\end{aligned}$$

where $c_{tikn}(\Lambda_i(t))$ and $Q(n, \lambda_{tik})$ are defined in Appendix. Here $\lambda_{tik} \triangleq [\Lambda_i(t)]_{k,k}$ and $\bar{\lambda}_{tik} \triangleq [\bar{\Lambda}_i(t)]_{k,k}$, $k = 1, \dots, 2N$ denote the (k, k) th element of matrix $\Lambda_i(t)$ and $\bar{\Lambda}_i(t)$, respectively. In (32) of Appendix, $S_1(x) \triangleq \int_x^\infty e^{-\tau} / \tau d\tau$ is the exponential integral function of order 1 [45].

As shown in (8), the ergodic sum mutual information is now expressed as a finite summation involving rational functions and exponential integration functions of the power scheduling vectors $\mathbf{d}_i(t)$, $(i, t) \in \{1, 2\}$, of both transmitting nodes. The exponential integration function is available in many software such as MATLAB and Mathematica. Thus, (8) is easy to compute. Note that (8) is derived under the assumption that all λ_{tik} , $k = 1, \dots, 2N$, have distinct values. Under the condition that some of them are identical, the closed-form ergodic sum mutual information expression can be obtained by deriving the limit of (8) with respect to those common values of λ_{tik} using L'Hospital's rule. However, for numerical evaluation, it is sufficient to slightly and randomly perturb these identical values of λ_{tik} , since all functions are continuous and λ_{tik} is deterministic [46]. The same assumption holds for $\bar{\lambda}_{tik}$ as well.

IV. MAXIMIZATION OF THE SUM ERGODIC MUTUAL INFORMATION

In this section, we aim at maximizing the sum ergodic mutual information (5) by choosing the transmit covariance matrices $\mathbf{Q}_1(t)$ and $\mathbf{Q}_2(t)$, $t = 1, 2$ subject to per node average power constraints and subsequently optimize the time-sharing parameter τ . Note that we consider fast fading channels in which the instantaneous CSI is assumed to be unknown at the transmitting nodes. When the knowledge of the instantaneous CSI is absent, statistical properties of the CSI is necessary for designing optimal power schedules. The optimization problem can be formulated as

$$\begin{aligned}
& \max_{\mathbf{Q}_1, \mathbf{Q}_2, \tau(t)} \sum_{i=1}^2 \bar{I}_i(\mathbf{Q}_1, \mathbf{Q}_2) \quad (9) \\
& \text{s.t.} \quad \sum_{t=1}^2 \tau(t) \text{tr} \{ \mathbf{Q}_i(t) \} \leq P_i, \quad i = 1, 2 \\
& \quad \mathbf{Q}_i(t) \geq 0, \quad \forall i, t \in \{1, 2\},
\end{aligned}$$

where $\bar{I}_i(\mathbf{Q}_1, \mathbf{Q}_2)$ is given in (5) and P_i is the averaged transmit power from the i th transmitter⁶.

⁵Note that in this paper we assume the same set of antennas are used for both transmission and reception. If we assume different sets of antennas are used for transmission and reception [27], i.e., each of the two nodes uses N_t antennas for transmission and N_r different antennas for reception, then the upper limits of the first and second summations inside the bracket of (8) is replaced with $N_r - 1$ and $2N_t$, respectively.

⁶Note that in (9), $\tau(1) = \tau$ and $\tau(2) = 1 - \tau$, where $\tau \in [0, 1]$. And it can be optimized using a grid search. Similar discussion has been held in [26].

A. Gradient Projection Approach

For a fixed τ , the optimal \mathbf{d}_1 and \mathbf{d}_2 can be obtained by solving the following problem

$$\max_{\mathbf{d}_1, \mathbf{d}_2} \bar{I}(\mathbf{d}_1, \mathbf{d}_2) \quad (10)$$

$$\text{s.t.} \quad \sum_{t=1}^2 \tau(t) \|\mathbf{d}_i(t)\|_1 = P_i, \quad i = 1, 2 \quad (11)$$

$$\mathbf{d}_i \geq 0, \quad i = 1, 2, \quad (12)$$

where $\bar{I}(\mathbf{d}_1, \mathbf{d}_2)$ is given in (8) and (11) is the power constraint at the i th transmitter. Here $\|\cdot\|_1$ denotes the sum norm (or l_1 norm) of a vector. For a vector \mathbf{x} , $\mathbf{x} \geq 0$ means that each entry of \mathbf{x} is nonnegative.

The objective function (10) is highly non-convex and does not have a clear structure. We can develop numerical algorithms based on nonlinear programming techniques to obtain a locally optimal solution to the problem (10)-(12). We choose the GP method [47], which is an extension of the unconstrained steepest descent method to the convex constrained problems. The GP method is simple, efficient, and guarantees the convergence to a stationary point, provided that proper step sizes are chosen.

There are two important steps in the GP algorithm: the computation of the gradient of the objective function, and the projection of the updated optimization variable onto the convex set specified by constraint functions. To apply the GP method to solve the problem (10)-(12), we first take gradient steps for \mathbf{d}_1 and \mathbf{d}_2 , and then project the updated \mathbf{d}_1 and \mathbf{d}_2 onto the constraint set specified by (11) and (12). The gradient of the objective function (10) with respect to $d_{lm}(t)$, $l = 1, 2$, $m = 1, \dots, N$, $t = 1, 2$, is given by

$$\begin{aligned} & \frac{\partial \bar{I}(\mathbf{d}_1, \mathbf{d}_2)}{\partial d_{lm}(t)} \\ &= \tau(t) \log_2(e) \sum_{i=1}^2 \left[\sum_{n=0}^{N-1} \sum_{k=1}^{2N} (c'_{tikn}(\mathbf{\Lambda}_i(t)) Q(n, \lambda_{tik}) \right. \\ & \quad + c_{tikn}(\mathbf{\Lambda}_i(t)) Q'(n, \lambda_{tik}) - c'_{tikn}(\mathbf{\bar{\Lambda}}_i(t)) Q(n, \bar{\lambda}_{tik}) \\ & \quad \left. - c_{tikn}(\mathbf{\bar{\Lambda}}_i(t)) Q'(n, \bar{\lambda}_{tik}) \right]. \end{aligned} \quad (13)$$

The parameters in (13) are given in Appendix.

Let us first consider the gradient steps of the i th transmitter-receiver pair, $i \in \{1, 2\}$, and denote the $2N \times 1$ vector of gradient as

$$\mathbf{g}_i \triangleq \left[\frac{\partial \bar{I}(\mathbf{d}_1, \mathbf{d}_2)}{\partial d_{i1}(1)}, \dots, \frac{\partial \bar{I}(\mathbf{d}_1, \mathbf{d}_2)}{\partial d_{iN}(2)} \right]^T, \quad i = 1, 2. \quad (14)$$

Then taking a step along the positive gradient direction, the power allocation vector is updated as

$$\hat{\mathbf{d}}_i = \bar{\mathbf{d}}_i + s \mathbf{g}_i, \quad i = 1, 2,$$

where s is a scalar of step size, and $\bar{\mathbf{d}}_i$ is the previous power allocation vector.

