Ergodic Mutual Information of Full-Duplex MIMO Radios with Residual Self-Interference

Ali Cagatay Cirik*, Yue Rong† and Yingbo Hua*

* Department of Electrical Engineering, University of California Riverside, Riverside, CA, 92521
† Department of Electrical and Computer Engineering, Curtin University, Bentley, WA 6102, Australia
Email: {acirik, yhua}@ee.ucr.edu, y.rong@curtin.edu.au

Abstract—We study the theoretical performance of a full-duplex multiple-input multiple-output (MIMO) bi-directional communication system. We focus on the effect of the residual self-interference due to channel estimation errors and transmitter impairments. We assume that the instantaneous channel state information (CSI) at the transmitting nodes is not known and the CSI at the receiving nodes is imperfect. To maximize the system ergodic mutual information, which is a non-convex function of power allocation vectors at the nodes, a gradient projection (GP) 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 the nodes to achieve higher sum ergodic mutual information. It is observed through the simulations that the algorithm reduces to a full-duplex scheme when the nominal residual self-interference is low, or to a half-duplex scheme when the nominal residual self-interference is high.

I. INTRODUCTION

This paper concerns radio frequency (RF) bi-directional wireless communication systems, in which two radios can be used to communicate directly 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. The potential advantages of full-duplex radios over half-duplex radios have recently motivated active research in several different aspects, ranging from information theory and signal processing based on mathematical models to hardware experimentation and real system demonstration [1]-[9].

A fundamental enabler for full-duplex radios is known as the self-interference cancellation (SIC). When a full-duplex radio transmits, it causes self-interference which must be canceled satisfactorily. SIC can be done by different methods, at different stages, and to different degrees, along the receiving chain of a full-duplex radio. Recent experimental results [1]-[3] show that the simultaneous transmission and reception in the same frequency band is possible using advanced antenna designs and passive/active cancellation techniques.

The increased degrees of freedom offered by multiple-input multiple-output (MIMO) systems lead to a range of methods for self-interference cancellation. Most of these methods can be called transmit beamforming techniques [3]-[8]. All of the methods [1]-[8] have their limitations, and some level of residual self-interference is still expected for all SIC methods known to date.

In this paper, we assume that the residue self-interference is low enough for further baseband processing. We will focus on a theoretical performance of the full-duplex radios under the effect of the residual self-interference. We will also refer to residual self-interference simply as self-interference. Particularly, we develop an algorithm to maximize a lower bound on the ergodic mutual information of a full-duplex bi-directional MIMO system under a transmitter distortion model for fast fading channels where the instantaneous CSI is not known at the transmitters and imperfectly known at the receivers. We develop a Gradient Projection (GP) method to solve this non-convex optimization problem. The most relevant prior work is [9]. A main difference between this paper and that one is that we consider fast fading channels and they considered slow fading channels[1]. Note that unlike slow fading channels assumed in [9], fast fading channel does not require any instantaneous CSI feedback from the receiver. In the absence of instantaneous CSI at the transmitting nodes, the knowledge of some statistical properties of the CSI is necessary for designing optimal power schedules. We will aim to optimize an ergodic system mutual information with respect to the statistical distributions of the CSI. It is shown through numerical simulations that at a high self-interference power level, the optimal power schedule reduces to the half-duplex mode and at a low self-interference power level, the optimal power schedule switches to the full-duplex mode.

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. $\mathrm{E}_{\mathbf{H}}\{\cdot\}$ stands for the statistical expectation with respect to the channel matrix $\mathbf{H}$; $\mathbf{I}_N$ denotes an $N \times N$ identity matrix; $\mathrm{tr}\{\cdot\}$ 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; $diag\{a_1, \cdots, a_n\}$ denotes a diagonal matrix with the diagonal elements given by $a_1, \cdots, a_n$. $\mathcal{CN}\left(\mu, \sigma^2\right)$ 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.

