Weighted Sum-Rate Maximization for Full-Duplex MIMO Interference Channels

Ali Cagatay Cirik, Rui Wang, *Member, IEEE*, Yingbo Hua, *Fellow, IEEE*, and Matti Latva-aho, *Senior Member, IEEE*

***Abstract*—We consider a K link multiple-input multiple-output (MIMO) interference channel, where each link consists of two full-duplex (FD) nodes exchanging information simultaneously in a bi-directional communication fashion. The nodes in each pair suffer from self-interference due to operating in FD mode, and inter-user interference from other links due to simultaneous transmission at each link. We consider the transmit and receive filter design for weighted sum-rate (WSR) maximization problem subject to sum-power constraint of the system or individual power constraints at each node of the system. Based on the relationship between WSR and weighted minimum-mean-squared-error (WMMSE) problems for FD MIMO interference channels, we propose a low complexity alternating algorithm which converges to a local WSR optimum point. Moreover, we show that the proposed algorithm is not only applicable to FD MIMO interference channels, but also applicable to FD cellular systems in which a base station (BS) operating in FD mode serves multiple uplink (UL) and downlink (DL) users operating in half-duplex (HD) mode, simultaneously. It is shown in simulations that the sum-rate achieved by FD mode is higher than the sum-rate achieved by baseline HD schemes.**

***Index Terms*—Bi-directional, full-duplex, MIMO interference channel, multi-user, self interference, transceiver designs.**

I. Introduction

THE INCREASING demand for high data rates with the proliferation of wireless services is calling for powerful communication technologies that exploit the current spectrum more efficiently. Half-duplex wireless communication systems, or commonly known as time-division duplex or frequency-division duplex, employ two orthogonal channels to transmit and receive, and thus they cannot achieve the maximal spectral efficiency. Full-duplex wireless communication system, which enables simultaneous transmission and reception at the same time in the same frequency band, has recently gained considerable interest in academia [1]–[40], as a promising technique to potentially double the link capacity, and increase the spectral efficiency targeted by the next generation wireless communication systems.

The potential of high spectral efficiency gain of a full-duplex radio over half-duplex radio has recently attracted several research groups. In particular, full-duplex relaying technology has been investigated in [1], since relays are effective in mitigating the effects of multipath fading, pathloss and shadowing, and enhancing the quality of service of the users at the edge of the cells. In addition to relay nodes, small cells which provide improved cellular coverage is considered to be suitable for deployment of full-duplex technology due to low transmit powers, short transmission distances and low mobility. A small cell network where a full-duplex base station (BS) serves multiple uplink (UL) and downlink (DL) users simultaneously has been considered in [2]–[5]. Moreover, cognitive radios, which is also a promising technology in wireless communications to enhance the spectrum efficiency of the fixed spectrum allocation policies, can be deployed in full-duplex. A full-duplex cognitive radio can transmit and sense the transmission status of other nodes [6], [7]. Full-duplex technology has also been studied to mitigate the problems associated with medium access control (MAC) layer, such as hidden terminals, large delays, and congestion [8]–[10].

The limiting factor on the performance of full-duplex systems is the strong self-interference at the front-end of the receiver created by the signal leakage from the transmitter antennas of a full-duplex node to its own receiver antennas. Unless this self-interference is canceled satisfactorily, a radio transceiver cannot perform full-duplex operation. Recently several research groups have developed methods for self-interference cancelation. These works include the transmit beamforming (spatial domain suppression) methods in [11]–[16], in which the self-interference is cancelled at the front-end of the receiver by generating a cancellation signal based on the transmit signal in the baseband. Promising results from experimental research that demonstrate the feasibility of full-duplex transmission using the off-the-shelf hardware are also available in [9], [10], [17]–[24], although RF front-end interference cancellation is still an on-going research topic [25].

However, due to imperfections of radio devices such as amplifier non-linearity, phase noise, and I/Q channel imbalance, the self-interference cannot be canceled completely in reality. Therefore, optimization problems (power allocation, transceiver beamforming, etc.) related to full-duplex systems

Manuscript received September 15, 2014; revised November 26, 2014 and January 17, 2015; accepted January 21, 2015. Date of publication February 2, 2015; date of current version March 13, 2015. Part of this work was submitted to 21th European Wireless (EW) Conference, Budapest, Hungary, May 2015. The research leading to these results has received funding in part from Army Research Office under grant no. W911NF1210432 and University of California Office of the President under Grant no. PC-12-247260, in part from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement number 316369–project DUPLO, and in part from the Academy of Finland under Grant 260755–Project JULIET. The associate editor coordinating the review of this paper and approving it for publication was M. Matthaiou.

A. C. Cirik and M. Latva-aho are with the Centre for Wireless Communications, University of Oulu, 90570 Oulu, Finland (e-mail: acirik@ee.oulu.fi; matti.latvaaho@ee.oulu.fi).

R. Wang is with the Department of Electronics and Information Engineering, Tongji University, Shanghai 201804, China (e-mail: ruiwang@tongji.edu.cn).

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

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

Digital Object Identifier 10.1109/TCOMM.2015.2397886

under this residual self-interference were considered in [2], [3], [26], [27]–[37]. In particular, by exploiting both spatial and temporal freedoms of the source covariance matrices of the multiple-input multiple-output (MIMO) links, the sum-rate maximization problem for full-duplex bi-directional MIMO channels has been studied for slow fading and fast fading channels in [27] and [28], respectively. The authors in [29] study the impact of residual self-interference on sum-rate performance under two situations: channel state information (CSI) is available only at receiver, and CSI is available both at the transmitter and the receiver. Two sequential convex programming-based algorithms to solve the sum-rate maximization problem of full-duplex single user MIMO systems were proposed in [30], and the optimal power allocation algorithms to maximize the sum-rate of full-duplex orthogonal frequency division multiplexing (OFDM) bi-directional full-duplex systems are developed in [31]. Moreover, the weighted sum-rate maximization, mean-squared-error (MSE) minimization, signal-to-leakage-plus-noise ratio maximization (SLNR), total transmit power minimization, and distributed sum-rate maximization problems for bi-directional full-duplex systems have been considered in [32]–[35], and [36], respectively. A transmit-receive antenna pair selection based on the maximization of the sum-rate, and the minimization of the symbol-error rate is proposed for a bi-directional full-duplex system in [37].

To the best of our knowledge, all the papers on full-duplex bi-directional systems consider a single pair of nodes, exchanging information simultaneously [16], [27]–[40], and no paper has studied full-duplex systems under multiple pair of nodes, i.e., full-duplex MIMO interference channel. Recently, the interest on MIMO channels has migrated from point-to-point MIMO and MIMO downlink channels, to the MIMO interference channels, since it is the inherent model behind many practical problems [41]. With the increase of wireless devices that share the same frequency and time resources, interference becomes the key bottleneck that limits the throughput of communication networks. A study on the performance of cellular communication systems (open spectrum, multi-cell systems, etc.) where each cell causes interference to other cells can be carried out by focusing on MIMO interference channels [42].

In this paper, we develop a low complexity algorithm to compute the maximum weighted sum-rate of a K-link full-duplex MIMO interference channel, where each link has two full-duplex nodes exchanging information simultaneously. The nodes in each pair suffer from self-interference due to operating in full-duplex mode, and inter-user interference due to simultaneous transmission at all links. Two types of power constraints are considered. One is a power constraint on the sum-power of the system. The other is a power constraint on the power at each radio node. Under either constraint, the problem is non-convex. Following a distributed approach used in [43] for broadcast channels, we turn the weighted sum-rate problem into a weighted minimum-mean-squared problem, the latter of which is easier to solve. The distributed implementation for MIMO interference channels has been implemented in [44], [45], but they both have considered only half-duplex models.

Moreover, we show that the proposed algorithm can also be applied to full-duplex cellular systems. DL and UL channels are currently designed to operate in either separate time slots or separate frequency bands, and thus the radio resources have not been efficiently used in current wireless communications systems. In this paper, we consider a scenario where a full-duplex capable BS communicates with half-duplex UL and DL users at the same time slot over the same frequency band. In addition to self-interference channel at the BS, the co-channel interference (CCI) caused by the UL users to DL users is also taken into account, which increases the difficulty of the optimization problem further. Full-duplex multi-user systems has been investigated in [2], [3], [26]. However the CCI is not taken into account in [3], and single-antenna users are assumed in [2]. Moreover, [2], [3], [26] ignores several fundamental impediments of FD systems, i.e., transmitter and receiver distortion caused by non-ideal amplifiers, oscillators, ADCs/DACs, etc., i.e., several system parameters were ideally assumed. These practical considerations are carefully examined in this paper.

Simulation results show that the proposed full-duplex system outperforms the baseline half-duplex systems under moderate interference levels. In addition, the self-interference and inter-user interference levels that reduce the performance of the full-duplex system to a level below that of the baseline schemes is investigated.

A. Notations

The following notations are used in this paper. Matrices and vectors are denoted as bold capital and lowercase letters, respectively. $(\cdot)^T$ is the transpose; $(\cdot)^H$ is the conjugate transpose. $\mathbb{E}\{\cdot\}$ means the statistical expectation; $\mathbf{I}_N$ is the N by N identity matrix; $\mathbf{0}_{N\times M}$ is the N by M zero matrix; $\mathrm{tr}(\cdot)$ is the trace; $|\cdot|$ is the determinant; $\mathrm{diag}(\mathbf{A})$ is the diagonal matrix with the same diagonal elements as $\mathbf{A}$. $\mathcal{CN}(\mu,\sigma^2)$ denotes a complex Gaussian distribution with mean μ and variance σ^2. $\mathbb{C}^{N\times M}$ denotes the set of complex matrices with a dimension of N by M. $\perp$ denotes the statistical independence. We will use the following abbreviations: FD-full-duplex, HD-half-duplex, WSR-weighted sum-rate, WMMSE-weighted minimum-mean-squared error.

II. System Model

In this section, we describe the system model of a FD MIMO interference channel as seen in Fig. 1. The signals mentioned below are defined in complex baseband. We consider MIMO wireless channels, where each pair is equipped with multiple antennas and exchanges information simultaneously in a two way communication. We assume that the FD nodes in ith link have N_i and M_i transmit and receive antennas, respectively.

We also take into account the limited dynamic range (DR). Limited-DR is caused by non-ideal amplifiers, oscillators, analog-to-digital converters (ADCs), and digital-to-analog converters (DACs). We adopt the limited DR model in [27], which has also been commonly used in [14], [31]–[34], [36]. Particularly, at each receive antenna an additive white Gaussian "receiver distortion" with variance β times the energy of the undistorted received signal on that receive antenna is applied,

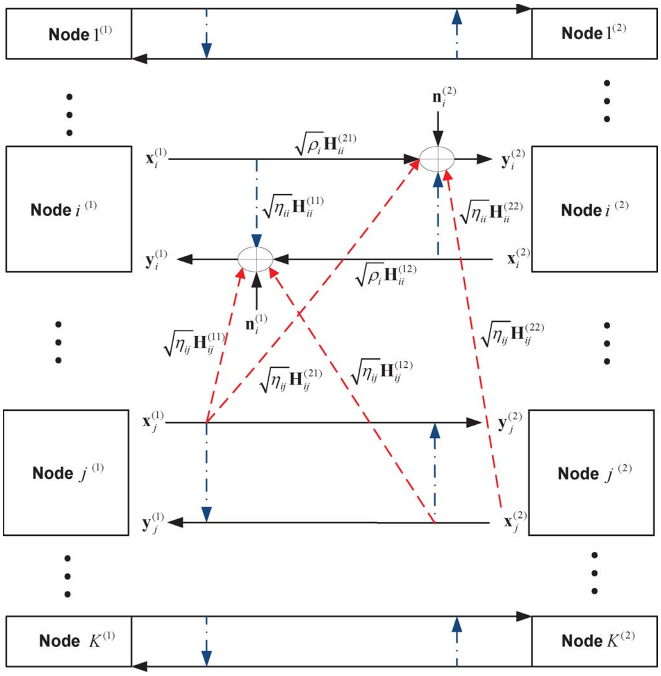

Fig. 1. Bi-directional full-duplex MIMO interference channel. $\longrightarrow$ denotes the interference between different pairs, and $-\cdot\rightarrow$ denotes the self-interference.

and at each transmit antenna, an additive white Gaussian "transmitter noise" with variance κ times the energy of the intended transmit signal is applied. This transmitter/receiver distortion model is valid, since it was shown by hardware measurements in [46] and [47] that the non-ideality of the transmitter and receiver chain can be approximated by an independent Gaussian noise model, respectively.

As illustrated in Fig. 1, the node $i^{(a)}, i \in \{1,\ldots,K\}$ and $a \in \{1,2\}$ receives signals from all the transmitters in the system via MIMO channels. $\mathbf{H}_{ii}^{(ab)} \in \mathbb{C}^{M_i\times N_i}$ is the channel between node a and b of the ith transmitter-receiver pair, where $b \in \{1,2\}$ and $b \neq a$. $\mathbf{H}_{ii}^{(aa)} \in \mathbb{C}^{M_i\times N_i}, a \in \{1,2\}$ denotes the self-interference channel of the node $i^{(a)}$. $\mathbf{H}_{ij}^{(ac)} \in \mathbb{C}^{M_i\times N_j}, (a,c) \in \{1,2\}$ denotes the interference channel from the transmitter antennas of the node c in the jth pair to the receiver antennas of the node a in the ith pair, $(i,j) \in \{1,\ldots,K\}$ and $j \neq i$. All the channel matrices are assumed to be mutually independent, and the entries of each matrix are independent and identically distributed (i.i.d.) circular complex Gaussian variables with zero mean and unit variance. We assume that CSI is only *locally* available at the transmitters, i.e., the transmitters are able to obtain the knowledge of channel coefficients which are directly connected to them [48]–[54], and the receiver at each link is able to obtain the channel coefficients and the interference-plus-noise covariance matrix at its only link.