The next step of the GP algorithm is to project $\hat{\mathbf{d}}_i$ onto the feasible region of power vector constraints (11)-(12). The projection operation is basically searching for a point $\bar{\mathbf{d}}_i$ in the region of (11)-(12), which has a minimum Euclidean

distance to the point $\hat{\mathbf{d}}_i$. Thus, the optimization problem for the projection operation can be written as

$$\min_{\bar{\mathbf{d}}_i} \left\| \bar{\mathbf{d}}_i - \hat{\mathbf{d}}_i \right\|^2 \quad (15)$$

$$\text{s.t.} \quad \sum_{t=1}^2 \tau(t) \|\bar{\mathbf{d}}_i(t)\|_1 = P_i, \quad \bar{\mathbf{d}}_i \geq 0, \quad i = 1, 2. \quad (16)$$

The problem (15)-(16) is convex and can be efficiently solved by the Lagrange multiplier method. It turns out that the problem (15)-(16) has a water-filling solution which is given by

$$\tilde{d}_{ik}(t) = \left[\hat{d}_{ik}(t) - \frac{\tau(t)\mu}{2} \right]^+, \quad k = 1, \dots, N, \quad (i, t) = 1, 2, \quad (17)$$

where $\mu \geq 0$ is the Lagrange multiplier, and for a real scalar x , $[x]^+ \triangleq \max\{x, 0\}$. The Lagrange multiplier μ can be obtained by substituting (17) back into (16) and solving the following nonlinear equation

$$\sum_{t=1}^2 \sum_{k=1}^N \tau(t) \left[\hat{d}_{ik}(t) - \frac{\tau(t)\mu}{2} \right]^+ = P_i, \quad i = 1, 2. \quad (18)$$

We can use the bisection method to solve (18), since the left hand side of (18) is a piecewise linear function and monotonically decreasing with respect to μ .

At the k th iteration, the power allocations vectors are updated as

$$\bar{\mathbf{d}}_i^{(k+1)} = \bar{\mathbf{d}}_i^{(k)} + \delta^{(k)} \left(\tilde{\mathbf{d}}_i^{(k)} - \bar{\mathbf{d}}_i^{(k)} \right), \quad i = 1, 2 \quad (19)$$

$$\tilde{\mathbf{d}}_i^{(k)} = \text{proj} \left[\bar{\mathbf{d}}_i^{(k)} + s^{(k)} \mathbf{g}_i^{(k)} \right], \quad i = 1, 2, \quad (20)$$

where $\text{proj}[\cdot]$ stands for the projection operation in (15)-(16), $\delta^{(k)}$ and $s^{(k)}$ are scalars of step size and can be chosen according to the Armijo rule [47]. In this rule, $s^{(k)} = s$ is a constant throughout the iterations, and $\delta^{(k)} = \theta^{m_k}$, where m_k is the minimal nonnegative integer that satisfies the following inequality

$$\begin{aligned} & \bar{I} \left(\bar{\mathbf{d}}^{(k+1)} \right) - \bar{I} \left(\bar{\mathbf{d}}^{(k)} \right) \\ & \geq \sigma \theta^{m_k} \sum_{i=1}^2 \left(\mathbf{g}_i^{(k)} \right)^T \left(\tilde{\mathbf{d}}_i^{(k)} - \bar{\mathbf{d}}_i^{(k)} \right), \end{aligned} \quad (21)$$

where σ and θ are constants and $\bar{\mathbf{d}}^{(k)} = \left[(\bar{\mathbf{d}}_1^{(k)})^T, (\bar{\mathbf{d}}_2^{(k)})^T \right]^T$. According to [47], usually σ is chosen close to 0, and a proper choice of θ is from 0.1 to 0.5.

The steps of (19) and (20) are performed for both nodes and continue until vector $\bar{\mathbf{d}}^{(k)}$ converges. The GP algorithm using the Armijo rule along the feasible direction guarantees such a convergence [47] and the convergence criterion is given as

$$\max \text{abs} \left\{ \bar{\mathbf{d}}^{(k+1)} - \bar{\mathbf{d}}^{(k)} \right\} \leq \epsilon, \quad (22)$$

where $\max \text{abs}\{\cdot\}$ denotes the maximal absolute value among all elements of a vector and ϵ is a positive constant close to 0. The procedure of applying the GP technique to solve the problem (10)-(12) is summarized in Table I. Subsequently, we optimize over $\tau \in [0, 1]$ using a grid-search [26].

TABLE I
PROCEDURE OF THE PROJECTED GRADIENT POWER
ALLOCATION APPROACH

- 1) Initialize power allocation vectors $\bar{\mathbf{d}}_i$.
Choose step sizes. Set $k = 0$.
- 2) Set $k := k + 1$.
Calculate the gradient of (8) $\mathbf{g}_i^{(k)}$ from (13) using (33)-(37).
Let $\hat{\mathbf{d}}_i^{(k)} = \bar{\mathbf{d}}_i^{(k)} + s\mathbf{g}_i^{(k)}$.
Project $\hat{\mathbf{d}}_i^{(k)}$ to obtain $\tilde{\mathbf{d}}_i^{(k)}$ using (17).
Update $\bar{\mathbf{d}}_i^{(k)}$ using (19) and (20).
- 3) If convergent, end.
Else go to step 2.

For the bi-directional case, the ergodic mutual information (8) and (23) are functions of averaged signal-to-noise ratio (SNR) and nominal interference-to-noise ratio (INR). Under the same INR for all interfering links, the desired link with the higher SNR gets the whole data transmission slot, i.e. $\tau = 1$ and the link with the lower SNR does not transmit, i.e. $\tau = 0$. In other words, the optimal τ is either one or zero depending on the average SNR. Though we presented a general transmission protocol and solved the optimization problem as a function of τ , this time-slot allocation is not fair for the bi-directional case, so we assumed $\tau = 0.5$ in our simulations for the bi-directional system.

B. Approximation of Sum Ergodic Mutual Information

In this subsection, we introduce a much simpler expression of $\bar{I}(\mathbf{d}_1, \mathbf{d}_2)$ than the one in (8), which in turn simplifies the computation in solving the problem (10)-(12). This simplification is based on an asymptotical form of $\bar{I}(\mathbf{d}_1, \mathbf{d}_2)$ when $N \rightarrow \infty$ as proposed in [36]. The proof of this asymptotical form is as follows: In [48], SNR at the output of an MMSE receiver is shown. And using the results in [48], the authors in [49] obtain the asymptotic capacity of an optimum receiver for randomly spread CDMA in fading channels. With a simple SNR normalization and by applying [49, Theorem IV.1], the asymptotic capacity of MIMO architectures impaired by AWGN as well as spatially colored interference can be easily found as the number of antennas go to infinity as shown in Appendix of [36]. Applying the result in [36], the sum ergodic mutual information in (7) can be approximated as

$$\begin{aligned}
& \bar{I}(\mathbf{d}_1, \mathbf{d}_2) \\
&= \sum_{i=1}^2 \sum_{t=1}^2 \tau(t) \mathbb{E}_{\tilde{\mathbf{H}}_{ii}, \tilde{\mathbf{H}}_{ij}} \left\{ \log_2 \left| \mathbf{I}_N + \rho_i \tilde{\mathbf{H}}_{ii} \mathbf{D}_i(t) \tilde{\mathbf{H}}_{ii}^H \tilde{\Sigma}_i(t)^{-1} \right| \right\} \\
&= \sum_{i=1}^2 \sum_{t=1}^2 \tau(t) \left[\mathbb{E}_{\tilde{\mathbf{H}}_i} \left\{ \log_2 \left| \tilde{\mathbf{H}}_i \mathbf{\Lambda}_i(t) \tilde{\mathbf{H}}_i^H + \mathbf{I}_N \right| \right\} \right. \\
&\quad \left. - \mathbb{E}_{\tilde{\mathbf{H}}_i} \left\{ \log_2 \left| \tilde{\mathbf{H}}_i \bar{\mathbf{\Lambda}}_i(t) \tilde{\mathbf{H}}_i^H + \mathbf{I}_N \right| \right\} \right] \\
&= \sum_{i=1}^2 \sum_{t=1}^2 \tau(t) \left[\sum_{k=1}^{2N} \log_2 \left(\frac{1 + N\alpha_{i,1}(t)\lambda_{tik}}{1 + N\alpha_{i,2}(t)\bar{\lambda}_{tik}} \right) \right. \\
&\quad \left. + N \log_2 \left(\frac{\alpha_{i,2}(t)}{\alpha_{i,1}(t)} \right) + N(\alpha_{i,1}(t) - \alpha_{i,2}(t)) \log_2 e \right], \quad (23)
\end{aligned}$$