---

[1]Unlike this work which only considers the residual transmitter distortion, the authors in [9] consider both transmitter and receiver distortion.
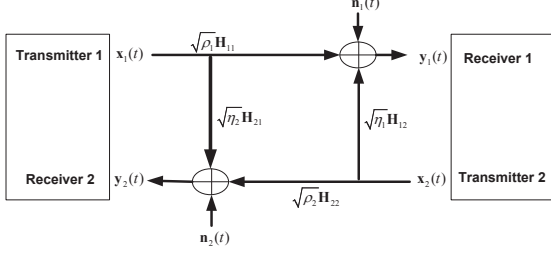
Fig. 1. System Model of Bi-directional Full-Duplex MIMO System.

II. SYSTEM MODEL

In this section, we describe the system model of a FD bi-directional MIMO system between two nodes. The signals mentioned below are defined in complex baseband. We consider MIMO wireless systems, where two nodes are equipped with multiple antennas and exchange information simultaneously in a two-way communication. We assume that each node has the same number of transmit and receive antennas, which is denoted by N. We partition the data transmission period under consideration or control into two normalized equal-length time slots. The reason of this partition will be clear in the simulation results.

As illustrated in Fig. 1, the receiver $i \in \{1, 2\}$ receives signals from both transmitters via MIMO channels $\mathbf{H}_{ij} \in \mathbb{C}^{N \times N}, (i, j) \in \{1, 2\}$. All the channel matrices are assumed to be random and independent, i.e., the entries of each matrix are independent and identically distributed (i.i.d.) circular complex Gaussian variables with zero mean and unit variance. We adopt the channel model used in [4], [5], [6], [7], where the estimation error is modeled as $\Delta\mathbf{H}_{ij} = \mathbf{H}_{ij} - \tilde{\mathbf{H}}_{ij}$, where $\tilde{\mathbf{H}}_{ij}$ and $\Delta\mathbf{H}_{ij}$ are uncorrelated, and the entries of $\Delta\mathbf{H}_{ij}$ are zero mean circularly symmetric complex Gaussian with variance $\sigma_{e,ij}^2$. 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.

Practical RF transmitters suffer from the presence of impairments resulting from non-linear distortions in the power amplifier, phase noise, or from IQ-imbalance. The impact of these impairments can be reduced through calibration and compensation algorithms. However, due to parameter estimation errors and the mismatch between the physical RF chain and the model of the compensation algorithm, some residual transmitter distortion still persists. The measurement results by [10] indicate that an i.i.d. additive Gaussian noise model accurately describes the sum of all such residual transmitter impairments. Furthermore, the authors assume sufficient decoupling of the transmitter RF chains such that the relevant impairments are statistically independent across transmit antennas. The resulting Gaussian model for the ith transmitter noise is given by $\mathbf{c}_i(t) \sim \mathcal{CN}\left(\mathbf{0}, \sigma_t^2 \mathbf{I}_N\right), \ i = 1, 2$.

The $N \times 1$ received signal vector at the ith receiver can be written as

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

where ρ_i denotes the average power gain of the ith transmitter-receiver link, η_i denotes the average power gain of the self-interference channel, $\mathbf{x}_i(t) \in \mathbb{C}^N, \ i = 1, 2$ is the signal vector transmitted by node i within time slot t and is Gaussian distributed with zero mean and covariance matrix $\mathbb{E}\left\{\mathbf{x}_i(t)\mathbf{x}_i(t)^H\right\} = \mathbf{Q}_i(t)$, and $\mathbf{n}_i(t) \in \mathbb{C}^N$ is the additive white Gaussian noise (AWGN) vector with zero mean and identity covariance matrix $\mathbb{E}\left\{\mathbf{n}_i(t)\mathbf{n}_i(t)^H\right\} = \mathbf{I}_N$. We assume that $\mathbf{n}_i(t)$ is independent of $\mathbf{x}_i(t)$.