The transmitted data streams of size d_i at the the node $i^{(a)}$ is denoted as $\mathbf{d}_i^{(a)} \in \mathbb{C}^{d_i}, i \in \{1,\ldots,K\}, a \in \{1,2\}$, and are assumed to be complex, zero mean, i.i.d. with

$$\mathbb{E}\left\{\mathbf{d}_i^{(a)}\right\} = \mathbf{0}_{d_i\times 1}, \tag{1}$$

$$\mathbb{E}\left\{\mathbf{d}_i^{(a)}\left(\mathbf{d}_j^{(b)}\right)^H\right\} = \begin{cases}\mathbf{I}_{d_i} & i=j \text{ and } a=b,\\ \mathbf{0}_{d_i\times d_j} & i\neq j \text{ or } a\neq b.\end{cases} \tag{2}$$

The $N_i \times 1$ signal vector transmitted by node $i^{(a)}$ is given by

$$\mathbf{x}_i^{(a)} = \mathbf{V}_i^{(a)}\mathbf{d}_i^{(a)}, \quad i=1,\ldots,K, a\in\{1,2\} \tag{3}$$

where $\mathbf{V}_i^{(a)} \in \mathbb{C}^{N_i\times d_i}$ represents the precoding matrix. $\mathbf{x}_i^{(a)}$ is assumed to be Gaussian distributed with zero mean and covariance matrix,

$$\mathbb{E}\left\{\mathbf{x}_i^{(a)}\left(\mathbf{x}_i^{(a)}\right)^H\right\} = \mathbf{V}_i^{(a)}\left(\mathbf{V}_i^{(a)}\right)^H.$$

We consider a FD bi-directional MIMO interference channel that suffers from self-interference and interference from other pairs. Thus, node $i^{(a)}$ receives a combination of the signals transmitted by all the transmitters and noise. The $M_i \times 1$ received signal at node $i^{(a)}$ is written as

$$\begin{aligned}\mathbf{y}_i^{(a)} &= \sqrt{\rho_i}\mathbf{H}_{ii}^{(ab)}\left(\mathbf{x}_i^{(b)}+\mathbf{c}_i^{(b)}\right)+\sqrt{\eta_{ii}}\mathbf{H}_{ii}^{(aa)}\left(\mathbf{x}_i^{(a)}+\mathbf{c}_i^{(a)}\right)\\ &\quad+\sum_{j\neq i}^{K}\sum_{c=1}^{2}\sqrt{\eta_{ij}}\,\mathbf{H}_{ij}^{(ac)}\left(\mathbf{x}_j^{(c)}+\mathbf{c}_j^{(c)}\right)+\mathbf{e}_i^{(a)}\\ &\quad+\mathbf{n}_i^{(a)}, i\in\{1,\ldots,K\},(a,b)\in\{1,2\},a\neq b.\end{aligned} \tag{4}$$

Here, $\mathbf{n}_i^{(a)} \in \mathbb{C}^{M_i}$ is the additive white Gaussian noise (AWGN) vector at node $i^{(a)}$ with zero mean and unit covariance matrix, and it is uncorrelated to all the transmitted signals. In (4), ρ_i denotes the average power gain of the ith transmitter-receiver pair, η_{ii} denotes the average power gain of the self-interference channel at the ith pair, and η_{ij} denotes the average power gain of the interference channel between the nodes at the ith and jth pair.

In (4), $\mathbf{c}_i^{(a)} \in \mathbb{C}^{N_i}, i \in \{1,\ldots,K\}, a \in \{1,2\}$ is the transmitter noise at the transmitter antennas of node $i^{(a)}$, which models the effect of limited transmitter DR and closely approximates the effects of additive power-amplifier noise, non-linearities in the DAC and phase noise. The covariance matrix of $\mathbf{c}_i^{(a)}$ is given by $\kappa(\kappa \ll 1)$ times the energy of the intended signal at each transmit antenna [27]. In particular $\mathbf{c}_i^{(a)}$ is modeled as

$$\mathbf{c}_i^{(a)} \sim \mathcal{CN}\left(\mathbf{0}, \kappa\mathrm{diag}\left(\mathbf{V}_i^{(a)}\left(\mathbf{V}_i^{(a)}\right)^H\right)\right), \tag{5}$$

$$\mathbf{c}_i^{(a)} \perp \mathbf{x}_i^{(a)}, \tag{6}$$

where, as mentioned before, $\perp$ denotes the statistical independence.

In (4), $\mathbf{e}_i^{(a)} \in \mathbb{C}^{M_i}, i \in \{1,\ldots,K\}, a \in \{1,2\}$ is the additive receiver distortion at the receiver antennas of node $i^{(a)}$, which models the effect of limited receiver DR and closely approximates the combined effects of additive gain-control noise, non-linearities in the ADC and phase noise. Each diagonal element of the covariance matrix of $\mathbf{e}_i^{(a)}$ is given by $\beta(\beta \ll 1)$ times the energy of the undistorted received signal at each receive antenna [27]. In particular, $\mathbf{e}_i^{(a)}$ is modeled as

$$\mathbf{e}_i^{(a)} \sim \mathcal{CN}\left(\mathbf{0}, \beta\mathrm{diag}\left(\mathbf{\Phi}_i^{(a)}\right)\right), \tag{7}$$

$$\mathbf{e}_i^{(a)} \perp \mathbf{u}_i^{(a)}, \tag{8}$$

where $\Phi_i^{(a)} = \text{Cov}\{\mathbf{u}_i^{(a)}\}$ and $\mathbf{u}_i^{(a)}$ is the undistorted received vector at the node $i^{(a)}$, i.e., $\mathbf{u}_i^{(a)} = \mathbf{y}_i^{(a)} - \mathbf{e}_i^{(a)}$.

Node $i^{(a)}$ knows the interfering codewords $\mathbf{x}_i^{(a)}$, so the self-interference term $\sqrt{\eta_{ii}}\mathbf{H}_{ii}^{(aa)}\mathbf{x}_i^{(a)}$ is known, and thus can be cancelled [27]. The interference canceled signal can then be written as

$$\begin{aligned}\tilde{\mathbf{y}}_i^{(a)} &= \mathbf{y}_i^{(a)} - \sqrt{\eta_{ii}}\mathbf{H}_{ii}^{(aa)}\mathbf{x}_i^{(a)} \\ &= \sqrt{\rho_i}\mathbf{H}_{ii}^{(ab)}\mathbf{x}_i^{(b)} + \mathbf{v}_i^{(a)},\end{aligned} \tag{9}$$

where $\mathbf{v}_i^{(a)}$ is the unknown interference-plus noise component of (9) after self-interference cancellation, and is given by

$$\begin{aligned}\mathbf{v}_i^{(a)} = \sqrt{\rho_i}\mathbf{H}_{ii}^{(ab)}\mathbf{c}_i^{(b)} + \sqrt{\eta_{ii}}\mathbf{H}_{ii}^{(aa)}\mathbf{c}_i^{(a)} + \mathbf{e}_i^{(a)} + \mathbf{n}_i^{(a)} \\ + \sum_{j\neq i}^{K}\sum_{c=1}^{2}\sqrt{\eta_{ij}}\mathbf{H}_{ij}^{(ac)}\left(\mathbf{x}_j^{(c)} + \mathbf{c}_j^{(c)}\right).\end{aligned} \tag{10}$$

Using (5)–(8), similar to [27], $\Sigma_i^{(a)}$, the covariance matrix of $\mathbf{v}_i^{(a)}$, is approximated, under $\kappa \ll 1$ and $\beta \ll 1$, as in (11). (See equation at the bottom of the page) Note that despite the fact $\Sigma_i^{(a)}$ at link i and node a depends on non-local parameters such as channel matrices and pre-coding matrices at other links, this covariance matrix can be determined locally provided that there is a sufficient coherence time window within which all channel matrices and pre-coding matrices do not change. An example of such an algorithm is shown later.

As a result of the limited DR in (10), the noise $\mathbf{v}_i^{(a)}$ is generally non-Gaussian. To the best of our knowledge, the exact capacity of MIMO channels with a non-Gaussian noise is still an open problem even for point-to-point MIMO systems [55], [56]. However, it is known that the Gaussian distribution is the worst case from a mutual-information perspective among all noise distributions [56], and thus assuming $\mathbf{v}_i^{(a)}$ as Gaussian, gives us the lower bound of mutual information [56]. The lower bound of the achievable rate of the node $i^{(a)}$, under Gaussian signaling, can be written as

$$I_i^{(a)} = \log_2\left|\mathbf{I}_{M_i} + \rho_i\mathbf{H}_{ii}^{(ab)}\mathbf{V}_i^{(b)}\left(\mathbf{V}_i^{(b)}\right)^H\left(\mathbf{H}_{ii}^{(ab)}\right)^H\left(\Sigma_i^{(a)}\right)^{-1}\right|. \tag{12}$$

WSR optimization scheme is formulated as follows:

$$\max_{\mathbf{V}_i^{(b)}} \quad \sum_{i=1}^{K}\sum_{a=1}^{2}\mu_i^{(a)} I_i^{(a)} \tag{13}$$

$$\text{s.t.} \quad \text{tr}\left\{\mathbf{V}_i^{(b)}\left(\mathbf{V}_i^{(b)}\right)^H\right\} \leq P_i^{(b)}, \ \forall(i,b) \tag{14}$$

$$\text{or s.t.} \quad \sum_{i=1}^{K}\sum_{b=1}^{2}\text{tr}\left\{\mathbf{V}_i^{(b)}\left(\mathbf{V}_i^{(b)}\right)^H\right\} \leq P_T, \tag{15}$$

where $P_i^{(b)}$ is the transmit power constraint at the node $i^{(b)}$, P_T is the sum-power constraint of the system, and $\mu_i^{(a)} \geq 0$ denotes the weight.

III. Weighted Minimum Mean-Squared-Error Minimization

To understand the link between the WSR maximization and the WMMSE minimization problems in the FD bi-directional MIMO channels, we need to establish the relationship between the achievable rate and the error covariance matrix. This argument is parallel to the one given in [43] for the MIMO broadcast channel and in [45] for the MIMO interference channel.

We assume that node $i^{(a)}$ applies the linear receiver $\mathbf{R}_i^{(a)} \in \mathbb{C}^{d_i \times M_i}$ to estimate the signal transmitted from node $i^{(b)}$. That is

$$\begin{aligned}\hat{\mathbf{d}}_i^{(b)} &= \mathbf{R}_i^{(a)}\tilde{\mathbf{y}}_i^{(a)} \\ &= \sqrt{\rho_i}\mathbf{R}_i^{(a)}\mathbf{H}_{ii}^{(ab)}\mathbf{V}_i^{(b)}\mathbf{d}_i^{(b)} + \mathbf{R}_i^{(a)}\mathbf{v}_i^{(a)}.\end{aligned} \tag{16}$$

We can now formulate the MSE of the node $i^{(a)}$. Using (16), the MSE matrix of the node $i^{(a)}$ can be written as

$$\begin{aligned}&\mathbf{MSE}_i^{(a)} \\ &= \mathbb{E}\left\{\left(\hat{\mathbf{d}}_i^{(b)} - \mathbf{d}_i^{(b)}\right)\left(\hat{\mathbf{d}}_i^{(b)} - \mathbf{d}_i^{(b)}\right)^H\right\} \\ &= \left(\sqrt{\rho_i}\mathbf{R}_i^{(a)}\mathbf{H}_{ii}^{(ab)}\mathbf{V}_i^{(b)} - \mathbf{I}_{d_i}\right)\left(\sqrt{\rho_i}\mathbf{R}_i^{(a)}\mathbf{H}_{ii}^{(ab)}\mathbf{V}_i^{(b)} - \mathbf{I}_{d_i}\right)^H \\ &\quad + \mathbf{R}_i^{(a)}\Sigma_i^{(a)}\left(\mathbf{R}_i^{(a)}\right)^H.\end{aligned} \tag{17}$$

$$\begin{aligned}\Sigma_i^{(a)} \approx\ & \rho_i\kappa\mathbf{H}_{ii}^{(ab)}\text{diag}\left(\mathbf{V}_i^{(b)}\left(\mathbf{V}_i^{(b)}\right)^H\right)\left(\mathbf{H}_{ii}^{(ab)}\right)^H + \eta_{ii}\kappa\mathbf{H}_{ii}^{(aa)}\text{diag}\left(\mathbf{V}_i^{(a)}\left(\mathbf{V}_i^{(a)}\right)^H\right)\left(\mathbf{H}_{ii}^{(aa)}\right)^H \\ & + \beta\rho_i\text{diag}\left(\mathbf{H}_{ii}^{(ab)}\mathbf{V}_i^{(b)}\left(\mathbf{V}_i^{(b)}\right)^H\left(\mathbf{H}_{ii}^{(ab)}\right)^H\right) + \beta\eta_{ii}\text{diag}\left(\mathbf{H}_{ii}^{(aa)}\mathbf{V}_i^{(a)}\left(\mathbf{V}_i^{(a)}\right)^H\left(\mathbf{H}_{ii}^{(aa)}\right)^H\right) \\ & + \sum_{j\neq i}^{K}\sum_{c=1}^{2}\eta_{ij}\left[\mathbf{H}_{ij}^{(ac)}\left(\mathbf{V}_j^{(c)}\left(\mathbf{V}_j^{(c)}\right)^H + \kappa\text{diag}\left(\mathbf{V}_j^{(c)}\left(\mathbf{V}_j^{(c)}\right)^H\right)\right)\left(\mathbf{H}_{ij}^{(ac)}\right)^H\right] \\ & + \sum_{j\neq i}^{K}\sum_{c=1}^{2}\beta\eta_{ij}\text{diag}\left(\mathbf{H}_{ij}^{(ac)}\mathbf{V}_j^{(c)}\left(\mathbf{V}_j^{(c)}\right)^H\left(\mathbf{H}_{ij}^{(ac)}\right)^H\right) + \mathbf{I}_{M_i}.\end{aligned} \tag{11}$$