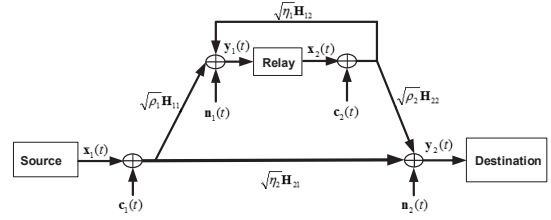


Fig. 2. The signal flow diagram of a two-hop full-duplex MIMO relay system.

where λ_{tik} and $\bar{\lambda}_{tik}$ is defined in (8) and $0 < \alpha_{i,1}(t), \alpha_{i,2}(t) < 1$ satisfies the following nonlinear equation

$$\alpha_{i,1}(t) + \sum_{k=1}^{2N} \frac{\alpha_{i,1}(t)\lambda_{tik}}{N\alpha_{i,1}(t)\lambda_{tik} + 1} = 1, \quad (24)$$

$$\alpha_{i,2}(t) + \sum_{k=1}^{2N} \frac{\alpha_{i,2}(t)\bar{\lambda}_{tik}}{N\alpha_{i,2}(t)\bar{\lambda}_{tik} + 1} = 1. \quad (25)$$

We can use the bisection method to compute $\alpha_{i,1}(t)$, since the left hand side of (24) is monotonically increasing functions of $\alpha_{i,1}(t)$. Same argument also holds for $\alpha_{i,2}(t)$. It is shown in the simulations that (23) is an accurate approximation of (8) even when N is as small as three.

With the simplified closed-form expression (23), the problem (10)-(12) can be solved by the GP method similar to the one developed in Section IV-A. We only need the gradient of the objective function (23) with respect to $d_{lm}(t)$, $l = 1, 2$, $m = 1, \dots, N$, $t = 1, 2$, which is given by

$$\begin{aligned}
\frac{\partial \bar{I}(\mathbf{d}_1, \mathbf{d}_2)}{\partial d_{lm}(t)} &= \tau(t) \sum_{i=1}^2 \sum_{k=1}^{2N} \left[\frac{N\alpha_{i,1}(t)\lambda'_{tik}}{1 + N\alpha_{i,1}(t)\lambda_{tik}} \right. \\
&\quad \left. - \frac{N\alpha_{i,2}(t)\bar{\lambda}'_{tik}}{1 + N\alpha_{i,2}(t)\bar{\lambda}_{tik}} \right], \quad (26)
\end{aligned}$$

where λ'_{tik} and $\bar{\lambda}'_{tik}$ are defined in (36) and (37) in Appendix, respectively. Note that since $\alpha_{i,1}(t)$ and $\alpha_{i,2}(t)$ are coefficients and are not functions of $d_{lm}(t)$, they can be treated as constants in the gradient expression [50].

V. FULL-DUPLEX RELAY SYSTEMS

In this section, we study the performance of a decode-and-forward FD relay system that suffers from self-interference, where all nodes are equipped with multiple antennas. The source node transmits signal streams to the destination node via the relay node and the direct link as shown in Fig. 2. We assume that the instantaneous CSI is not used by the transmitters and an imperfect CSI is used by the receiver. It can be seen from Fig. 2 that the system model of a FD relay is similar to that of the bi-directional FD system in Fig. 1.

For the relay system, we still assume that the relay uses N transmit antennas and N receive antennas in either FD mode or HD mode. For relay, the direction of reception is generally different from the direction of transmission. If directional antennas are used, the transmit antennas and the receive antennas should face different directions. And hence,

even if the HD mode is considered, the relay still should use N antennas for transmission and N antennas for reception at any given time. For power efficiency, directional antennas are a much better choice than omnidirectional antennas.

After the partial self-interference cancelation at the relay node, the received signal at the relay node and the destination is given by

$$\mathbf{y}_R(t) = \tilde{\mathbf{y}}_1(t), \quad (27)$$

$$\mathbf{y}_D(t) = \tilde{\mathbf{y}}_2(t) + \sqrt{\eta_2} \tilde{\mathbf{H}}_{21} \mathbf{x}_1(t), \quad (28)$$

where $\tilde{\mathbf{y}}_i(t)$, $i = 1, 2$ is defined in (2). Unlike the relay node, where the partial self-interference is possible, the destination node can not cancel the interference term $\sqrt{\eta_2} \tilde{\mathbf{H}}_{21} \mathbf{x}_1(t)$ resulting from the direct link, but adds it to the total noise $\mathbf{v}_2(t)$ in (3). (If the direct link is strong, the optimal scheme may switch to direct transmission as shown in [35]. But we do not consider this scenario). For fixed τ , the lower bound of the averaged ergodic mutual information of the decode-and-forward FD relay system over two time slots can be written as [51]

$$\bar{I}(\mathbf{Q}_1, \mathbf{Q}_2) = \min \{ \bar{I}_1(\mathbf{Q}_1, \mathbf{Q}_2), \bar{I}_2(\mathbf{Q}_1, \mathbf{Q}_2) \}, \quad (29)$$

where $\bar{I}_i(\mathbf{Q}_1, \mathbf{Q}_2)$, $i = 1, 2$ is defined in (5). The only difference is that the covariance matrix of the total noise $\tilde{\Sigma}_2(t)$ (4) in $\bar{I}_2(\mathbf{Q}_1, \mathbf{Q}_2)$ has the additional term $\eta_2 \tilde{\mathbf{H}}_{21} \mathbf{Q}_1(t) \tilde{\mathbf{H}}_{21}^H$ because of the additional term in (28).

A. Maximization of the Ergodic Mutual Information of the FD Relay System

In this subsection, we aim at maximizing the ergodic mutual information (29) by choosing the transmit covariance matrices $\mathbf{Q}_1(t)$ and $\mathbf{Q}_2(t)$, $t = 1, 2$, subject to per link power constraints and subsequently optimize over τ . Similar to Section IV, we consider fast fading channels in which the statistical CSI is assumed to be known at the transmitting nodes to design the optimal power schedules. This problem can be formulated as

$$\max_{\mathbf{Q}_1, \mathbf{Q}_2} \min \{ \bar{I}_1(\mathbf{Q}_1, \mathbf{Q}_2), \bar{I}_2(\mathbf{Q}_1, \mathbf{Q}_2) \} \quad (30)$$

$$\text{s.t.} \quad \sum_{t=1}^2 \tau(t) \text{tr} \{ \mathbf{Q}_i(t) \} \leq P_i, \quad i = 1, 2, \quad (31)$$

$$\mathbf{Q}_i(t) \geq 0, \quad i = 1, 2. \quad (32)$$

Applying the link equalizing algorithm proposed in [26], the objective function $\min \{ \bar{I}_1(\mathbf{Q}_1, \mathbf{Q}_2), \bar{I}_2(\mathbf{Q}_1, \mathbf{Q}_2) \}$ in (30) can be replaced with a ζ -weighted sum-rate problem, i.e., $\zeta \bar{I}_1(\mathbf{Q}_1, \mathbf{Q}_2) + (1 - \zeta) \bar{I}_2(\mathbf{Q}_1, \mathbf{Q}_2)$, where ζ is computed using bisection method (see Section IV-A of [26] for more details about the link-equalizing algorithm). Therefore, the ζ -weighted sum-rate optimization problem can be expressed as