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 $\mathbf{y}_i(t)$ [9]. The interference cancelled signal can then be written as

$$
\tilde{\mathbf{y}}_i(t) = \sqrt{\rho_i}\tilde{\mathbf{H}}_{ii}\mathbf{x}_i(t) + \mathbf{v}_i(t) \tag{2}
$$

where

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

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

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

where 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. 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 [11]. 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 [11], which was also used in [9]. The lower bound of the ergodic sum mutual information of the system averaged over two time slots can be written as

$$
\bar{I}\left(\mathbf{Q}_1, \mathbf{Q}_2\right) \tag{5}
$$
$$
= \frac{1}{2}\sum_{i=1}^{2}\sum_{t=1}^{2}\mathbb{E}_{\tilde{\mathbf{H}}_{ii}, \tilde{\mathbf{H}}_{ij}}\left\{\log_2\left|\mathbf{I}_N + \rho_i\tilde{\mathbf{H}}_{ii}\mathbf{Q}_i(t)\tilde{\mathbf{H}}_{ii}^H\tilde{\mathbf{\Sigma}}_i(t)^{-1}\right|\right\}
$$

where $\mathbf{Q}_i \triangleq \left[\mathbf{Q}_i^T(1), \mathbf{Q}_i^T(2)\right]^T$, $i = 1, 2$.

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 $\mathbf{Q}_i(t) = \mathbf{U}_i(t)\mathbf{D}_i(t)\mathbf{U}_i(t)^H$, where $\mathbf{U}_i(t)$ is the unitary matrix of eigenvectors, and $\mathbf{D}_i(t) = diag\{d_{i1}(t), d_{i2}(t), \ldots, d_{iN}(t)\}$ is a diagonal matrix of all eigenvalues. For convenience, we will use the column vectors $\mathbf{d}_i(t) \triangleq [d_{i1}(t), d_{i2}(t), \ldots, d_{iN}(t)]^T$, $i = 1, 2$. Since $\tilde{\mathbf{H}}_{ij}$, $(i, j) \in \{1, 2\}$ has i.i.d. Gaussian entries and $\mathbf{U}_j(t)$ is unitary, the statistics of $\tilde{\mathbf{H}}_{ij}\mathbf{U}_j(t)$ is identical to that of $\tilde{\mathbf{H}}_{ij}$ [12, Lemma 5]. We can then rewrite (5) as

$$
\bar{I}(\mathbf{d}_1, \mathbf{d}_2)
$$
$$
= \frac{1}{2}\sum_{i=1}^{2}\sum_{t=1}^{2} \mathrm{E}_{\tilde{\mathbf{H}}_{ii}, \tilde{\mathbf{H}}_{ij}}\left\{\log_2\left|\mathbf{I}_N + \rho_i\tilde{\mathbf{H}}_{ii}\mathbf{D}_i(t)\tilde{\mathbf{H}}_{ii}^H\boldsymbol{\Sigma}_i(t)^{-1}\right|\right\}
$$
$$
= \frac{1}{2}\sum_{i=1}^{2}\sum_{t=1}^{2}\left[\mathrm{E}_{\tilde{\mathbf{H}}_i}\left\{\log_2\left|\tilde{\mathbf{H}}_i\boldsymbol{\Lambda}_i(t)\tilde{\mathbf{H}}_i^H + \mathbf{I}_N\right|\right\}\right.
$$
$$
\left. - \mathrm{E}_{\tilde{\mathbf{H}}_i}\left\{\log_2\left|\tilde{\mathbf{H}}_i\bar{\boldsymbol{\Lambda}}_i(t)\tilde{\mathbf{H}}_i^H + \mathbf{I}_N\right|\right\}\right] \quad (6)
$$

where

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

Here $\mathbf{1}_N$ is an $N \times 1$ column vector of ones.

Because of the space limitation, we omit the exact closed form expression of (6) which can be found in the journal version of this paper. For simplicity, herein we consider a much simpler closed form expression of $\bar{I}(\mathbf{d}_1, \mathbf{d}_2)$ which is shown to be an accurate approximation even for systems with a small number of antennas [13]. This simplification is based on an asymptotical form of $\bar{I}(\mathbf{d}_1, \mathbf{d}_2)$ when $N \to \infty$ as proposed in [13]. Applying the result in [13], the sum ergodic mutual information in (6) can be approximated as

$$
\bar{I}(\mathbf{d}_1, \mathbf{d}_2) \approx \sum_{i=1}^{2}\sum_{t=1}^{2}\tau(t)\left[\sum_{k=1}^{2N}\log_2\left(\frac{1 + N\alpha_{i,1}(t)\lambda_{tik}}{1 + N\alpha_{i,2}(t)\bar{\lambda}_{tik}}\right)\right.
$$
$$
\left. + N\log_2\left(\frac{\alpha_{i,2}(t)}{\alpha_{i,1}(t)}\right) + N\left(\alpha_{i,1}(t) - \alpha_{i,2}(t)\right)\log_2 e\right] \quad (7)
$$