The minimum mean-squared error (MMSE) receiver filter applied at the node $i^{(a)}$ can be expressed as

$$\begin{aligned}\mathbf{R}_i^{(a)*} &= \arg\min_{\mathbf{R}_i^{(a)}} \operatorname{tr}\left\{\mathbf{MSE}_i^{(a)}\right\}\\ &= \sqrt{\rho_i}\left(\mathbf{V}_i^{(b)}\right)^H\left(\mathbf{H}_{ii}^{(ab)}\right)^H\\ &\quad\times\left(\rho_i\mathbf{H}_{ii}^{(ab)}\mathbf{V}_i^{(b)}\left(\mathbf{V}_i^{(b)}\right)^H\left(\mathbf{H}_{ii}^{(ab)}\right)^H+\Sigma_i^{(a)}\right)^{-1}. \quad (18)\end{aligned}$$

Plugging (18) in (17) gives us the MSE matrix for the node $i^{(a)}$ given that the MMSE receive filter is applied, and it can be written as

$$\begin{aligned}\mathbf{E}_i^{(a)} &= \mathbf{I}_{d_i}-\rho_i\left(\mathbf{V}_i^{(b)}\right)^H\left(\mathbf{H}_{ii}^{(ab)}\right)^H\\ &\quad\times\left(\rho_i\mathbf{H}_{ii}^{(ab)}\mathbf{V}_i^{(b)}\left(\mathbf{V}_i^{(b)}\right)^H\left(\mathbf{H}_{ii}^{(ab)}\right)^H+\Sigma_i^{(a)}\right)^{-1}\\ &= \left(\mathbf{I}_{d_i}+\rho_i\left(\mathbf{V}_i^{(b)}\right)^H\left(\mathbf{H}_{ii}^{(ab)}\right)^H\left(\Sigma_i^{(a)}\right)^{-1}\mathbf{H}_{ii}^{(ab)}\mathbf{V}_i^{(b)}\right)^{-1}, \quad (19)\end{aligned}$$

where the matrix inversion lemma $(\mathbf{A}+\mathbf{BCD})^{-1}=\mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{B}(\mathbf{DA}^{-1}\mathbf{B}+\mathbf{C}^{-1})^{-1}\mathbf{DA}^{-1}$ [57] is applied in the second equality. We refer to $\mathbf{E}_i^{(a)}$ as the MMSE-matrix.

Comparing (12) and (19), and using $\log_2|\mathbf{I}_N+\mathbf{AB}|=\log_2|\mathbf{I}_M+\mathbf{BA}|$ identity [57], where $\mathbf{A}\in\mathbb{C}^{N\times M}$, and $\mathbf{B}\in\mathbb{C}^{M\times N}$, it is easy to see the relationship between the achievable rate and the MMSE-matrix as

$$I_i^{(a)}=\log_2\left|\left(\mathbf{E}_i^{(a)}\right)^{-1}\right|. \quad (20)$$

Now, we first formulate WMMSE optimization problem, and then show that there is a simple relationship between the Karush-Kuhn-Tucker (KKT) conditions of the WSR and WMMSE problems. Assuming that MMSE receive filtering is applied, WMMSE problem can be formulated as

$$\min_{\mathbf{V}_i^{(b)}} \sum_{i=1}^{K}\sum_{a=1}^{2}\operatorname{tr}\left\{\mathbf{W}_i^{(a)}\mathbf{E}_i^{(a)}\right\} \quad (21)$$

$$\text{s.t.} \quad \operatorname{tr}\left\{\mathbf{V}_i^{(b)}\left(\mathbf{V}_i^{(b)}\right)^H\right\}\le P_i^{(b)},\ \forall(i,b) \quad (22)$$

$$\text{or s.t.} \quad \sum_{i=1}^{K}\sum_{b=1}^{2}\operatorname{tr}\left\{\mathbf{V}_i^{(b)}\left(\mathbf{V}_i^{(b)}\right)^H\right\}\le P_T, \quad (23)$$

where $\mathbf{W}_i^{(a)}\in\mathbb{C}^{d_i\times d_i}$ is a constant weight matrix associated with node $i^{(a)}$.

As shown in Appendix A, the gradient of WSR and the gradient of WMMSE problems are equal if the MSE-weights $\mathbf{W}_i^{(a)}$ are chosen as:

$$\mathbf{W}_i^{(a)}=\frac{\mu_i^{(a)}}{\ln 2}\left(\mathbf{E}_i^{(a)}\right)^{-1}. \quad (24)$$

Since the KKT conditions of the WSR and WMMSE problems can be satisfied simultaneously with the choice of MSE-weights (24), we can solve the WSR problem (13)–(15) through solving WMMSE problem (21)–(23).

A. Sum-Power Constrained Transceiver Design

The problem to find the optimal transmit filters $\mathbf{V}_i^{(b)}$ for fixed receive filters under the sum-power constraint of the system is formulated as below:

$$\begin{aligned}\min_{\mathbf{V}_i^{(b)}} &\sum_{i=1}^{K}\sum_{a=1}^{2}\operatorname{tr}\left\{\mathbf{W}_i^{(a)}\mathbb{E}\left\{\left(\mathbf{d}_i^{(b)}-\alpha^{-1}\hat{\mathbf{d}}_i^{(b)}\right)\right.\right.\\ &\quad\left.\left.\times\left(\mathbf{d}_i^{(b)}-\alpha^{-1}\hat{\mathbf{d}}_i^{(b)}\right)^H\right\}\right\} \quad (25)\end{aligned}$$

$$\text{s.t.} \quad \sum_{i=1}^{K}\sum_{b=1}^{2}\operatorname{tr}\left\{\mathbf{V}_i^{(b)}\left(\mathbf{V}_i^{(b)}\right)^H\right\}\le P_T, \quad (26)$$

where $\mathbf{W}_i^{(a)}$ is chosen according to (24) and α is a scaling parameter. Similar to [58], where the optimal transmit filters are computed for the unweighted case, the WMMSE transmit filter of (25), (26) can be shown to be

$$\mathbf{V}_i^{(b)}=\alpha\bar{\mathbf{V}}_i^{(b)}. \quad (27)$$

Here, α is defined as

$$\alpha=\sqrt{\frac{P_T}{\sum_{i=1}^{K}\sum_{b=1}^{2}\operatorname{tr}\left\{\bar{\mathbf{V}}_i^{(b)}\left(\bar{\mathbf{V}}_i^{(b)}\right)^H\right\}}} \quad (28)$$

and as shown in Appendix B, $\bar{\mathbf{V}}_i^{(b)}$ is computed as

$$\begin{aligned}\bar{\mathbf{V}}_i^{(b)} &= \sqrt{\rho_i}\left(\mathbf{X}_i^{(b)}+\frac{\sum_{i=1}^{K}\sum_{a=1}^{2}\operatorname{tr}\left\{\mathbf{A}_i^{(a)}\right\}}{P_T}\mathbf{I}_{N_i}\right)^{-1}\\ &\quad\times\left(\mathbf{H}_{ii}^{(ab)}\right)^H\left(\mathbf{R}_i^{(a)}\right)^H\mathbf{W}_i^{(a)}, \quad (29)\end{aligned}$$

$$\begin{aligned}\mathbf{X}_i^{(b)} &= \rho_i\left(\mathbf{H}_{ii}^{(ab)}\right)^H\mathbf{A}_i^{(a)}\mathbf{H}_{ii}^{(ab)}+\rho_i\kappa\operatorname{diag}\left(\left(\mathbf{H}_{ii}^{(ab)}\right)^H\mathbf{A}_i^{(a)}\mathbf{H}_{ii}^{(ab)}\right)+\rho_i\beta\left(\mathbf{H}_{ii}^{(ab)}\right)^H\operatorname{diag}\left(\mathbf{A}_i^{(a)}\right)\mathbf{H}_{ii}^{(ab)}\\ &\quad+\eta_{ii}\kappa\operatorname{diag}\left(\left(\mathbf{H}_{ii}^{(bb)}\right)^H\mathbf{A}_i^{(b)}\mathbf{H}_{ii}^{(bb)}\right)+\eta_{ii}\beta\left(\mathbf{H}_{ii}^{(bb)}\right)^H\operatorname{diag}\left(\mathbf{A}_i^{(b)}\right)\mathbf{H}_{ii}^{(bb)}\\ &\quad+\sum_{j\neq i}^{K}\sum_{c=1}^{2}\left(\eta_{ji}\left(\mathbf{H}_{ji}^{(cb)}\right)^H\mathbf{A}_j^{(c)}\mathbf{H}_{ji}^{(cb)}+\eta_{ji}\kappa\operatorname{diag}\left(\left(\mathbf{H}_{ji}^{(cb)}\right)^H\mathbf{A}_j^{(c)}\mathbf{H}_{ji}^{(cb)}\right)+\eta_{ji}\beta\left(\mathbf{H}_{ji}^{(cb)}\right)^H\operatorname{diag}\left(\mathbf{A}_j^{(c)}\right)\mathbf{H}_{ji}^{(cb)}\right). \quad (30)\end{aligned}$$

where $\mathbf{X}_i^{(b)}$ is shown in (30), (See equation at the bottom of the previous page) and $\mathbf{A}_i^{(a)} = (\mathbf{R}_i^{(a)})^H \mathbf{W}_i^{(a)} \mathbf{R}_i^{(a)}, i = 1, \ldots, K, a = 1, 2$.

B. Individual-Power-Constrained Transceiver Design

The problem to find the optimal transmit filters $\mathbf{V}_i^{(b)}$ for fixed receive filters under the individual-power constraint at each node of the system is formulated as below:

$$\min_{\mathbf{V}_i^{(b)}} \sum_{i=1}^{K} \sum_{a=1}^{2} \operatorname{tr}\left\{\mathbf{W}_i^{(a)} \mathbb{E}\left\{\left(\mathbf{d}_i^{(b)} - \hat{\mathbf{d}}_i^{(b)}\right) \times \left(\mathbf{d}_i^{(b)} - \hat{\mathbf{d}}_i^{(b)}\right)^H\right\}\right\} \tag{31}$$

$$\text{s.t.} \quad \operatorname{tr}\left\{\mathbf{V}_i^{(b)} \left(\mathbf{V}_i^{(b)}\right)^H\right\} \le P_i^{(b)}, \ \forall (i, b). \tag{32}$$

Taking the partial derivative of the Lagrange function of (31), (32) with respect to the matrix $\mathbf{V}_i^{(b)}$, we can obtain the optimal $\mathbf{V}_i^{(b)}$ as

$$\mathbf{V}_i^{(b)} = \sqrt{\rho_i}\left(\lambda_i^{(b)} \mathbf{I}_{N_i} + \mathbf{X}_i^{(b)}\right)^{-1} \left(\mathbf{R}_i^{(a)} \mathbf{H}_{ii}^{(ab)}\right)^H \mathbf{W}_i^{(a)} \tag{33}$$

where $\mathbf{X}_i^{(b)}$ is defined in (30).

The values of the Lagrange multiplier $\lambda_i^{(b)}$ in (33) are calculated by taking the singular value decomposition of $\mathbf{X}_i^{(b)} = \mathbf{U}_i^{(b)} \Delta_i^{(b)} (\mathbf{U}_i^{(b)})^H$ and writing the power constraint in (32), after simple steps, as

$$\operatorname{tr}\left\{\mathbf{V}_i^{(b)} \left(\mathbf{V}_i^{(b)}\right)^H\right\} = \rho_i \sum_{k=1}^{N} \frac{g_{ik}^{(b)}}{\left(\lambda_i^{(b)} + \Delta_{ik}^{(b)}\right)^2} = P_i^{(b)}, \tag{34}$$

where $g_{ik}^{(b)}$ denotes the kth row and kth column element of $(\mathbf{U}_i^{(b)})^H (\mathbf{H}_{ii}^{(ab)})^H (\mathbf{R}_i^{(a)})^H \mathbf{W}_i^{(a)} (\mathbf{W}_i^{(a)})^H \mathbf{R}_i^{(a)} \mathbf{H}_{ii}^{(ab)} \mathbf{U}_i^{(b)}$ and $\Delta_{ik}^{(b)}$ denotes the kth row and kth column element of the matrix $\Delta_i^{(b)}$. We can compute $\lambda_i^{(b)}$ from (34) numerically. If the values of the Lagrange multipliers $\lambda_i^{(b)}$ are negative, we assign $\lambda_i^{(b)}$ as zeros.