$$\max_{\mathbf{Q}_1, \mathbf{Q}_2} \sum_{i=1}^2 \zeta(i) \bar{I}_i(\mathbf{Q}_1, \mathbf{Q}_2) \quad (33)$$

$$\text{s.t.} \quad \sum_{t=1}^2 \tau(t) \text{tr} \{ \mathbf{Q}_i(t) \} \leq P_i, \quad i = 1, 2, \quad (34)$$

$$\mathbf{Q}_i(t) \geq 0, \quad i = 1, 2, \quad (35)$$

where $\zeta(1) = \zeta$ and $\zeta(2) = 1 - \zeta$. Since the optimization problem (33)-(35) has a similar structure with (9), GP method proposed in Section IV-A can be applied to solve (33)-(35). Note that at each bisection step to compute ζ , GP method is used. The closed-form ergodic mutual information expression of the relay system can be obtained similar to (8). Due to the additional term in (28), the only modification required is on the term $\lambda_{2,2}(t)$ in $\Lambda_2(t)$ and $\bar{\Lambda}_2(t)$, which is modified as

$$\lambda_{2,2}(t) = \eta_2 \frac{\mathbf{d}_1(t) + \sigma_t^2 \mathbf{1}_N}{c_2(t)}.$$

Similarly the gradient of the objective function (33) can be obtained similar to (13). The only modification is on the terms λ'_{t2k} and $\bar{\lambda}'_{t2k}$, which are given at the top of the following page.

VI. SIMULATION RESULTS

In this section, we study the performance of the proposed FD MIMO bi-directional communication system through numerical simulations as a function of the averaged SNR, the nominal INR, the number of antennas N , the channel estimation errors $\sigma_{e,ij}^2$ and the transmitter impairments σ_t^2 . For all simulation examples, we set the same channel estimation error for all links, i.e., $\sigma_{e,ij}^2 = \sigma_e^2$, $i, j \in \{1, 2\}$. The Armijo parameters are selected as $\sigma = 0.1$, $\theta = 0.5$, and the stopping threshold of the GP algorithm is chosen as $\epsilon = 10^{-5}$. For simplicity, we focus on the case of $\eta_1 = \eta_2 = \eta$ and the same average transmit power for each node (i.e., $P_i = N$, $i = 1, 2$). Thus, the averaged SNR for all desired links is defined as $\text{SNR}_i = \rho_i N$, $i = 1, 2$ and the nominal INR for all interfering links $\text{INR}_i = \text{INR} = \eta N$, $i = 1, 2$. Since the nominal INR and the averaged SNR_i , $i = 1, 2$ are quasi static, we assume that their values can be obtained with relatively high precision, so we treat them as deterministic parameters. To optimize the HD scheme, we use the GP method to solve the problem (9) with the HD constraint of $\mathbf{Q}_1(2) = \mathbf{Q}_2(1) = \mathbf{0}$. To show the importance of using two time slots, we compare our FD system using two data transmission slots (FD2) with the FD system using only one data transmission slot (FD1). In the FD1 scheme, the same source covariance matrices are used for both time slots, i.e., $\mathbf{Q}_1(1) = \mathbf{Q}_1(2)$ and $\mathbf{Q}_2(1) = \mathbf{Q}_2(2)$. Since the GP algorithm only converges to a locally optimal solution, we use the output of the HD scheme as the initialization of the FD scheme. For the maximization problem (9), the time-sharing coefficient τ can be optimized over the grid $\tau \in \{0.1, 0.2, \dots, 0.9\}$ [26].

In the first example, we compare the exact and approximate closed-form expressions of the lower bound ergodic mutual information of the FD2 system using (8) and (23), respectively, for different number of antennas. We set $\text{SNR}_i = \text{SNR} = 20\text{dB}$, $i = 1, 2$, $\sigma_e^2 = 0.01$ and $\sigma_t^2 = -30\text{dB}$. It can be seen from Fig. 3 that the ergodic mutual information of the FD2 system is always equal to or greater than that of the HD system (the reason is explained in Fig. 4). It can also be seen from Fig. 3 that the asymptotic closed-form expression for the ergodic mutual information is an accurate approximation even when the number of antennas is as small as $N = 3$. Unless

$$\lambda'_{t2k} = \begin{cases} -\rho_2^2 c_2(t)^{-2} \sigma_{e,22}^2 (d_{2k}(t) + \sigma_t^2), & l = 2 \text{ and } k \neq m \text{ and } k \leq N \\ \frac{\rho_2^2}{c_2(t)} - \rho_2^2 c_2(t)^{-2} \sigma_{e,22}^2 (d_{2k}(t) + \sigma_t^2), & l = 2 \text{ and } k = m \text{ and } k \leq N \\ -\rho_2 \eta_2 c_2(t)^{-2} \sigma_{e,22}^2 (d_{1k}(t) + \sigma_t^2), & l = 2 \text{ and } k > N \\ -\rho_2 \eta_2 c_2(t)^{-2} \sigma_{e,21}^2 (d_{2k}(t) + \sigma_t^2), & l \neq 2 \text{ (} l = 1 \text{) and } k \leq N \\ -\eta_2^2 c_2(t)^{-2} \sigma_{e,21}^2 (d_{1(k-N)}(t) + \sigma_t^2), & l \neq i \text{ (} l = j \text{) and } k \neq N + m \text{ and } k > N \\ \frac{\eta_2}{c_2(t)} - \eta_2^2 c_2(t)^{-2} \sigma_{e,21}^2 (d_{1(k-N)}(t) + \sigma_t^2), & l \neq i \text{ (} l = j \text{) and } k = N + m \text{ and } k > N \end{cases}$$

$$\bar{\lambda}'_{t2k} = \begin{cases} -\rho_2^2 c_2(t)^{-2} \sigma_{e,22}^2 \sigma_t^2, & l = 2 \text{ and } k \leq N \\ -\rho_2 \eta_2 c_2(t)^{-2} \sigma_{e,22}^2 (d_{1(k-N)}(t) + \sigma_t^2), & l = 2 \text{ and } k > N \\ -\rho_2 \eta_2 c_2(t)^{-2} \sigma_{e,21}^2 \sigma_t^2, & l \neq 2 \text{ (} l = 1 \text{) and } k \leq N \\ -\eta_2^2 c_2(t)^{-2} \sigma_{e,21}^2 (d_{1(k-N)}(t) + \sigma_t^2), & l \neq i \text{ (} l = j \text{) and } k \neq N + m \text{ and } k > N \\ \frac{\eta_2}{c_2(t)} - \eta_2^2 c_2(t)^{-2} \sigma_{e,21}^2 (d_{1(k-N)}(t) + \sigma_t^2), & l \neq i \text{ (} l = j \text{) and } k = N + m \text{ and } k > N \end{cases}$$

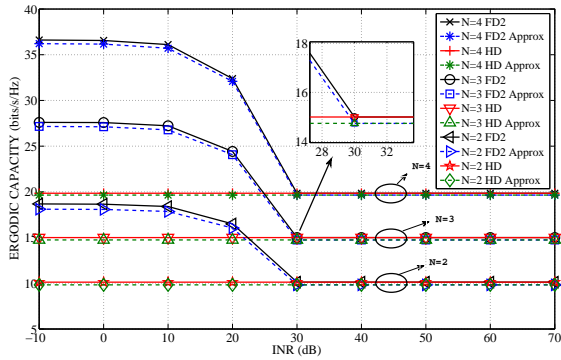


Fig. 3. Ergodic mutual information comparison of the FD2 and HD systems with different number of antennas versus INR. Here $\text{SNR} = 20\text{dB}$, $\sigma_e^2 = 0.01$, $\sigma_t^2 = -30\text{dB}$.