where $\lambda_{tik} \triangleq [\boldsymbol{\Lambda}_i(t)]_{k,k}$ and $\bar{\lambda}_{tik} \triangleq [\bar{\boldsymbol{\Lambda}}_i(t)]_{k,k}$, $k = 1, \ldots, 2N$ denote the (k, k)th element of matrix $\boldsymbol{\Lambda}_i(t)$ and

$\bar{\boldsymbol{\Lambda}}_i(t)$, respectively. Here $0 < \alpha_{i,1}(t), \alpha_{i,2}(t) < 1$ satisfies the following nonlinear equation

$$
\alpha_{i,1}(t) + \sum_{k=1}^{2N}\frac{\alpha_{i,1}(t)\lambda_{tik}}{N\alpha_{i,1}(t)\lambda_{tik} + 1} = 1. \quad (8)
$$
$$
\alpha_{i,2}(t) + \sum_{k=1}^{2N}\frac{\alpha_{i,2}(t)\bar{\lambda}_{tik}}{N\alpha_{i,2}(t)\bar{\lambda}_{tik} + 1} = 1. \quad (9)
$$

We can use the bisection method to compute $\alpha_{i,1}(t)$, since the left hand side of (8) is monotonically increasing functions of $\alpha_{i,1}(t)$. Same argument also holds for $\alpha_{i,2}(t)$.

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 $\mathbf{Q}_i(t)$, $(i, t) \in \{1, 2\}$ subject to per node average power constraints $\frac{1}{2}\sum_{t=1}^{2}\mathrm{tr}\{\mathbf{Q}_i(t)\} \leq P_i$, where P_i is the averaged transmit power from the ith transmitter. 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. Particularly, we will design the power schedule to maximize an ergodic system mutual information which is averaged over the statistical distribution of the channel matrices. The optimization problem can be formulated as

$$
\max_{\mathbf{d}_1, \mathbf{d}_2} \quad \bar{I}(\mathbf{d}_1, \mathbf{d}_2) \quad (10)
$$
$$
\text{s.t.} \quad \|\mathbf{d}_i\|_1 = 2P_i, \quad i = 1, 2 \quad (11)
$$
$$
\mathbf{d}_i \geq 0, \quad i = 1, 2 \quad (12)
$$

where $\bar{I}(\mathbf{d}_1, \mathbf{d}_2)$ is given in (7). Here $\|.\|_1$ 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.

We now develop a numerical algorithm to obtain a locally optimal solution to the problem (10)-(12). by employing GP method. 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 $d_{lm}(t)$, $l = 1, 2$, $m = 1, \ldots, N$, $t = 1, 2$, is given by

$$
\frac{\partial\bar{I}(\mathbf{d}_1, \mathbf{d}_2)}{\partial d_{lm}(t)} = \frac{1}{2}\sum_{i=1}^{2}\sum_{k=1}^{2N}\left[\frac{N\alpha_{i,1}(t)\lambda'_{tik}}{1 + N\alpha_{i,1}(t)\lambda_{tik}}\right.
$$
$$
\left. - \frac{N\alpha_{i,2}(t)\bar{\lambda}'_{tik}}{1 + N\alpha_{i,2}(t)\bar{\lambda}_{tik}}\right] \quad (13)
$$

where λ'_{tik} and $\bar{\lambda}'_{tik}$ are defined at the bottom of the following page.