C. Convergence

The iterative alternating algorithm for solving the WSR optimization problem (13)–(15) through WMMSE minimization problem is given in Table I. The algorithm in Table I holds for both the sum-power constraint and the individual-power-constraint WSR problems. The proof of the convergence of the algorithm is based on the proof of a more general equivalent optimization problem, which includes the MSE weights and receive filters as new optimization variables in addition to transmit filters. The objective function of the new optimization problem is formulated as

TABLE I
WSR MAXIMIZATION ALGORITHM

1) Set the iteration number $n = 0$ and initialize the transmit filters $\mathbf{V}_i^{(b),[0]}$, $i \in \{1, \ldots, K\}$, $b \in \{1, 2\}$.
2) $n \leftarrow n + 1$. Update the receive filter $\mathbf{R}_i^{(a),[n]}$ using (18).
3) Calculate and update the weighting matrix $\mathbf{W}_i^{(a),[n]}$ using (24):
4) Calculate and update the transmit filter $\mathbf{V}_i^{(b),[n]}$ using (27) for the sum-power constraint problem and using (33) for the individual power constraint problem.
5) Repeat steps 2, 3 and 4 until convergence, or a predefined number of iterations is reached.

$$\min_{\mathbf{R}_i^{(a)}, \mathbf{V}_i^{(b)}, \mathbf{W}_i^{(a)}} \sum_{i=1}^{K} \sum_{a=1}^{2} \operatorname{tr}\left\{\mathbf{W}_i^{(a)} \mathbf{MSE}_i^{(a)}\right\} - \mu_i^{(a)} \log_2 \left|\frac{\ln 2}{\mu_i^{(a)}} \mathbf{W}_i^{(a)}\right| - d_i \frac{\mu_i^{(a)}}{\ln 2}. \tag{35}$$

When the weight matrices $\mathbf{W}_i^{(a)}$, and the transmit filters $\mathbf{V}_i^{(b)}$ are fixed, the optimal receiving filter $\mathbf{R}_i^{(a)}$ is MMSE receving filter given in (18). Substituting the optimal receiving filter $\mathbf{R}_i^{(a)}$ into $\mathbf{MSE}_i^{(a)}$ in the objective function (35) gives a new cost function for the weights and the transmit filters given as

$$\min_{\mathbf{V}_i^{(b)}, \mathbf{W}_i^{(a)}} \sum_{i=1}^{K} \sum_{a=1}^{2} \operatorname{tr}\left\{\mathbf{W}_i^{(a)} \mathbf{E}_i^{(a)}\right\} - \mu_i^{(a)} \log_2 \left|\frac{\ln 2}{\mu_i^{(a)}} \mathbf{W}_i^{(a)}\right| - d_i \frac{\mu_i^{(a)}}{\ln 2}. \tag{36}$$

Optimizing the cost function (36) with respect to weight matrices, $\mathbf{W}_i^{(a)}$ results in $\mathbf{W}_i^{(a)} = \frac{\mu_i^{(a)}}{\ln 2} (\mathbf{E}_i^{(a)})^{-1}$. Plugging in the optimal weight matrix in the cost function (36), we obtain a new cost function for the transmit filters expressed as

$$\min_{\mathbf{V}_i^{(b)}} \sum_{i=1}^{K} \sum_{a=1}^{2} -\mu_i^{(a)} \log_2 \left|\left(\mathbf{E}_i^{(a)}\right)^{-1}\right|, \tag{37}$$

which is exactly same as the cost function of the original WSR problem (13). Therefore, the new optimization problem reduces to the original optimization problem when minimized with respect to the transmit filters $\mathbf{V}_i^{(b)}$ under fixed receiving filters and weight matrices.

The alternating minimization algorithm decreases the cost function (35) monotonically at each step of the iteration. And since the cost function (35) is lower bounded under a fixed individual power constraint, it is concluded that the algorithm in Table I is guaranteed to converge to a local optimum.

D. Remarks

1) Since the WMMSE minimization (21)–(23) problem is not jointly convex over optimization variables, the proposed algorithm does not ensure to converge to the global optimal solution.

Because of the non-convexity of the optimization problems we are dealing with, we need to choose good initialization points to have a suboptimal solution with a good performance. In [59], several reasonable choices such as right singular matrices, random matrices and interference alignment (IA) initialization have been proposed. Optimal initialization method that guarantees global optimal solution is a very challenging problem and will be left to our future work.

2) Note that WSR optimization problem (13)–(15) can also be solved using gradient-based (GB) search [60], but the iterative scheme based on the WMMSE applied in this paper requires considerably less computational complexity and a smaller amount of feedback resources [45]. Also, our scheme requires only the *local* CSI (i.e., each transmitter needs to know only the CSI of the links originating from itself), whereas GB search method requires the CSI for all links (global CSI). Particularly, in our scheme, a node does not need to know the self-interference channel of the other nodes. It is challenging in practice, to acquire the strong self-interference channel of the other nodes accurately since its large DR requires heavy feedback information between the two nodes [36]. And it will be shown in the simulation results that the sum-rate achieved by the proposed algorithm is very close to that of the GB algorithm.

Now we will elaborate on the CSI assumption considered in this paper. Although there exists various definitions of *local* CSI, in this paper, with *local* CSI, we assume that the transmitters are able to obtain the knowledge of channel coefficients which are directly connected to them. The same definition has also been adopted and commonly used in other papers [48]–[54]. In particular, we assume each transmitter has perfect knowledge of the channel matrices only between itself and all receivers. This information can be obtained easily by overhearing signaling packets at the MAC layer. For example, in the IEEE 802.11n scheme, assuming the channel reciprocity, a transmitter can estimate the channel between itself and the unintended receiver by capturing the "Clear-to-Send" message, which contains a training sequence from an unintended receiver [53], [61]. Since only *local* CSI is required at each user, the proposed method fits for distributed implementation. Similar to [48], [49], [53], [61], we assume that each receiver $i^{(a)}$ locally estimates the received interference-plus-noise covariance matrix $\Sigma_i^{(a)}$ through measurements, and directly uses it in the calculations.[1] Using the estimate of $\Sigma_i^{(a)}$, the receiver $i^{(a)}$ updates the matrices $\mathbf{R}_i^{(a)}$ and $\mathbf{W}_i^{(a)}$. Then, it feeds back the updated $\mathbf{R}_i^{(a)}$ and $\mathbf{W}_i^{(a)}$ to the transmitters via wireless broadcast [49].

Note that in a centralized scheme, a centralized scheduler first collects all MIMO channels of all links by employing multihop routing to receive the CSI feedback from the nodes located far from the centralized scheduler, and then calibrates, computes and distributes the optimum filtering matrices of all links. The complexity of the centralized algorithm increases substantially as the number of links increases and it comes at the cost of signaling overhead. On the contrary, the proposed distributed scheme requires each link to collect only *local* CSI, and thus possesses improved scalability and less complexity.

Assuming the same number of transmit antennas, receive antennas, and data streams at each node, i.e., $N_i = N$, $M_i = M$, and $d_i = d$, the required feedback information to compute the transmit filtering matrix $\mathbf{V}_i^{(b)}$ at the node $i^{(b)}$ is computed as follows. The centralized scheme requires the CSI of all links, which has a matrix size of $4MNK^2$, and it requires transmit filtering coefficients, $\mathbf{V}_j^{(a)}$, of all nodes, except node $i^{(b)}$, which has a matrix size of $Nd(2K-1)$, giving a total of $4MNK^2 + Nd(2K-1)I_{iter}$ feedback requirement, where I_{iter} is the average number of iterations for convergence. In our simulations, I_{iter} is usually around 10. On the other hand, the proposed method for the individual power constrained problem requires only local CSI, which has a matrix dimension of $2MNK$, and requires coefficients of $\mathbf{R}_j^{(a)}$ and $\mathbf{W}_j^{(a)}$ from all nodes, which has a matrix dimension of $2K(Md+d^2)$, resulting in a total of $2MNK + 2K(Md+d^2)I_{iter}$ feedback requirement.[2] As the number of pairs, K, increases, the feedback requirement for the centralized and proposed algorithm increases quadratically and linearly, respectively. Therefore, the proposed method has less feedback requirement for large networks.

IV. Extension to Full-Duplex Cellular Systems

In this section, we show that the algorithm proposed for the FD MIMO interference channel also holds for FD cellular systems, in which a FD BS communicates with HD mode UL and DL users, simultaneously as seen in Fig. 2.[3] The BS serves K UL users and J DL users simultaneously. The BS is equipped with M_0 and N_0 transmit and receive antennas, respectively. The number of antennas of the k-th UL user and the j-th DL user are denoted by M_k and N_j, respectively. The number of data streams transmitted from the k-th UL user (to the j-th DL user) is denoted by $d_k^{UL}(d_j^{DL})$.

$\mathbf{H}_k^{UL} \in \mathbb{C}^{N_0 \times M_k}$ and $\mathbf{H}_j^{DL} \in \mathbb{C}^{N_j \times M_0}$ represent the k-th UL channel and the j-th DL channel, respectively. $\mathbf{H}_0 \in \mathbb{C}^{N_0 \times M_0}$ is the self-interference channel from the transmitter antennas of BS to the receiver antennas of BS. $\mathbf{H}_{jk}^{DU} \in \mathbb{C}^{N_j \times M_k}$ denotes the CCI channel from the k-th UL user to the j-th DL user.

[1] Using a very small fraction of channel coherence time, all links can estimate their desired local channel matrices (the channel between their transmitter and receiver) sequentially. With the desired local channel matrices, and given set of the precoding matrices at all links, all links first undergo a training phase to estimate $\Sigma_i^{(a)}$. In this phase, all links transmit synchronously/simultaneously training sequences with their given precoding matrices, and the node a of link i uses (9) to measure $\mathbf{v}_i^{(a)}$, and hence estimates $\Sigma_i^{(a)}$. With a sufficiently long training time, the error in estimated $\Sigma_i^{(a)}$ can be negligible. With a given set of $\Sigma_i^{(a)}$ at all links, all links then follow the principle shown in this paper to update their precoding matrices locally. After each update of the precoding matrices, all links undergo the training phase again to estimate their $\Sigma_i^{(a)}$. The (local) estimation of $\Sigma_i^{(a)}$ at all links is indeed an overhead. But this is entirely feasible in theory. In the simulation part of this paper, we assume the exact knowledge of $\Sigma_i^{(a)}$ as is justified.

[2] Note that in sum-power constrained problem, the total feedback requirement is $2MNK + (2K(Md+d^2)+1)I_{iter}$, since it also requires the feedback of $\sum_{i=1}^{K}\sum_{b=1}^{2}\mathrm{tr}\{\bar{\mathbf{V}}_i^{(b)}(\bar{\mathbf{V}}_i^{(b)})^H\}$.

[3] In this section, we will use the same notations as in [26].

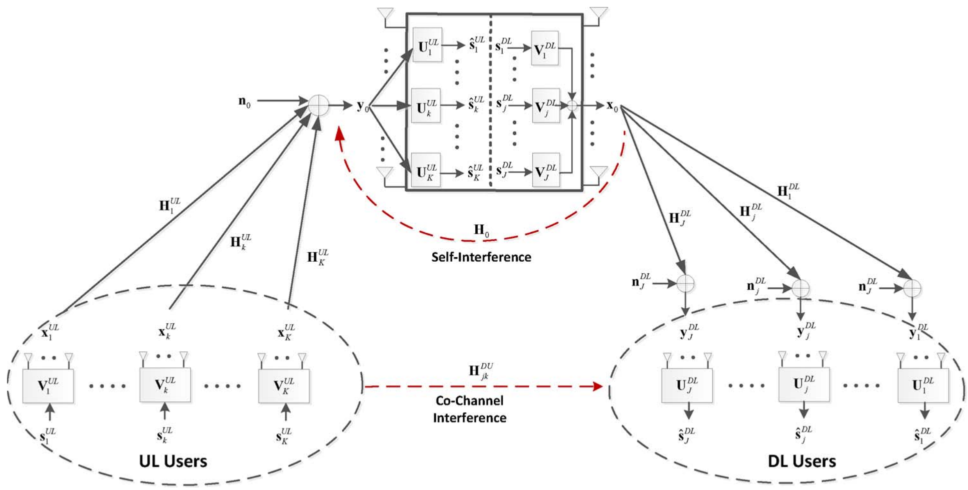

Fig. 2. Full-duplex multi-user MIMO system model.