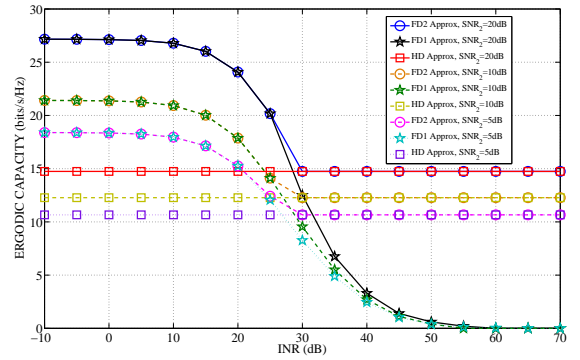


Fig. 4. Ergodic mutual information comparison of the FD2, FD1, and HD systems versus INR for different SNR_2 values. Here $N = 3$, $\text{SNR}_1 = 20\text{dB}$, $\sigma_e^2 = 0.01$, $\sigma_t^2 = -30\text{dB}$.

otherwise stated, hereafter we adopt the asymptotic closed-form ergodic mutual information expression, since it has a much lower computational complexity.

In the next example, we investigate the impact of INR on the ergodic mutual information of the FD2, FD1, and HD schemes with $N = 3$, $\text{SNR}_2 = 20\text{dB}$, $\sigma_e^2 = 0.01$ and $\sigma_t^2 = -30\text{dB}$ for different SNR_1 values. As expected, it can be observed from Fig. 4 that the HD scheme is invariant to INR. For the low-to-mid values of INR, the FD2 scheme has the FD system behavior and it switches to the HD scheme at the high values

of INR. The FD1 scheme performs similar to the FD2 scheme at low-to-mid values of INR, but its performance drops below that of the HD scheme for larger values of INR. The use of two distinct data time slots gives the freedom to switch to the HD signaling when the power of the self-interference channel is high (where the HD scheme is optimal), while the FD1 system forces FD signaling at each time slot, regardless of the strength of the self-interference channel [26], [27].

In our third example, we examine the value of INR that FD2 converges to HD. Fig. 5 demonstrates that the behavior

$$c_{tikn}(\mathbf{\Lambda}_i(t)) = \frac{(-1)^{N-n-1}}{n!} \lambda_{tik}^{N-1} \left(\prod_{h \neq k} (\lambda_{tik} - \lambda_{tih}) \right)^{-1} b_{tikn}(\mathbf{\Lambda}_i(t)), \quad (30)$$

$$b_{tikn}(\mathbf{\Lambda}_i(t)) = \begin{cases} \sum_{1 \leq j_1 < \dots < j_{N-n-1} \leq 2N}^{j_r \neq k} \lambda_{tj_1} \dots \lambda_{tj_{N-n-1}}, & n = 0, \dots, N-2 \\ 1, & n = N-1, \end{cases} \quad (31)$$

$$Q(n, \lambda_{tik}) = \int_0^\infty \ln(1+x) x^n e^{-x/\lambda_{tik}} dx \quad (32)$$

$$= \sum_{r=0}^n \frac{n! (-1)^{(n-r)}}{(n-r)!} \lambda_{tik}^{r+1} e^{1/\lambda_{tik}} S_1\left(\frac{1}{\lambda_{tik}}\right) + \sum_{r=1}^n \sum_{s=0}^{r-1} \sum_{h=0}^{r-s-1} \frac{n! (-1)^{(n-r)} \lambda_{tik}^{h+s+2}}{(n-r)! (r-s-h-1)! (r-s)}.$$

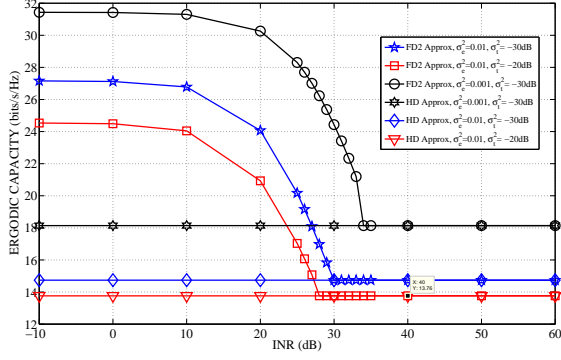


Fig. 5. Ergodic mutual information comparison of the FD2 and HD systems versus INR for different σ_e^2 and σ_t^2 values. Here $N = 3$, SNR = 20dB.

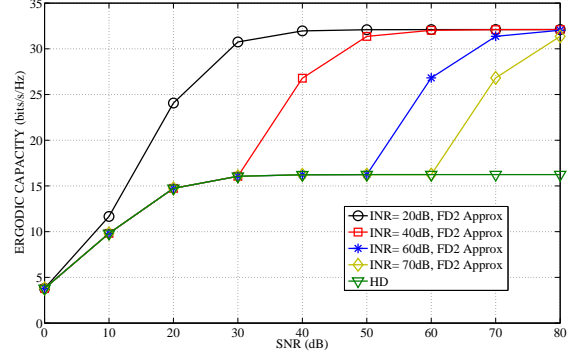


Fig. 6. Ergodic mutual information comparison of the FD2 and HD systems for different INR values versus SNR. Here $N = 3$, $\sigma_e^2 = 0.01$, $\sigma_t^2 = -30$ dB.

of convergence depends on σ_e^2 and σ_t^2 values.

In our fourth example, we examine the ergodic mutual information of the FD2 and HD systems versus SNR_i = SNR, $i = 1, 2$ for various fixed values of INR. We choose $N = 3$, $\sigma_e^2 = 0.01$ and $\sigma_t^2 = -30$ dB. It can be observed from Fig. 6 that at low INR, the system operates in the FD mode

for all values of SNR, since SNR mostly dominates INR. At high INR, the system operates in the HD mode at low values of SNR (since INR dominates SNR), but switches to the FD mode as SNR increases, since SNR starts to dominate INR.

For our fifth example, in Fig. 7, we examine the ergodic

$$\begin{aligned}
 c'_{tikn}(\mathbf{\Lambda}_i(t)) &= \frac{(-1)^{N-n-1}(N-1)}{n!} \lambda_{tik}^{N-2} \lambda'_{tik} \left(\prod_{h \neq k} (\lambda_{tik} - \lambda_{tih}) \right)^{-1} b_{tikn}(\mathbf{\Lambda}_i(t)) \\
 &\quad - \frac{(-1)^{N-n-1}}{n!} \lambda_{tik}^{N-1} \sum_{j=1, j \neq k}^{2N} \left[\left(\prod_{h \neq k} (\lambda_{tik} - \lambda_{tih}) \right)^{-1} (\lambda_{tik} - \lambda_{tij})^{-1} (\lambda'_{tik} - \lambda'_{tij}) \right] \\
 &\quad \times b_{tikn}(\mathbf{\Lambda}_i(t)) + \frac{(-1)^{N-n-1}}{n!} \lambda_{tik}^{N-1} \left(\prod_{h \neq k} (\lambda_{tik} - \lambda_{tih}) \right)^{-1} b'_{tikn}(\mathbf{\Lambda}_i(t)), \quad (33)
 \end{aligned}$$