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

gradient as $\mathbf{g}_i \triangleq \left[\frac{\partial \bar{I}(\mathbf{d}_1, \mathbf{d}_2)}{\partial d_{i1}(1)}, \ldots, \frac{\partial \bar{I}(\mathbf{d}_1, \mathbf{d}_2)}{\partial d_{iN}(2)} \right]^T$. Then taking a step along the positive gradient direction, the power allocation vector is updated as $\hat{\mathbf{d}}_i = \mathbf{d}_i + s\mathbf{g}_i$, $i = 1, 2$ where s is a scalar of step size. 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 $\tilde{\mathbf{d}}_i$ 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

$$\min_{\tilde{\mathbf{d}}_i} \quad \left\| \tilde{\mathbf{d}}_i - \hat{\mathbf{d}}_i \right\|^2 \tag{14}$$

$$\text{s.t.} \quad \|\tilde{\mathbf{d}}_i\|_1 = 2P_i, \qquad \tilde{\mathbf{d}}_i \geq 0, \ i = 1, 2. \tag{15}$$

The problem (14)-(15) is convex and can be efficiently solved by the Lagrange multiplier method. It turns out that the problem (14)-(15) has a water-filling solution which is given by $\tilde{d}_{ik} = \left[\hat{d}_{ik} - \frac{\mu}{2} \right]^+$, $k = 1, \ldots, 2N$, $i = 1, 2$, 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 $\tilde{d}_{ik}$ back into (15) and solving the nonlinear equation $\sum_{k=1}^{2N} \left[\hat{d}_{ik} - \frac{\mu}{2} \right]^+ = 2P_i$, $i = 1, 2$. The left hand side of this equation is a piecewise linear function and monotonically decreasing with respect to μ, so it can solved using the bisection method. After computing the two most important steps of the GP algorithm, the details of the rest of the GP algorithm can be found in [9].

V. 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 interference-to-noise ratio (INR), the channel estimation errors $\sigma_{e,ij}^2$ and the transmitter impairments σ_t^2. For simplicity, we consider $\sigma_{e,ij}^2 = \sigma_e^2$, $i, j \in \{1, 2\}$, $\rho_i = \rho$ and $\eta_i = \eta$ and $P_i = N$, $i = 1, 2$. Thus, we have the same averaged SNR for all desired links $\text{SNR}_i = \text{SNR} = \rho N$, $i = 1, 2$ and the same nominal INR for all interfering links $\text{INR}_i = \text{INR} = \eta N$, $i = 1, 2$. The Armijo parameters are selected as $\sigma = 0.1$, $\theta = 0.5$, and the stopping threshold of the GP algorithm is chosen as $\epsilon = 10^{-5}$. To show the
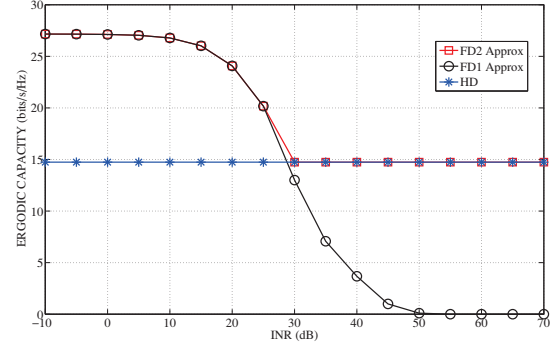

Fig. 2. Ergodic mutual information comparison of FD2, FD1, and HD systems versus INR. Here $N = 3$, SNR = 20dB, $\sigma_e^2 = 0.01$, $\sigma_t^2 = -30$dB.

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). 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.

In our first example, we investigate the impact of INR on the ergodic mutual information of the FD2, FD1, and HD schemes. As expected, it can be observed from Fig. 2 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. This is similar to FD systems for slow fading channels [9].

In the second example, we investigate the role of channel estimation errors on the lower bound of the ergodic mutual information (7). It can be seen from Fig. 3 that as the channel estimation errors increases, the ergodic mutual information of