The vector of source symbols transmitted by the k-th UL user is denoted as $\mathbf{s}_k^{UL} = [s_{k,1}^{UL}, \ldots, s_{k,d_k^{UL}}^{UL}]^T$. It is assumed that the symbols are i.i.d. with unit power, i.e., $\mathbb{E}[\mathbf{s}_k^{UL}(\mathbf{s}_k^{UL})^H] = \mathbf{I}_{d_k^{UL}}$. Similarly, the transmit symbols for the j-th DL user is denoted by $\mathbf{s}_j^{DL} = [s_{j,1}^{DL}, \ldots, s_{j,d_j^{DL}}^{DL}]^T$, with $\mathbb{E}[\mathbf{s}_j^{DL}(\mathbf{s}_j^{DL})^H] = \mathbf{I}_{d_j^{DL}}$. Denoting the precoders for the data streams of k-th UL and j-th DL user as $\mathbf{V}_k^{UL} = [\mathbf{v}_{k,1}^{UL}, \ldots, \mathbf{v}_{k,d_k^{UL}}^{UL}] \in \mathbb{C}^{M_k \times d_k^{UL}}$, and $\mathbf{V}_j^{DL} = [\mathbf{v}_{j,1}^{DL}, \ldots, \mathbf{v}_{j,d_j^{DL}}^{DL}] \in \mathbb{C}^{M_0 \times d_j^{DL}}$, respectively, the transmitted signal of the k-th UL user and that of the BS can be written, respectively, as

$$\mathbf{x}_k^{UL} = \mathbf{V}_k^{UL}\mathbf{s}_k^{UL}, \tag{38}$$

$$\mathbf{x}_0 = \sum_{j=1}^{J} \mathbf{V}_j^{DL}\mathbf{s}_j^{DL}. \tag{39}$$

We consider a FD multi-user MIMO system that suffers from self-interference and CCI. The signal received by the BS and that received by the j-th DL user can be written, respectively, as

$$\begin{aligned}\mathbf{y}_0 = &\sum_{k=1}^{K} \mathbf{H}_k^{UL}\left(\mathbf{x}_k^{UL} + \mathbf{c}_k^{UL}\right) + \mathbf{H}_0\left(\mathbf{x}_0 + \mathbf{c}_0\right) \\ &+ \mathbf{e}_0 + \mathbf{n}_0,\end{aligned} \tag{40}$$

$$\begin{aligned}\mathbf{y}_j^{DL} = &\mathbf{H}_j^{DL}\left(\mathbf{x}_0 + \mathbf{c}_0\right) + \sum_{k=1}^{K} \mathbf{H}_{jk}^{DU}\left(\mathbf{x}_k^{UL} + \mathbf{c}_k^{UL}\right) \\ &+ \mathbf{e}_j^{DL} + \mathbf{n}_j^{DL},\end{aligned} \tag{41}$$

where $\mathbf{n}_0 \in \mathbb{C}^{N_0}$ and $\mathbf{n}_j^{DL} \in \mathbb{C}^{N_j}$ denote the AWGN vector with zero mean and unit covariance matrix at the the BS and the j-th DL user, respectively. In (40) and (41), $\mathbf{c}_k^{UL}(\mathbf{c}_0)$ is the transmitter distortion at the k-th UL user (BS), which is modeled as in (5), (6), and $\mathbf{e}_j^{DL}(\mathbf{e}_0)$ is the receiver distortion at the j-th DL user (BS), which is modeled as in (7), (8).

From (40), (41), the aggregate interference-plus-noise terms at the k-th UL and the j-th DL user are written, respectively as

$$\begin{aligned}\mathbf{m}_k^{UL} = &\sum_{j=1, j\neq k}^{K} \mathbf{H}_j^{UL}\mathbf{x}_j^{UL} + \sum_{j=1}^{K} \mathbf{H}_j^{UL}\mathbf{c}_j^{UL} + \mathbf{H}_0\left(\mathbf{x}_0 + \mathbf{c}_0\right) \\ &+ \mathbf{e}_0 + \mathbf{n}_0, k = 1, \ldots, K,\end{aligned}$$

$$\begin{aligned}\mathbf{m}_j^{DL} = &\mathbf{H}_j^{DL}\sum_{k=1, k\neq j}^{J} \mathbf{V}_k^{DL}\mathbf{s}_k^{DL} + \sum_{k=1}^{K} \mathbf{H}_{jk}^{DU}\left(\mathbf{x}_k^{UL} + \mathbf{c}_k^{UL}\right) \\ &+ \mathbf{H}_j^{DL}\mathbf{c}_0 + \mathbf{e}_j^{DL} + \mathbf{n}_j^{DL}, j = 1, \ldots, J.\end{aligned}$$

Similar to [27], the covariance matrix of $\mathbf{m}_k^{UL}$, i.e., Σ_k^{UL} can be approximated, under $\beta \ll 1$ and $\kappa \ll 1$, as in (42). (See equation at the bottom of the page) The covariance matrix of $\mathbf{m}_j^{DL}$, i.e., Σ_j^{DL} can be defined similarly, i.e., by replacing $\mathbf{H}_j^{UL}$, $\mathbf{V}_j^{UL}$, and $\mathbf{H}_0$ in (11) with $\mathbf{H}_j^{DL}$, $\mathbf{V}_k^{DL}$, and $\mathbf{H}_{jk}^{DU}$, respectively.

The received signals are processed by linear decoders, denoted as $\mathbf{U}_k^{UL} = [\mathbf{u}_{k,1}^{UL}, \ldots, \mathbf{u}_{k,d_k^{UL}}^{UL}] \in \mathbb{C}^{N_0 \times d_k^{UL}}$, and $\mathbf{U}_j^{DL} = [\mathbf{u}_{j,1}^{DL}, \ldots, \mathbf{u}_{j,d_j^{DL}}^{DL}] \in \mathbb{C}^{N_j \times d_j^{DL}}$ by the BS and the j-th DL user,

$$\begin{aligned}\Sigma_k^{UL} = &\sum_{j\neq k}^{K} \mathbf{H}_j^{UL}\mathbf{V}_j^{UL}\left(\mathbf{V}_j^{UL}\right)^H\left(\mathbf{H}_j^{UL}\right)^H + \kappa\sum_{j=1}^{K}\mathbf{H}_j^{UL}\mathrm{diag}\left(\mathbf{V}_j^{UL}\left(\mathbf{V}_j^{UL}\right)^H\right)\left(\mathbf{H}_j^{UL}\right)^H \\ &+ \sum_{j=1}^{J}\mathbf{H}_0\left(\mathbf{V}_j^{DL}\left(\mathbf{V}_j^{DL}\right)^H + \kappa\mathrm{diag}\left(\mathbf{V}_j^{DL}\left(\mathbf{V}_j^{DL}\right)^H\right)\right)\mathbf{H}_0^H + \beta\sum_{j=1}^{K}\mathrm{diag}\left(\mathbf{H}_j^{UL}\mathbf{V}_j^{UL}\left(\mathbf{V}_j^{UL}\right)^H\left(\mathbf{H}_j^{UL}\right)^H\right) \\ &+ \beta\sum_{j=1}^{J}\mathrm{diag}\left(\mathbf{H}_0\mathbf{V}_j^{DL}\left(\mathbf{V}_j^{DL}\right)^H\mathbf{H}_0^H\right) + \mathbf{I}_{N_0}.\end{aligned} \tag{42}$$

respectively. Therefore, the estimate of data streams of the k-th UL user at the BS is given as $\hat{\mathbf{s}}_k^{UL} = (\mathbf{U}_k^{UL})^H \mathbf{y}_0$, and similarly, the estimate of the date stream of the j-th DL user is $\hat{\mathbf{s}}_j^{DL} = (\mathbf{U}_j^{DL})^H \mathbf{y}_j^{DL}$. Using these estimates, the signal-to-interference-plus-noise ratio (SINR) values of the m-th stream associated with k-th UL and j-th DL user can be, respectively, written as

$$\gamma_{k,m}^{UL} = \frac{\left|\left(\mathbf{u}_{k,m}^{UL}\right)^H \mathbf{H}_k^{UL}\mathbf{v}_{k,m}^{UL}\right|^2}{\left(\mathbf{u}_{k,m}^{UL}\right)^H \Sigma_k^{UL}\mathbf{u}_{k,m}^{UL} + \sum_{n\neq m}^{d_k^{UL}} \left|\left(\mathbf{u}_{k,m}^{UL}\right)^H \mathbf{H}_k^{UL}\mathbf{v}_{k,n}^{UL}\right|^2},$$

$$\gamma_{j,m}^{DL} = \frac{\left|\left(\mathbf{u}_{j,m}^{DL}\right)^H \mathbf{H}_j^{DL}\mathbf{v}_{j,m}^{DL}\right|^2}{\left(\mathbf{u}_{j,m}^{DL}\right)^H \Sigma_j^{DL}\mathbf{u}_{j,m}^{DL} + \sum_{n\neq m}^{d_j^{DL}} \left|(\mathbf{u}_{j,m}^{DL})^H \mathbf{H}_j^{DL}\mathbf{v}_{j,n}^{DL}\right|^2}.$$

A. Joint Beamforming Design

The optimization problem can be formulated as:

$$\max_{\substack{\mathbf{v}_{k,m}^{UL},\mathbf{u}_{k,m}^{UL}\\ \mathbf{v}_{j,m}^{DL},\mathbf{u}_{j,m}^{DL}}} \sum_{k=1}^{K} \mu_k^{UL} \sum_{m=1}^{d_k^{UL}} \log_2\left(1+\gamma_{k,m}^{UL}\right) + \sum_{j=1}^{J} \mu_j^{DL} \sum_{m=1}^{d_j^{DL}} \log_2\left(1+\gamma_{j,m}^{DL}\right) \quad (43)$$

$$\text{s.t.} \quad \sum_{m=1}^{d_k^{UL}} \left(\mathbf{v}_{k,m}^{UL}\right)^H \mathbf{v}_{k,m}^{UL} \leq P_k, k\in \mathcal{S}^{UL}, \quad (44)$$

$$\sum_{j=1}^{J}\sum_{m=1}^{d_j^{DL}} \left(\mathbf{v}_{j,m}^{DL}\right)^H \mathbf{v}_{j,m}^{DL} \leq P_0, \quad (45)$$

where P_k in (44) is the transmit power constraint at the k-th UL user, P_0 in (45) is the total power constraint at the BS, and $\mu_k^{UL}(\mu_j^{DL})$ is the weight at the kth (jth) UL (DL) user. We use $\mathcal{S}^{UL}$ and $\mathcal{S}^{DL}$ to represent the set of K UL and J DL channels, respectively.

Using Theorem 1 in [62], the optimization problem (43)–(45) can be expressed in terms of only linear precoders as

$$\max_{\mathbf{V}} \sum_{k=1}^{K} \mu_k^{UL} \log_2\left|\mathbf{I}_{N_0} + \mathbf{A}_k^{UL}\left(\mathbf{A}_k^{UL}\right)^H \left(\Sigma_k^{UL}\right)^{-1}\right| + \sum_{j=1}^{J} \mu_j^{DL} \log_2\left|\mathbf{I}_{N_j} + \mathbf{A}_j^{DL}\left(\mathbf{A}_j^{DL}\right)^H \left(\Sigma_j^{DL}\right)^{-1}\right| \quad (46)$$

$$\text{s.t.} \quad \mathrm{tr}\left\{\mathbf{V}_k^{UL}\left(\mathbf{V}_k^{UL}\right)^H\right\} \leq P_k, k\in \mathcal{S}^{UL}, \quad (47)$$

$$\sum_{j=1}^{J} \mathrm{tr}\left\{\mathbf{V}_j^{DL}\left(\mathbf{V}_j^{DL}\right)^H\right\} \leq P_0, \quad (48)$$

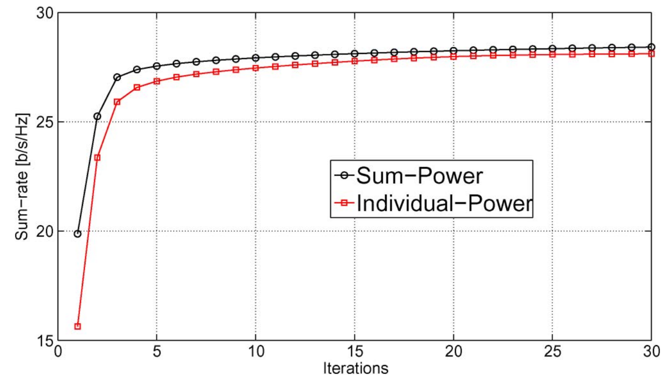

Fig. 3. Convergence behaviour of the proposed algorithm. Here $K = 3$, $N = 2$, SNR $= 20$ dB, INR $= 10$ dB, $\mathrm{INR_{SI}} = 40$ dB, $\mu = 1$, $\kappa = \beta = -40$ dB.

where $\mathbf{A}_k^{UL} = \mathbf{H}_k^{UL}\mathbf{V}_k^{UL}$, and $\mathbf{A}_j^{DL} = \mathbf{H}_j^{DL}\mathbf{V}_j^{DL}$, and $\mathbf{V}$ denotes a matrix obtained by stacking the precoding matrices of all UL and DL users. The optimization problem (46)–(48) has the same formulation as (13)–(15), and thus we can apply the individual-power constrained transceiver design proposed in Section III-B for uplink users, $i \in \mathcal{S}^{UL}$, and apply the sum-power constrained transceiver design proposed in Section III-A for downlink users, $i \in \mathcal{S}^{DL}$.

V. Numerical Results

In this section, we numerically investigate the WSR optimization problem for FD MIMO interference channel as a function of signal-to-noise ratio (SNR) and nominal interference to noise ratio (INR). For brevity, we set the same number of transmit and receive antennas at each node, i.e., $M_i = N_i = N, i = 1,\ldots,K$.[4] The transmit power constraint for each node in the system, i.e., $P_i^{(b)} = N$, $\forall(i,b)$, and $P_T = 2KN$. We also assumed the same weights, i.e., $\mu_i^{(a)} = \mu$, $\forall(i,a)$.

We define SNR of the nodes in the i-th pair as $\mathrm{SNR}_i = \mathrm{SNR} \triangleq \rho_i N, i = 1,\ldots,K$, and the nominal INR from the nodes in the j-th pair to the nodes in the i-th pair as $\mathrm{INR}_{ij} = \mathrm{INR} \triangleq \eta_{ij}N, i,j = 1,\ldots,K, i\neq j$. The INR of the self-interference channel at the nodes in the i-th pair INR_{ii} is denoted as $\mathrm{INR_{SI}}$. We choose right singular matrices initialization, and average the results over 100 independent channel realizations.

Fig. 3 illustrates the convergence behavior of the proposed algorithm. It shows that both individual and sum-power constrained problems converge in few steps, and they do so monotonically.

Fig. 4 compares the achievable sum-rate for individual power constrained problem (13), (14) and sum-power constraint problem (13), (15) under different number of antennas. As it is seen in Fig. 4, at low SNR, sum-power constraint problem achieves higher sum-rate than the individual-power constraint problem,

[4]Note that although the nodes in ith link have $N_i + M_i$ antennas in total, similar to [26], [27], we assume that only $N_i(M_i)$ antennas can be used for transmission (reception) in HD mode. The reason is that in practical systems RF front-ends are scarce resources, since they are much more expensive than antennas. Therefore we assume that the nodes in the ith link only has N_i transmission front-ends and M_i receiving front-ends, and do not carry out antenna partitioning.