$$\begin{aligned}
 Q'(n, \lambda_{tik}) &= \sum_{r=0}^n \frac{n!(-1)^{(n-r)}}{(n-r)!} \lambda_{tik}^r \lambda'_{tik} \left[(r+1)e^{1/\lambda_{tik}} S_1\left(\frac{1}{\lambda_{tik}}\right) - \frac{1}{\lambda_{tik}} e^{1/\lambda_{tik}} S_1\left(\frac{1}{\lambda_{tik}}\right) + 1 \right] \\
 &\quad + \sum_{r=1}^n \sum_{s=0}^{r-1} \sum_{h=0}^{r-s-1} \frac{n!(-1)^{(n-r)}(h+s+2)\lambda_{tik}^{h+s+1}}{(n-r)!(r-s-h-1)!(r-s)} \lambda'_{tik}, \quad (34)
 \end{aligned}$$

$$b'_{tikn}(\mathbf{\Lambda}_i(t)) = \begin{cases} \sum_{1 \leq j_1 < \dots < j_{N-n-1} \leq 2N} \lambda'_{tij_1} \lambda_{tij_2} \dots \lambda_{tij_{N-n-1}} \\ + \sum_{1 \leq j_1 < \dots < j_{N-n-1} \leq 2N} \lambda_{tij_1} \lambda'_{tij_2} \dots \lambda_{tij_{N-n-1}} \\ + \dots + \sum_{1 \leq j_1 < \dots < j_{N-n-1} \leq 2N} \lambda_{tij_1} \lambda_{tij_2} \dots \lambda'_{tij_{N-n-1}}, & n = 0, \dots, N-2 \\ 0, & n = N-1, \end{cases} \quad (35)$$

$$\lambda'_{tik} = \begin{cases} -\rho_i^2 c_i(t)^{-2} \sigma_{e,ii}^2 (d_{ik}(t) + \sigma_t^2), & l = i \text{ and } k \neq m \text{ and } k \leq N \\ \frac{\rho_i}{c_i(t)} - \rho_i^2 c_i(t)^{-2} \sigma_{e,ii}^2 (d_{ik}(t) + \sigma_t^2), & l = i \text{ and } k = m \text{ and } k \leq N \\ -\rho_i \eta_i c_i(t)^{-2} \sigma_{e,ii}^2 \sigma_t^2, & l = i \text{ and } k > N \\ -\rho_i \eta_i c_i(t)^{-2} \sigma_{e,ij}^2 (d_{ik}(t) + \sigma_t^2), & l \neq i \text{ (} l = j \text{) and } k \leq N \\ -\eta_i^2 c_i(t)^{-2} \sigma_{e,ij}^2 \sigma_t^2, & l \neq i \text{ (} l = j \text{) and } k > N \end{cases}. \quad (36)$$

$$\bar{\lambda}'_{tik} = \begin{cases} -\rho_i^2 c_i(t)^{-2} \sigma_{e,ii}^2 \sigma_t^2, & l = i \text{ and } k \leq N \\ -\rho_i \eta_i c_i(t)^{-2} \sigma_{e,ii}^2 \sigma_t^2, & l = i \text{ and } k > N \\ -\rho_i \eta_i c_i(t)^{-2} \sigma_{e,ij}^2 \sigma_t^2, & l \neq i \text{ (} l = j \text{) and } k \leq N \\ -\eta_i^2 c_i(t)^{-2} \sigma_{e,ij}^2 \sigma_t^2, & l \neq i \text{ (} l = j \text{) and } k > N \end{cases}. \quad (37)$$

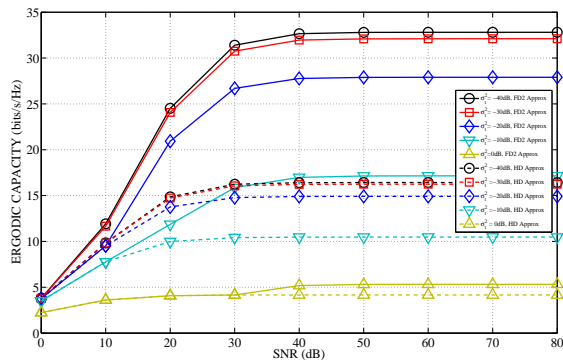


Fig. 7. Ergodic mutual information comparison of the FD2 and HD systems with different σ_e^2 values versus SNR. Here $N = 3$, $\text{INR} = 20\text{dB}$, $\sigma_e^2 = 0.01$.

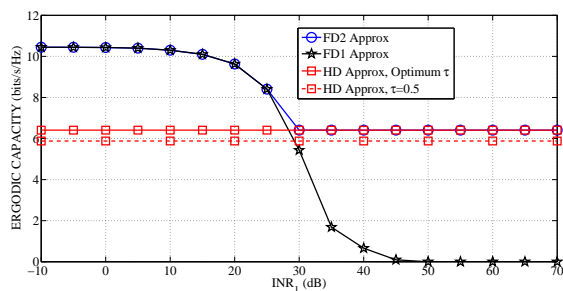


Fig. 8. Ergodic mutual information comparison of the FD2, FD1, and HD relay systems versus INR_1 . Here $N = 3$, $\text{SNR}_1 = 20\text{dB}$, $\text{SNR}_2 = 10\text{dB}$, $\text{INR}_2 = 0\text{dB}$, $\sigma_e^2 = 0.01$, $\sigma_t^2 = -30\text{dB}$.

mutual information of the FD2 and HD systems versus $\text{SNR}_i = \text{SNR}$, $i = 1, 2$ for various values of σ_e^2 . We choose $N = 3$, $\text{INR} = 20\text{dB}$ and $\sigma_e^2 = 0.01$.

In the next example, we consider MIMO FD relay systems. We obtain similar results as MIMO bi-directional FD system as shown in Fig. 8. In particular, the relay node operates in the FD mode when the self-interference is weak, and as the self-interference increases, we observe a transition of the relay node to the HD mode. Similar to [26], we can also observe that, compared to using fixed value $\tau = 0.5$, the optimization of τ gives a small rate improvement.

For our last example, in Fig. 9, we investigate the role of channel estimation errors on the lower bound of the ergodic mutual information for MIMO FD relay systems.

VII. CONCLUSION

In this work, we have studied the ergodic mutual information maximization of two FD MIMO radio systems (bi-directional system and relay system) that suffer from a (digitally manageable residual) self-interference under a fast fading channel model. The source covariance matrices are treated as a function of time and/or frequency within any given time/frequency band so that both spatial and temporal freedoms of the source covariance matrices can be exploited. Since the globally optimal solution is difficult to obtain due to the non-convex nature of the problem, a gradient projection algorithm is developed to optimize the power allocation vectors

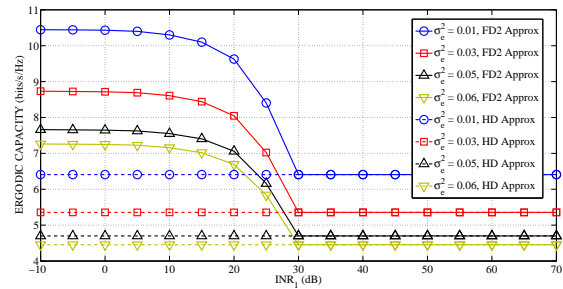


Fig. 9. Ergodic mutual information comparison of the FD2 and HD systems with different channel estimation errors versus INR_1 . Here $N = 3$, $\text{SNR}_1 = 20\text{dB}$, $\text{SNR}_2 = 10\text{dB}$, $\text{INR}_2 = 0\text{dB}$, $\sigma_t^2 = -30\text{dB}$.

at two respective nodes with the knowledge of statistical CSI useful for the transmitters. In addition to an exact closed-form ergodic mutual information expression, we introduced a much simpler asymptotic closed-form ergodic mutual information expression, which is shown to be an accurate approximation and in turn simplifies the computation of the power allocation vectors. It is shown through numerical simulations that the ergodic mutual information increases with the number of antennas, decreases as the channel estimation error and/or the transmitter distortion increases. Moreover, it is demonstrated that at a high self-interference power level, the optimal power schedule reduces to the HD mode, and at a low self-interference power level, the optimal power schedule switches to the FD mode.