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

$$\bar{\lambda}'_{tik} = \begin{cases} -\rho_i^2 c_i(t)^{-2} \sigma_{e,ii}^2 \sigma_t^2, & l = i \text{ and } k \leq N \\ -\rho_i \eta_i c_i(t)^{-2} \sigma_{e,ii}^2 \sigma_t^2, & l = i \text{ and } k > N \\ -\rho_i \eta_i c_i(t)^{-2} \sigma_{e,ij}^2 \sigma_t^2, & l \neq i \ (l = j) \text{ and } k \leq N \\ -\eta_i^2 c_i(t)^{-2} \sigma_{e,ij}^2 \sigma_t^2, & l \neq i \ (l = j) \text{ and } k > N \end{cases} .$$
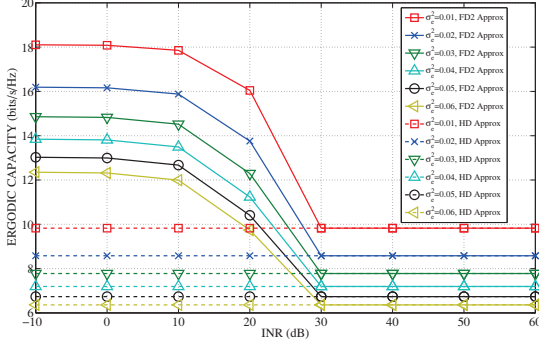
Fig. 3. Ergodic mutual information comparison of the FD2 and HD systems with different channel estimation errors versus INR. Here $N = 2$, SNR $= 20$dB, $\sigma_t^2 = -30$dB.

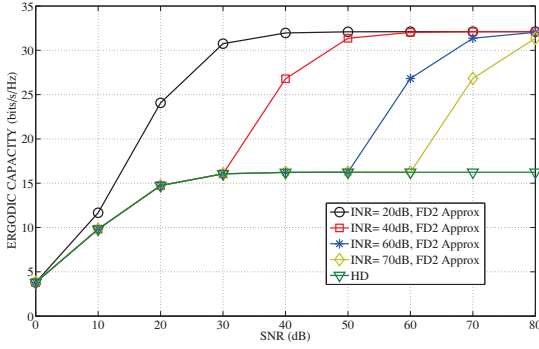Fig. 5. Ergodic mutual information comparison of the FD2 system with different σ_t^2 values versus SNR. Here $N = 3$, INR $= 20$dB, $\sigma_e^2 = 0.01$.

Fig. 4. Ergodic mutual information comparison of the FD2 and HD systems versus SNR. Here $N = 3$, $\sigma_e^2 = 0.01$, $\sigma_t^2 = -30$dB.

both the FD2 and HD systems decreases. The gap between ergodic mutual information curves diminishes as the σ_e^2 increases.

In our third example, we examine the ergodic mutual information of the FD2 and HD systems versus SNR for various fixed values of INR. It can be observed from Fig. 4 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.

In our last example, we examine the ergodic mutual information of the FD2 systems versus SNR for various values of σ_t^2. It can be observed from Fig. 5 that as σ_t^2 decreases, the ergodic mutual information increases and the gap between the curves diminishes.

VI. CONCLUSION

In this work, we studied the maximization of the asymptotic ergodic mutual information for FD MIMO bi-directional system that suffers from a residual self-interference. Since the globally optimal solution is difficult to obtain due to the non-convex nature of the problem, a gradient projection algorithm
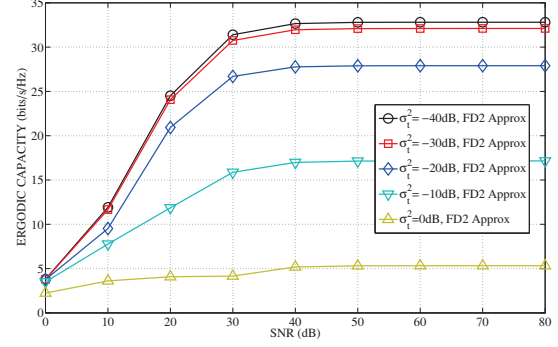
is developed to optimize the power allocation vectors at two nodes with the knowledge of statistical CSI at the transmitters. It is shown through numerical simulations that at a high self-interference power level, the optimal power schedule reduces to the HD mode and at a low self-interference power level, the optimal power schedule switches to the FD mode.

REFERENCES

[1] 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.

[2] M. Duarte and A. Sabharwal, "Full-duplex wireless communications using off-the-shelf radios: Feasibility and first results," in *Proc. 2010 Asilomar Conf. Signals, Syst. Computers*, pp. 1558-1562, Nov. 2010.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] B. P. Day, D. W. Bliss, A. R. Margetts, 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.

[10] 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.

[11] 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.

[12] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Europ. Trans. Telecommun.*, vol. 10, pp. 585-595, Nov. 1999.

[13] 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.