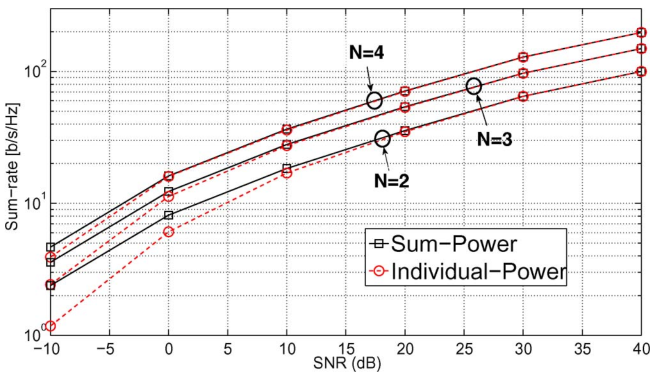

Fig. 4. Sum-rate comparison of the individual and sum-power constraint problems versus SNR. Here $K = 3$, INR $= 5$ dB, $\mathrm{INR_{SI}} = 20$ dB, $\mu = 1$, $\kappa = \beta = -40$ dB.

because the sum-power constraint problem is more relaxed than individual power constraint, and thus can allocate more power to the node that contributes more to achieve higher sum-rate, but both solutions converge at high SNR.

We now compare our proposed system, in which all the pairs operate in FD mode, and transmit at the same time (in particular, we have both self-interference, and inter-user interference) with the following baseline systems:

- **FD-TDMA**: All the pairs in the system operate in FD mode, but transmit on different time slots, like TDMA. Therefore, we have only self-interference, and do not have inter-user interference between different pairs. Particularly, in the first time slot, only the first pair transmits and receives in FD mode. In the second time slot, only the second pair transmits and receives in FD mode, and in the Kth time slot, only the Kth pair transmits and receives in FD mode. In this case, the sum-rate should be divided by the number of time slots (or pairs), in our case K.
- **HD**: All the pairs in the system operate in HD mode, but transmit at the same time. Particularly, in the first time slot, all the nodes on the left hand side in Fig. 1 transmit to their pairs on the right. And in the second time slot the nodes on the right hand side in Fig. 1 transmit to their pairs on the left. So in this case, we do not have self-interference, but we have inter-user interference, and sum-rate should be divided by 2 because of the HD scheme.
- **HD-TDMA**: All the nodes transmit sequentially, so we need $2K$ time slots. Therefore, we have neither self-interference nor inter-user interference.

The comparison of FD scheme with the baseline schemes are shown for in Fig. 5 and Fig. 6 for sum-power and individual power constraint problems, respectively. It is seen that under moderate interference levels, FD and HD-TDMA systems achieve the higher and lowest sum-rate, respectively, and FD-TDMA and HD schemes give similar performances. Since individual power and sum-power constrained problems have similar performance, from now on we will consider individual power constrained problem in our figures.

The effect of the self-interference, $\mathrm{INR_{SI}}$ on the proposed and baseline schemes is examined in Fig. 7. It is seen in Fig. 7 that when INR $= 10$ dB, the performance of FD system drops below that of HD and HD-TDMA schemes around $\mathrm{INR_{SI}} =$ 50 dB and $\mathrm{INR_{SI}} = 65$ dB, respectively. Moreover, the performance of FD-TDMA system drops below that of HD and HD-TDMA schemes around $\mathrm{INR_{SI}} = 35$ dB and $\mathrm{INR_{SI}} = 50$ dB, respectively.

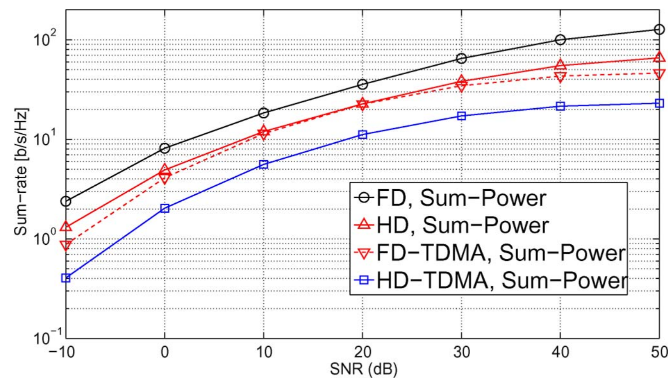

Fig. 5. Sum-rate comparison of the sum power constraint problem for different schemes versus SNR. Here $K = 3$, $N = 2$, INR $= 5$ dB, $\mathrm{INR_{SI}} = 20$ dB, $\mu = 1$, $\kappa = \beta = -40$ dB.

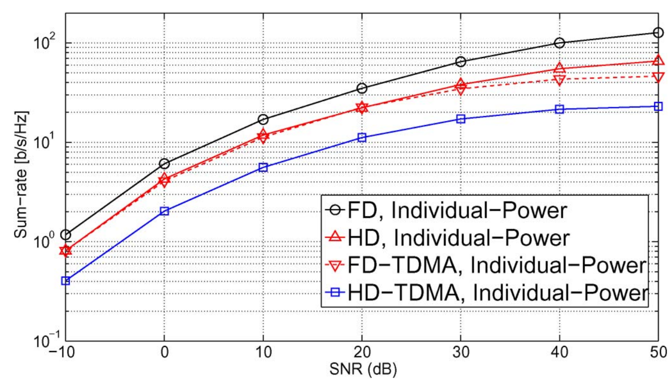

Fig. 6. Sum-rate comparison of the individual power constraint problem for different schemes versus SNR. Here $K = 3$, $N = 2$, INR $= 5$ dB, $\mathrm{INR_{SI}} = 20$ dB, $\mu = 1$, $\kappa = \beta = -40$ dB.

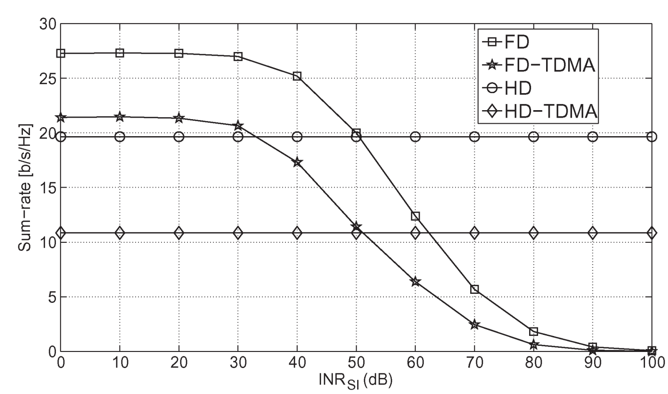

Fig. 7. Sum-rate comparison of different schemes versus $\mathrm{INR_{SI}}$. Here $K = 3$, $N = 2$, SNR $= 20$ dB, INR $= 10$ dB, $\mu = 1$, $\kappa = \beta = -40$ dB.

In Fig. 8, the performance of FD and HD-TDMA (the best and worst schemes under moderate interference levels, see Fig. 6) are compared with respect to INR for various $\mathrm{INR_{SI}}$ values to detect the self-interference and inter-user interference level that results in lower achievable rate for FD system than HD-TDMA system. It is seen that the performance gap between $\mathrm{INR_{SI}} = 0$ dB and $\mathrm{INR_{SI}} = 20$ dB is indistinguishable, and

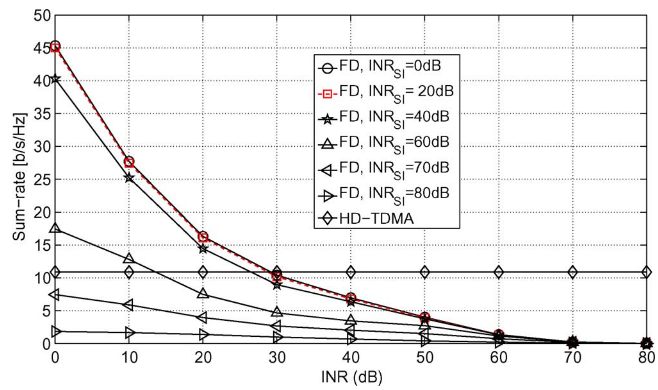

Fig. 8. Sum-rate comparison of different schemes versus INR. Here $K = 3$, $N = 2$, SNR $= 20$ dB, $\mu = 1$, $\kappa = \beta = -40$ dB.

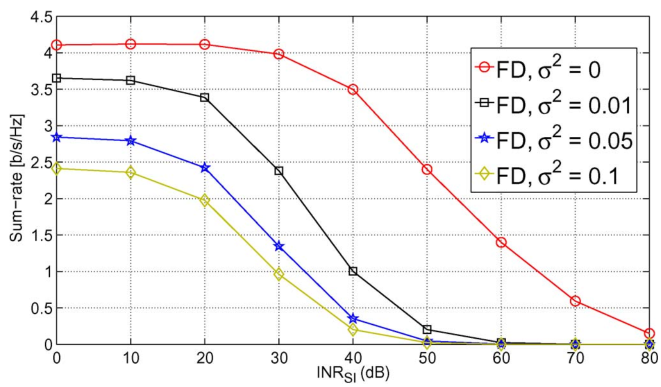

Fig. 9. Sum-rate of the proposed FD algorithm with different channel estimation errors versus $\mathrm{INR_{SI}}$. Here $N = 2$, $K = 1$, SNR $= 20$ dB, $\mu = 0.25$, $\kappa = \beta = -40$ dB.

around $\mathrm{INR_{SI}} = 65$ dB, HD-TDMA scheme starts outperforming FD scheme for all INR levels.

In Fig. 9, we investigate the role of channel estimation errors on the sum-rate of the individual power constrained problem. We adopt the channel model used in [11], [12], [28], [55], where the estimation error is modeled as $\Delta\mathbf{H} = \mathbf{H} - \tilde{\mathbf{H}}$, where $\tilde{\mathbf{H}}$ and $\Delta\mathbf{H}$ are uncorrelated, and the entries of $\Delta\mathbf{H}$ are zero mean circularly symmetric complex Gaussian with variance σ^2. It can be seen that as the channel estimation error, σ^2, increases, the sum-rate decreases, and the gap between sum-rate curves diminishes.

In our last example, we compare our proposed algorithm with the GB algorithm [60]. The simulation results in Fig. 10 confirm that the proposed algorithm provides the sum rate performance close to the GB algorithm, but as it is mentioned before GB algorithm requires global CSI, while our algorithm requires only the *local* CSI.

VI. Conclusion

In this work, we have addressed the transmit and receive filter design for WSR maximization problem in FD bi-directional MIMO interference channel and in FD cellular systems that suffer from self-interference and interference from other links under the limited DR at the transmitters and receivers. Both sum-power and individual power constraints were considered. Since the globally optimal solution is difficult to obtain due to

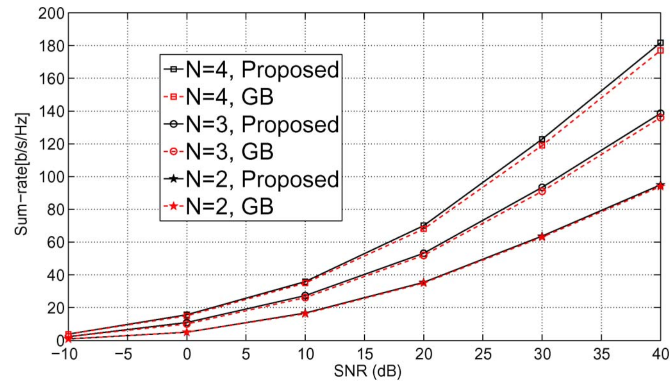

Fig. 10. Sum-rate comparison of the proposed and GB algorithms with different antenna numbers N values versus SNR. Here $K = 3$, INR $= 5$ dB, $\mathrm{INR_{SI}} = 20$ dB, $\mu = 1$, $\kappa = \beta = -40$ dB.

the non-convex nature of the problems, an alternating iterative algorithm to find a local WSR optimum was proposed based on the relationship between WSR and WMMSE problems. It is shown in simulations that the sum-rate achieved by FD mode is higher than the sum-rate achieved by baseline HD schemes at moderate interference levels, but its performance is outperformed by baseline schemes at high interference levels.