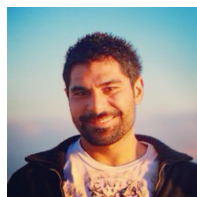
APPENDIX

See (30)-(32) shown at the bottom of page 9 for the parameters $c_{tikn}(\Lambda_i(t))$ and $Q(n, \lambda_{tik})$ in (8). See (33)-(37) shown at the bottom of the previous page for the definition of the parameters in (13).

REFERENCES

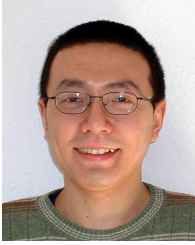
- [1] A. C. Cirik, Y. Rong, and Y. Hua, "Ergodic mutual information of full-duplex MIMO radios with residual self-interference," in *Proc. 2013 Asilomar Conf. Signals, Syst. Computers*, Nov. 2013.
- [2] T. Riihonen, S. Werner, and R. Wichman, "Comparison of full-duplex and half-duplex modes with a fixed amplify-and-forward relay," in *Proc. IEEE Wireless Communications and Networking Conference*, April 2009.
- [3] T. Riihonen, S. Werner, R. Wichman, and E. Zaccarias B., "On the feasibility of full-duplex relaying in the presence of loop interference," in *Proc. 10th IEEE Workshop on Signal Processing Advances in Wireless Communications*, June 2009.
- [4] D. Kim, H. Ju, S. Park, D. Hong, "Effects of channel estimation error on full-duplex two-way networks," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 9, pp. 4666-4672, Nov. 2013.
- [5] W. Cheng, X. Zhang, H. Zhang, "Optimal dynamic power control for full-duplex bidirectional-channel based wireless networks," *IEEE INFOCOM*, pp. 3120-3128, April 2013.
- [6] T. Riihonen, M. Vehkaperä, R. Wichman, "Large-system analysis of rate regions in bidirectional full-duplex MIMO link: Suppression versus cancellation," *47th Annual Conference on Information Sciences and Systems (CISS)*, pp. 1-6, March 2013.
- [7] S. Chen, M.A. Beach and J.P. McGeehan, "Division-free duplex for wireless applications," *Electronics Letters*, vol. 34, no. 2, 1998.
- [8] D. W. Bliss, P. A. Parker, and A. R. Margetts, "Simultaneous transmission and reception for improved wireless network performance," in *Proc. IEEE 14th Workshop on Statistical Signal Processing*, pp. 26-29, Aug. 2007.

- [9] J. I. Choi, M. Jain, K. Srinivasan, P. Levis, and S. Katti, "Achieving single channel, full duplex wireless communication," in *Proc. Mobicom 2010*, pp. 1-12, Sep. 2010.
- [10] M. Jain, J. I. Choi, T. Kim, D. Bharadia, K. Srinivasan, S. Seth, P. Levis, S. Katti, and P. Sinha, "Practical, real-time, full duplex wireless," in *Proc. Mobicom*, 2011.
- [11] 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.
- [12] M. Knox, "Single antenna full duplex communications using a common carrier," in *13th Annual IEEE Wireless and Microwave Technology Conference (WAMICON)*, pp. 1-6, April 2012.
- [13] S. Hong, J. Mehlman, and S. Katti, "Picasso: Flexible RF and spectrum slicing," in *SIGCOMM*, 2012.
- [14] D. Bharadia, E. McMillin, and S. Katti, "Full duplex radios," in *ACM SIGCOMM'13*, Aug. 2013.
- [15] Y. Hua, P. Liang, Y. Ma, A. Cirik and Q. Gao, "A method for broadband full-duplex MIMO radio," *IEEE Signal Processing Letters*, vol. 19, no. 12, pp. 793-796, Dec 2012.
- [16] D. Bharadia and S. Katti, "Full duplex MIMO radios," *USENIX NSDI*, 2014.
- [17] A. Sahai, G. Patel, and A. Sabharwal, "Pushing the limits of full-duplex: Design and real-time implementation", Online at arXiv, 2011.
- [18] B. Chun, E. Jeong, J. Joung, Y. Oh, and Y. H. Lee, "Pre-nulling for self-interference suppression in full-duplex relays," in *Proc. Asia-Pacific Signal and Information Processing Association Annual Summit Conference (APSIPA ASC)*, Sapporo, Oct. 2009.
- [19] P. Lioliou, M. Viberg, M. Coldrey, and F. Athley, "Self-interference suppression in full-duplex MIMO relays," in *Proc. Asilomar Conf. Signals Syst. Comput.*, (Pacific Grove, CA), pp. 658-662, Oct. 2010.
- [20] T. Riihonen, S. Werner, and R. Wichman, "Residual self-interference in full-duplex MIMO relays after null-space projection and cancellation," in *Proc. 44th Annual Asilomar Conference on Signals, Systems, and Computers*, November 2010.
- [21] T. Riihonen, S. Werner, and R. Wichman, "Transmit power optimization for multiantenna decode-and-forward relays with loopback self-interference from full-duplex operation," in *45th Annual Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, pp. 1408-1412, Nov. 2011.
- [22] T. Riihonen, S. Werner, and R. Wichman, "Mitigation of loopback self-interference in full-duplex MIMO relays," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 5983-5993, Dec. 2011.
- [23] B. Chun and H. Park; , "A spatial-domain joint-nulling method of self-interference in full-duplex relays," *IEEE Communications Letters*, vol.16, no.4, pp.436-438, April 2012.
- [24] Y. Hua, "An overview of beamforming and power allocation for MIMO relays," in *Proc. IEEE Military Commun. Conf.*, (San Jose, CA), pp. 375-380, Nov. 2010.
- [25] T. Riihonen, A. Balakrishnan, K. Haneda, S. Wyne, S. Werner, and R. Wichman, "Optimal eigenbeamforming for suppressing self interference in full-duplex MIMO relays", *45th Annual Conference on Information Sciences and Systems (CISS)*, Baltimore, Maryland, March 2011.
- [26] B. P. Day, A. R. Margetts, D. W. Bliss, and P. Schniter, "Full-duplex MIMO relaying: Achievable rates under limited dynamic range," *IEEE J. Selected Areas Commun.*, vol. 30, no. 8, pp. 1541-1553, Sep. 2012.
- [27] B. P. Day, A. R. Margetts, D. W. Bliss, and P. Schniter, "Full-duplex bidirectional MIMO: Achievable rates under limited dynamic range," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3702-3713, July 2012.
- [28] S. P. Herath, T. Le-Ngoc, "Sum-rate performance and impact of self-interference cancellation on full-duplex wireless systems," *IEEE 24th International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, pp. 881-885, Sept. 2013.
- [29] D. W. Bliss, T. M. Hancock, and P. Schniter, "Hardware phenomenological effects on cochannel full-duplex MIMO relay performance," in *IEEE Asilomar Conference on Signals, Systems and Computers*, pp. 34-39, Nov. 2012.
- [30] M. Vehkaper, T. Riihonen, and R. Wichman, "Asymptotic analysis of full-duplex bidirectional MIMO link with transmitter noise", *24th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pp. 1265-1270, Sept. 2013.
- [31] B. Yin, M. Wu, C. Studer, J. R. Cavallaro, and J. Lilleberg, "Full-duplex in large-scale wireless systems", in *47th Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, Nov. 2013.
- [32] M. Vehkaper, M. Girnyk, T. Riihonen, R. Wichman, and L. Rasmussen, "On achievable rate regions at large-system limit in full-duplex wireless local access", *First International Black Sea Conference on Communications and Networking (BlackSeaCom)*, pp. 7-11, July 2013.
- [33] A. C. Cirik, R. Wang, and Y. Hua, "Weighted-sum-rate maximization for bi-directional full-duplex MIMO systems," in *Proc. 2013 Asilomar Conf. Signals, Syst. Computers*, Nov. 2013.
- [34] A. C. Cirik, J. Zhang, M. Haardt, and Y. Hua, "Sum-rate maximization for bi-directional full-duplex MIMO systems under multiple linear constraints", in *Proc. 15th IEEE Workshop on Signal Processing Advances in Wireless Communication (SPAWC'2014)*, Toronto, June, 2014.
- [35] T. Riihonen, S. Werner, R. Wichman, "Hybrid full-duplex/half-duplex relaying with transmit power adaptation," *IEEE Trans. on Wireless Communications*, vol. 10, no. 9, pp. 3074-3085, Sept. 2011.
- [36] A. Lozano and A. M. Tulino, "Capacity of multiple-transmit multiple-receive antenna architectures," *IEEE Trans. Inf. Theory*, vol. 48, pp. 3117-3128, Dec. 2002.
- [37] Y. Rong and Y. Hua, "Optimal power schedule for distributed MIMO links," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 2896-2900, Aug. 2008.
- [38] D. W. K. Ng, E. S. Lo, and R. Schober, "Dynamic resource allocation in MIMO-OFDMA systems with full-duplex and hybrid relaying," *IEEE Transactions on Communications*, vol. 60, no. 5, pp. 1291-1304, 2012.
- [39] T. Yoo and A. J. Goldsmith, "Capacity and optimal power allocation for fading MIMO channels with channel estimation error," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2203-2214, May 2006.
- [40] H. Suzuki, T. V. A. Tran, I. B. Collings, G. Daniels, and M. Hedley, "Transmitter noise effect on the performance of a MIMO-OFDM hardware implementation achieving improved coverage," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 6, pp. 867-876, Aug. 2008.
- [41] C. Studer, M. Wenk, and A. Burg, "MIMO transmission with residual Tx-RF impairments", in *International ITG Workshop on Smart Antennas*, pp. 189-196, Feb. 2010.
- [42] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, pp. 951-963, Apr. 2003.
- [43] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Europ. Trans. Telecommun.*, vol. 10, pp. 585-595, Nov. 1999.
- [44] P. J. Smith, S. Roy, and M. Shafi, "Capacity of MIMO systems with semicorrelated flat fading," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2781-2788, Oct. 2003.
- [45] I. Abramowitz and M. Stegun, Eds., *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*. Washington, DC: U.S. Government Printing Office, 1972.
- [46] Y. Rong, Y. Hua, A. Swami, and A. L. Swindlehurst, "Space-time power schedule for distributed MIMO links without instantaneous channel state information at the transmitting nodes," *IEEE Trans. Signal Process.*, vol. 56, no. 2, pp. 686-700, Feb. 2008.
- [47] D. P. Bertsekas, *Nonlinear Programming*, 2nd. edition. Belmont, MA: Athena Scientific, 1995.
- [48] D. N. C. Tse and S. V. Hanly, "Linear multiuser receivers: Effective interference, effective bandwidth and user capacity," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 641-657. Mar. 1999.
- [49] S. Shamai (Shitz) and S. Verdú, "The effect of frequency-flat fading on the spectral efficiency of CDMA," *IEEE Trans. Inform. Theory*, vol. 47, May 2001.
- [50] Y. Rong and Y. Hua, "Space-time power scheduling of MIMO links – Fairness and QoS considerations," *IEEE J. Selected Topics Signal Process.*, vol. 2, no. 2, pp. 171-180, Apr. 2008.
- [51] B. Wang, J. Zhang, and A. Host-Madsen, "On the capacity of MIMO relay channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 29-43, Jan. 2005.



Ali Cagatay Cirik received the B.S and M.S. degrees in telecommunications and electronics engineering from Sabanci University, Istanbul, Turkey, in 2007 and 2009, respectively, and Ph.D. degree in electrical engineering from University of California, Riverside in 2014. Currently, he is working as a research scientist at Centre for Wireless Communications, University of Oulu, Finland.

His industry experience includes internships at Mitsubishi Electric Research Labs, Cambridge, MA, in 2012 and at Broadcom Corporation, Irvine, CA, in 2013. His primary research interests are full-duplex communication, MIMO signal processing, and convex optimization.



Yue Rong (S'03-M'06-SM'11) received the Ph.D. degree (summa cum laude) in electrical engineering from the Darmstadt University of Technology, Darmstadt, Germany, in 2005.

He was a Post-Doctoral Researcher with the Department of Electrical Engineering, University of California, Riverside, from February 2006 to November 2007. Since December 2007, he has been with the Department of Electrical and Computer Engineering, Curtin University, Bentley, Australia, where he is currently an Associate Professor. His

research interests include signal processing for communications, wireless communications, underwater acoustic communications, applications of linear algebra and optimization methods, and statistical and array signal processing.

Dr. Rong was a recipient of the Best Paper Award at the 2011 International Conference on Wireless Communications and Signal Processing, the Best Paper Award at the 2010 Asia-Pacific Conference on Communications, and the Young Researcher of the Year Award of the Faculty of Science and Engineering at Curtin University in 2010. He is an Editor of the IEEE WIRELESS COMMUNICATIONS LETTERS, a Guest Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS special issue on theories and methods for advanced wireless relays, and was a TPC Member for the IEEE ICC, WCSP, IWCMC, and ChinaCom.



Yingbo Hua (S'86-M'88-SM'92-F'02) received a B.S. degree (1982) from Southeast University, Nanjing, China, a M.S. degree (1983) and a Ph.D. degree (1988) from Syracuse University, Syracuse, NY. He was a Lecturer (1990-1992), a Senior Lecturer (1993-1995), and a Reader and Associate Professor (1996-2000) with the University of Melbourne, Australia. He was a Visiting Faculty Member with Hong Kong University of Science and Technology (1999-2000), and a Consultant with Microsoft Research, WA (summer 2000). Since 2001, he has been with

the University of California at Riverside, where he is a Senior Full Professor. Dr. Hua has served as Editor, Guest Editor, Member of Editorial Board and/or Member of Steering Committee for IEEE Transactions on Signal Processing, IEEE Signal Processing Letters, EURASIP Signal Processing, IEEE Signal Processing Magazine, IEEE Journal of Selected Areas in Communications, and IEEE Wireless Communication Letters. He has been a Member of IEEE Signal Processing Society's Technical Committees for Underwater Acoustic Signal Processing, Sensor Array and Multichannel Signal Processing, and Signal Processing for Communication and Networking. He has served as member of Technical and/or Advisory Committees for over forty international conferences and workshops. He has authored over three hundreds of articles and coedited three volumes of books, with more than six thousands of citations, in the fields of Sensing, Signal Processing and Communications. He is a Fellow of IEEE and AAAS.