Appendix A
The Relationship Between WSR and WMMSE Problems

To investigate stationary points of the optimization problems, the Lagrangian functions of the optimization problems (13)–(15) and (21)–(23) under the fixed weight matrix $\mathbf{W}_i^{(a)}$ can be written, respectively, as

$$\begin{aligned}\mathcal{L}_1 = &-\sum_{i=1}^{K}\sum_{a=1}^{2}\mu_i^{(a)}\, I_i^{(a)} \\ &+Q\lambda\left(\sum_{i=1}^{K}\sum_{b=1}^{2}\mathrm{tr}\left\{\mathbf{V}_i^{(b)}\left(\mathbf{V}_i^{(b)}\right)^H\right\}-P_T\right) \\ &+(1-Q)\sum_{i=1}^{K}\sum_{b=1}^{2}\lambda_i^{(b)}\left(\mathrm{tr}\left\{\mathbf{V}_i^{(b)}\left(\mathbf{V}_i^{(b)}\right)^H\right\}-P_i^{(b)}\right),\end{aligned} \tag{49}$$

$$\begin{aligned}\mathcal{L}_2 = &\sum_{i=1}^{K}\sum_{a=1}^{2}\mathrm{tr}\left\{\mathbf{W}_i^{(a)}\mathbf{E}_i^{(a)}\right\} \\ &+Q\lambda\left(\sum_{i=1}^{K}\sum_{b=1}^{2}\mathrm{tr}\left\{\mathbf{V}_i^{(b)}\left(\mathbf{V}_i^{(b)}\right)^H\right\}-P_T\right) \\ &+(1-Q)\sum_{i=1}^{K}\sum_{b=1}^{2}\lambda_i^{(b)}\left(\mathrm{tr}\left\{\mathbf{V}_i^{(b)}\left(\mathbf{V}_i^{(b)}\right)^H\right\}-P_i^{(b)}\right),\end{aligned} \tag{50}$$

where Q selects the desired power constraint ($Q = 1$ for the sum power constraint and $Q = 0$ for the individual power constraint), λ and $\lambda_i^{(b)}$ denote the Lagrange multipliers for the sum power constraint and individual power constraint at the node $i^{(b)}$, respectively. The gradients of both Lagrangian functions (49) and (50) with respect to $\mathbf{V}_i^{(b)}$ can be written, respectively, as

$$\frac{\partial \mathcal{L}_1}{\partial \mathbf{V}_i^{(b)*}} = -\frac{1}{\ln 2}\left(\sum_{i=1}^{K}\sum_{a=1}^{2}\mu_i^{(a)}\frac{\mathrm{tr}\left\{\mathbf{E}_i^{(a)}\partial\left(\mathbf{E}_i^{(a)}\right)^{-1}\right\}}{\partial \mathbf{V}_i^{(b)*}}\right) + \lambda Q\mathbf{V}_i^{(b)} + \lambda_i^{(b)}(1-Q)\mathbf{V}_i^{(b)}, \tag{51}$$

$$\frac{\partial \mathcal{L}_2}{\partial \mathbf{V}_i^{(b)*}} = -\left(\sum_{i=1}^{K}\sum_{a=1}^{2}\frac{\mathrm{tr}\left\{\mathbf{W}_i^{(a)}\mathbf{E}_i^{(a)}\partial\left(\mathbf{E}_i^{(a)}\right)^{-1}\mathbf{E}_i^{(a)}\right\}}{\partial \mathbf{V}_i^{(b)*}}\right) + \lambda Q\mathbf{V}_i^{(b)} + \lambda_i^{(b)}(1-Q)\mathbf{V}_i^{(b)}, \tag{52}$$

where we have used the matrix derivative formulas $\partial \ln|\mathbf{X}| = \mathrm{tr}\{\mathbf{X}^{-1}\partial\mathbf{X}\}$ and $\partial\mathbf{X}^{-1} = -\mathbf{X}^{-1}\partial\mathbf{X}\mathbf{X}^{-1}$.

Comparing (51) and (52), it is easy to see that given transmit filters $\mathbf{V}_i^{(b)}$ and MMSE-matrices $\mathbf{E}_i^{(a)}$, the gradient of WSR and the gradient of WMMSE problems are equal if the MSE-weights $\mathbf{W}_i^{(a)}$ are chosen as:

$$\mathbf{W}_i^{(a)} = \frac{\mu_i^{(a)}}{\ln 2}\left(\mathbf{E}_i^{(a)}\right)^{-1}. \tag{53}$$

Appendix B
The Computation of the Optimum Transmit Filter for the Sum-Power Constrained Transceiver Design

The Lagrangian function of the optimization problem (25), (26) can be expressed as

$$\begin{aligned}\mathcal{L} = \sum_{i=1}^{K}\sum_{a=1}^{2}\mathrm{tr}\Bigg\{\mathbf{W}_i^{(a)}\Bigg(&\mathbf{I}_{d_i} - \sqrt{\rho_i}\left(\bar{\mathbf{V}}_i^{(b)}\right)^H\left(\mathbf{H}_{ii}^{(ab)}\right)^H\left(\mathbf{R}_i^{(a)}\right)^H \\ &-\sqrt{\rho_i}\mathbf{R}_i^{(a)}\mathbf{H}_{ii}^{(ab)}\bar{\mathbf{V}}_i^{(b)} + \mathbf{R}_i^{(a)}\bar{\Sigma}_i^{(a)}\left(\mathbf{R}_i^{(a)}\right)^H \\ &+\rho_i\mathbf{R}_i^{(a)}\mathbf{H}_{ii}^{(ab)}\bar{\mathbf{V}}_i^{(b)}\left(\bar{\mathbf{V}}_i^{(b)}\right)^H\left(\mathbf{H}_{ii}^{(ab)}\right)^H\left(\mathbf{R}_i^{(a)}\right)^H\Bigg)\Bigg\} \\ &+\lambda\left(\alpha^2\sum_{i=1}^{K}\sum_{b=1}^{2}\mathrm{tr}\left\{\bar{\mathbf{V}}_i^{(b)}\left(\bar{\mathbf{V}}_i^{(b)}\right)^H\right\} - P_T\right),\end{aligned} \tag{54}$$

where $\bar{\Sigma}_i^{(a)}$ is defined at the bottom of the following page.

Taking the partial derivative of (54) with respect to the matrix $\bar{\mathbf{V}}_i^{(b)}$, we obtain

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \bar{\mathbf{V}}_i^{(b)*}} = &-\sqrt{\rho_i}\left(\mathbf{H}_{ii}^{(ab)}\right)^H\left(\mathbf{R}_i^{(a)}\right)^H\mathbf{W}_i^{(a)} \\ &+\rho_i\left(\mathbf{H}_{ii}^{(ab)}\right)^H\left(\mathbf{R}_i^{(a)}\right)^H\mathbf{W}_i^{(a)}\mathbf{R}_i^{(a)}\mathbf{H}_{ii}^{(ab)}\bar{\mathbf{V}}_i^{(b)} \\ &+\frac{\mathrm{tr}\left\{\left(\mathbf{R}_i^{(a)}\right)^H\mathbf{W}_i^{(a)}\mathbf{R}_i^{(a)}\partial\bar{\Sigma}_i^{(a)}\right\}}{\partial \bar{\mathbf{V}}_i^{(b)*}} \\ &+\frac{\mathrm{tr}\left\{\left(\mathbf{R}_i^{(b)}\right)^H\mathbf{W}_i^{(b)}\mathbf{R}_i^{(b)}\partial\bar{\Sigma}_i^{(b)}\right\}}{\partial \bar{\mathbf{V}}_i^{(b)*}} + \lambda\alpha^2\bar{\mathbf{V}}_i^{(b)} \\ &+\sum_{j\neq i}^{K}\sum_{c=1}^{2}\frac{\mathrm{tr}\left\{\left(\mathbf{R}_j^{(c)}\right)^H\mathbf{W}_j^{(c)}\mathbf{R}_j^{(c)}\partial\bar{\Sigma}_j^{(c)}\right\}}{\partial \bar{\mathbf{V}}_i^{(b)*}}.\end{aligned} \tag{56}$$

Using (55), (see equation at the bottom of the page) we can write

$$\begin{aligned}\frac{\mathrm{tr}\left\{\mathbf{A}_i^{(a)}\partial\bar{\Sigma}_i^{(a)}\right\}}{\partial \bar{\mathbf{V}}_i^{(b)*}} &= \Bigg(\rho_i\kappa\mathrm{diag}\left(\left(\mathbf{H}_{ii}^{(ab)}\right)^H\mathbf{A}_i^{(a)}\mathbf{H}_{ii}^{(ab)}\right) \\ &\quad+\rho_i\beta\left(\mathbf{H}_{ii}^{(ab)}\right)^H\mathrm{diag}\left(\mathbf{A}_i^{(a)}\right)\mathbf{H}_{ii}^{(ab)}\Bigg)\bar{\mathbf{V}}_i^{(b)},\end{aligned} \tag{57}$$

$$\begin{aligned}\frac{\mathrm{tr}\left\{\mathbf{A}_i^{(b)}\partial\bar{\Sigma}_i^{(b)}\right\}}{\partial \bar{\mathbf{V}}_i^{(b)*}} &= \Bigg(\eta_{ii}\kappa\mathrm{diag}\left(\left(\mathbf{H}_{ii}^{(bb)}\right)^H\mathbf{A}_i^{(b)}\mathbf{H}_{ii}^{(bb)}\right) \\ &\quad+\eta_{ii}\beta\left(\mathbf{H}_{ii}^{(bb)}\right)^H\mathrm{diag}\left(\mathbf{A}_i^{(b)}\right)\mathbf{H}_{ii}^{(bb)}\Bigg)\bar{\mathbf{V}}_i^{(b)},\end{aligned} \tag{58}$$

$$\begin{aligned}\frac{\mathrm{tr}\left\{\mathbf{A}_j^{(c)}\partial\bar{\Sigma}_j^{(c)}\right\}}{\partial \bar{\mathbf{V}}_i^{(b)*}} &= \Bigg(\eta_{ji}\left(\mathbf{H}_{ji}^{(cb)}\right)^H\mathbf{A}_j^{(c)}\mathbf{H}_{ji}^{(cb)} \\ &\quad+\eta_{ji}\kappa\mathrm{diag}\left(\left(\mathbf{H}_{ji}^{(cb)}\right)^H\mathbf{A}_j^{(c)}\mathbf{H}_{ji}^{(cb)}\right) \\ &\quad+\eta_{ji}\beta\left(\mathbf{H}_{ji}^{(cb)}\right)^H\mathrm{diag}\left(\mathbf{A}_j^{(c)}\right)\mathbf{H}_{ji}^{(cb)}\Bigg)\bar{\mathbf{V}}_i^{(b)},\end{aligned} \tag{59}$$

$$\begin{aligned}\bar{\Sigma}_i^{(a)} \approx\; &\rho_i\kappa\mathbf{H}_{ii}^{(ab)}\mathrm{diag}\left(\bar{\mathbf{V}}_i^{(b)}\left(\bar{\mathbf{V}}_i^{(b)}\right)^H\right)\left(\mathbf{H}_{ii}^{(ab)}\right)^H + \eta_{ii}\kappa\mathbf{H}_{ii}^{(aa)}\mathrm{diag}\left(\bar{\mathbf{V}}_i^{(a)}\left(\bar{\mathbf{V}}_i^{(a)}\right)^H\right)\left(\mathbf{H}_{ii}^{(aa)}\right)^H \\ &+\beta\rho_i\mathrm{diag}\left(\mathbf{H}_{ii}^{(ab)}\bar{\mathbf{V}}_i^{(b)}\left(\bar{\mathbf{V}}_i^{(b)}\right)^H\left(\mathbf{H}_{ii}^{(ab)}\right)^H\right) + \beta\eta_{ii}\mathrm{diag}\left(\mathbf{H}_{ii}^{(aa)}\bar{\mathbf{V}}_i^{(a)}\left(\bar{\mathbf{V}}_i^{(a)}\right)^H\left(\mathbf{H}_{ii}^{(aa)}\right)^H\right) \\ &+\sum_{j\neq i}^{K}\sum_{c=1}^{2}\eta_{ij}\left[\mathbf{H}_{ij}^{(ac)}\left(\bar{\mathbf{V}}_j^{(c)}\left(\bar{\mathbf{V}}_j^{(c)}\right)^H + \kappa\mathrm{diag}\left(\bar{\mathbf{V}}_j^{(c)}\left(\bar{\mathbf{V}}_j^{(c)}\right)^H\right)\right)\left(\mathbf{H}_{ij}^{(ac)}\right)^H\right] \\ &+\sum_{j\neq i}^{K}\sum_{c=1}^{2}\beta\eta_{ij}\mathrm{diag}\left(\mathbf{H}_{ij}^{(ac)}\bar{\mathbf{V}}_j^{(c)}\left(\bar{\mathbf{V}}_j^{(c)}\right)^H\left(\mathbf{H}_{ij}^{(ac)}\right)^H\right) + \alpha^{-2}\mathbf{I}_{M_i}.\end{aligned} \tag{55}$$

where $\mathbf{A}_j^{(c)} = \left(\mathbf{R}_j^{(c)}\right)^H \mathbf{W}_j^{(c)} \mathbf{R}_j^{(c)}$. By plugging (57)–(59) into (56) and making it equal to zero, we can obtain the optimal $\bar{\mathbf{V}}_i^{(b)}$ as

$$\bar{\mathbf{V}}_i^{(b)} = \sqrt{\rho_i}\left(\mathbf{X}_i^{(b)} + \lambda\alpha^2 \mathbf{I}_{N_i}\right)^{-1}\left(\mathbf{R}_i^{(a)}\mathbf{H}_{ii}^{(ab)}\right)^H \mathbf{W}_i^{(a)} \tag{60}$$

where $\mathbf{X}_i^{(b)}$ is defined in (30).

Taking the derivative of the Lagrange function (54) with respect to α, we obtain

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \alpha} = &-2\alpha^{-3}\sum_{i=1}^{K}\sum_{a=1}^{2}\operatorname{tr}\left\{\mathbf{W}_i^{(a)}\mathbf{R}_i^{(a)}\left(\mathbf{R}_i^{(a)}\right)^H\right\}\\ &+2\lambda\alpha\sum_{i=1}^{K}\sum_{b=1}^{2}\operatorname{tr}\left\{\bar{\mathbf{V}}_i^{(b)}\left(\bar{\mathbf{V}}_i^{(b)}\right)^H\right\} = 0,\\ &\Rightarrow \lambda\alpha^2\sum_{i=1}^{K}\sum_{b=1}^{2}\operatorname{tr}\left\{\bar{\mathbf{V}}_i^{(b)}\left(\bar{\mathbf{V}}_i^{(b)}\right)^H\right\}\\ &= \alpha^{-2}\sum_{i=1}^{K}\sum_{a=1}^{2}\operatorname{tr}\left\{\mathbf{W}_i^{(a)}\mathbf{R}_i^{(a)}\left(\mathbf{R}_i^{(a)}\right)^H\right\}.\end{aligned} \tag{61}$$

The complementary slackness condition of (25), (26) is:

$$\begin{aligned}&\lambda\left(\alpha^2\sum_{i=1}^{K}\sum_{b=1}^{2}\operatorname{tr}\left\{\bar{\mathbf{V}}_i^{(b)}\left(\bar{\mathbf{V}}_i^{(a)}\right)^H\right\} - P_T\right) = 0,\\ &\Rightarrow \lambda\alpha^2\sum_{i=1}^{K}\sum_{b=1}^{2}\operatorname{tr}\left\{\bar{\mathbf{V}}_i^{(b)}\left(\bar{\mathbf{V}}_i^{(b)}\right)^H\right\} = \lambda P_T.\end{aligned} \tag{62}$$

Plugging (61) into (62), we obtain

$$\lambda\alpha^2 = \frac{\sum_{i=1}^{K}\sum_{a=1}^{2}\operatorname{tr}\left\{\mathbf{W}_i^{(a)}\mathbf{R}_i^{(a)}\left(\mathbf{R}_i^{(a)}\right)^H\right\}}{P_T}. \tag{63}$$

Substituting (63) into (60), we get the desired result (29).

REFERENCES

[1] H. Ju, E. Oh, and D. Hong, "Catching resource-devouring worms in next-generation wireless relay systems: Two-way relay and full-duplex relay," *IEEE Comm. Mag.*, vol. 47, no. 9, pp. 58–65, Sep. 2009.

[2] D. Nguyen, L. Tran, P. Pirinen, and M. Latva-aho, "On the spectral efficiency of full-duplex small cell wireless systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 9, pp. 4896–4910, Sep. 2014.

[3] D. Nguyen, L.-N. Tran, P. Pirinen, and M. Latva-aho, "Precoding for full duplex multiuser MIMO systems: Spectral and energy efficiency maximization," *IEEE Trans. Signal Process.*, vol. 61, no. 16, pp. 4038–4050, Aug. 2013.

[4] A. C. Cirik, K. Rikkinen, and M. Latva-aho, "Joint subcarrier and power allocation for sum-rate maximization in OFDMA full-duplex systems," in *Proc. IEEE VTC-Spring*, Glasgow, Scotland, Apr. 2015, accepted for publication.

[5] S. Goyal *et al.*, "Improving small cell capacity with common-carrier full duplex radios," in *Proc. IEEE ICC*, Jun. 2014, pp. 4987–4993.

[6] W. Afifi and M. Krunz, "Exploiting self-interference suppression for improved spectrum awareness/efficiency in cognitive radio systems," in *Proc. IEEE INFOCOM*, Apr. 2013, pp. 1258–1266.

[7] W. Cheng, X. Zhang, and H. Zhang, "Imperfect full duplex spectrum sensing in cognitive radio networks," in *Proc. 3rd ACM Workshop Cognitive Radio Netw.*, Sep. 2011, pp. 1–6.

[8] N. Singh *et al.*, "Efficient and fair MAC for wireless networks with self-interference cancellation," in *Proc. Int. Symp. WiOpt*, May 2011, pp. 94–101.

[9] J. I. Choi, M. Jain, K. Srinivasan, P. Levis, and S. Katti, "Achieving single channel, full duplex wireless communication," in *Proc. Mobicom*, Sep. 2010, pp. 1–12.

[10] M. Jain *et al.*, "Practical, real-time, full duplex wireless," in *Proc. Mobicom*, Sep. 2011, pp. 301–312.

[11] 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.

[12] P. Lioliou, M. Viberg, M. Coldrey, and F. Athley, "Self-interference suppression in full-duplex MIMO relays," in *Proc. Asilomar Conf. Signals Syst. Comput.*, Oct. 2010, pp. 658–662.

[13] B. Chun and H. Park, "A spatial-domain joint-nulling method of self-interference in full-duplex relays," *IEEE Commun. Lett.*, vol. 16, no. 4, pp. 436–438, Apr. 2012.

[14] E. Antonio-Rodriguez, R. Lopez-Valcarce, T. Riihonen, S. Werner, and R. Wichman, "SINR optimization in wideband full-duplex MIMO relays under limited dynamic range," in *Proc. IEEE SAM Signal Process. Workshop*, Jun. 2014, pp. 177–180.

[15] Y. Hua, "An overview of beamforming and power allocation for MIMO relays," in *Proc. IEEE Mil. Commun. Conf.*, Nov. 2010, pp. 375–380.

[16] M. Vehkapera, T. Riihonen, and R. Wichman, "Asymptotic analysis of full-duplex bidirectional MIMO link with transmitter noise," in *Proc. IEEE Int. Symp. Pers. Indoor Mobile Radio Commun.*, Sep. 2013, pp. 1265–1270.

[17] D. W. Bliss, P. A. Parker, and A. R. Margetts, "Simultaneous transmission and reception for improved wireless network performance," in *Proc. IEEE 14th Workshop Statist. Signal Process.*, Aug. 2007, pp. 478–482.

[18] 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.

[19] M. Knox, "Single antenna full duplex communications using a common carrier," in *Proc. 13th Annu. IEEE WAMICON*, Apr. 2012, pp. 1–6.

[20] S. Hong, J. Mehlman, and S. Katti, "Picasso: Flexible RF and spectrum slicing," in *Proc. SIGCOMM*, Aug. 2012, pp. 37–48.

[21] Y. Hua, P. Liang, Y. Ma, A. C. Cirik, and Q. Gao, "A method for broadband full-duplex MIMO radio," *IEEE Signal Process. Lett.*, vol. 19, no. 12, pp. 793–796, Dec. 2012.

[22] Y. Hua, Y. Ma, P. Liang, and A. C. Cirik, "Breaking the barrier of transmission noise in full-duplex radio," in *Proc. IEEE MILCOM*, Nov. 2013, pp. 1558–1563.

[23] B. Debaillie *et al.*, "Analog/RF solutions enabling compact full-duplex radios," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1662–1673, Sep. 2014.

[24] D. Bharadia and S. Katti, "Full duplex MIMO radios," in *Proc. 11th USENIX Conf. NSDI*, Apr. 2014, pp. 359–372.

[25] Y. Hua *et al.*, "Radio self-interference cancellation by transmit beamforming, all-analog cancellation and blind digital tuning," *Signal Process.*, vol. 108, pp. 322–340, Mar. 2015.

[26] S. Li, R. Murch, and V. Lau, "Linear transceiver design for full-duplex multi-user MIMO system," in *Proc. IEEE ICC*, Jun. 2014, pp. 4921–4926.

[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, Jul. 2012.

[28] A. C. Cirik, Y. Rong, and Y. Hua, "Achievable rates of full-duplex MIMO radios in fast fading channels with imperfect channel estimation," *IEEE Trans. Signal Process.*, vol. 62, no. 15, pp. 3874–3886, Aug. 2014.

[29] S. P. Herath and T. Le-Ngoc, "Sum-rate performance and impact of self-interference cancellation on full-duplex wireless systems," in *Proc. IEEE Int. Symp. PIMRC*, Sep. 2013, pp. 881–885.

[30] S. Huberman and T. Le-Ngoc, "Sequential convex programming for full-duplex single-user MIMO systems," in *Proc. IEEE ICC*, Jun. 2014, pp. 5078–5082.

[31] W. Li, J. Lilleberg, and K. Rikkinen, "On rate region analysis of half- and full-duplex OFDM communication links," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1688–1698, Sep. 2014.

[32] A. C. Cirik, R. Wang, and Y. Hua, "Weighted-sum-rate maximization for bi-directional full-duplex MIMO systems," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Nov. 2013, pp. 1632–1636.

[33] A. C. Cirik, R. Wang, Y. Rong, and Y. Hua, "MSE based transceiver designs for bi-directional full-duplex MIMO systems," in *Proc. 15th IEEE Workshop SPAWC*, Jun. 2014, pp. 384–388.

[34] J. Zhang, O. Taghizadeh, and M. Haardt, "Transmit strategies for full-duplex point-to-point systems with residual self-interference," in *Proc. Int. ITG WSA*, Mar. 2013, pp. 1–8.

[35] J. Zhang, O. Taghizadeh, and M. Haardt, "Robust transmit beamforming design for full-duplex point-to-point MIMO systems," in *Proc. 10th ISWCS*, Aug. 2013, pp. 1–5.

[36] T. M. Kim, H. J. Yang, and A. Paulraj, "Distributed sum-rate optimization for full-duplex MIMO system under limited dynamic range," *IEEE Signal Process. Lett.*, vol. 20, no. 6, pp. 555–558, Jun. 2013.

[37] M. Zhou, H. Cui, L. Song, and B. Jiao, "Transmit-receive antenna pair selection in full duplex systems," *IEEE Wireless Commun. Lett.*, vol. 3, no. 1, pp. 34–37, Feb. 2014.

[38] D. Kim, H. Ju, S. Park, and D. Hong, "Effects of channel estimation error on full-duplex two-way networks," *IEEE Trans. Veh. Tech.*, vol. 62, no. 9, pp. 4666–4672, Nov. 2013.

[39] S. Huberman and T. Le-Ngoc, "Self-interference pricing for full-duplex MIMO systems," in *Proc. IEEE Global Telecommun. Conf.*, Dec. 2013, pp. 3902–3906.

[40] W. Cheng, X. Zhang, and H. Zhang, "Optimal dynamic power control for full duplex bidirectional-channel based wireless networks," in *Proc. IEEE INFOCOM*, Apr. 2013, pp. 3120–3128.

[41] C.-E. Chen and W.-H. Chung, "An iterative minmax per-stream MSE transceiver design for MIMO interference channel," *IEEE Wireless Commun. Lett.*, vol. 1, no. 3, pp. 229–232, Jun. 2012.

[42] F. Negro, S. P. Shenoy, I. Ghauri, and D. T. M. Slock, "On the MIMO interference channel," in *Proc. Inf. Theory Appl. Workshop*, Jan. 2010, pp. 1–9.

[43] S. S. Christensen, R. Agarwal, E. Carvalho, and J. M. Cioffi, "Weighted sum rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4792–4799, Dec. 2008.

[44] X. Dong, Y. Rong, and Y. Hua, "Cooperative power scheduling for a network of MIMO links," *IEEE Trans. Wireless Commun.*, vol. 9, no. 3, pp. 939–944, Mar. 2010.

[45] J. Shin and J. Moon, "Weighted-sum-rate-maximizing linear transceiver filters for the K-user MIMO interference channel," *IEEE Trans. Commun.*, vol. 60, no. 10, pp. 2776–2783, Oct. 2012.

[46] 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 J. Sel. Areas Commun.*, vol. 26, no. 6, pp. 867–876, Aug. 2008.

[47] W. Namgoong, "Modeling and analysis of nonlinearities and mismatches in AC-coupled direct-conversion receiver," *IEEE Trans. Wireless Commun.*, vol. 4, no. 1, pp. 163–173, Jan. 2005.

[48] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331–4340, Sep. 2011.

[49] S. J. Kim and G. B. Giannakis, "Optimal resource allocation for MIMO ad-hoc cognitive radio networks," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 3117–3131, May 2011.

[50] R. Mochaourab and E. Jorswieck, "Optimal beamforming in interference networks with perfect local channel information," *IEEE Trans. Signal Process.*, vol. 59, no. 3, pp. 1142–1157, Mar. 2011.

[51] Z. K. M. Ho and D. Gesbert, "Balancing egoism and altruism on interference channel: The MIMO case," in *Proc. IEEE ICC*, May 2010, pp. 1–5.

[52] R. Zakhour, Z. Ho, and D. Gesbert, "Distributed beamforming coordination in multicell MIMO channels," in *Proc. IEEE VTC Spring*, Apr. 2009, pp. 1–5.

[53] D. Nguyen and M. Krunz, "Power-efficient spatial multiplexing for multi-antenna MANETs," in *Proc. IEEE ICC*, Jun. 2012, pp. 4016–4020.

[54] K. T. Truong and R. W. Heath, Jr., "Joint transmit precoding for the relay interference broadcast channel," *IEEE Trans. Veh. Technol.*, vol. 62, no. 3, pp. 1201–1215, Mar. 2013.

[55] T. Yoo and A. J. Goldsmith, "Capacity and power allocation for fading MIMO channels with channel estimation error," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2203–2214, May 2006.

[56] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[57] K. B. Petersen and M. S. Pedersen, *The Matrix Cookbook*. Lyngby, Denmark: Technical Univ. of Denmark, Feb. 2008.

[58] M. Joham, K. Kusume, M. Gzara, W. Utschick, and J. Nossek, "Transmit wiener filter for the downlink of TDDDS-CDMA systems," in *Proc. IEEE 7th Int. Symp. Spread Spectrum Tech. Appl.*, Sep. 2002, vol. 1, pp. 9–13.

[59] H. Shen, B. Li, M. Tao, and X. Wang, "MSE-based transceiver designs for the MIMO interference channel," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3480–3489, Nov. 2010.

[60] H. Sung, S. H. Park, K. J. Lee, and I. Lee, "Linear precoder designs for K-user interference channels," *IEEE Trans. Wireless Commun.*, vol. 9, no. 1, pp. 291–301, Jan. 2010.

[61] K. Gomadam, V. R. Cadambe, and S. A. Jafar, "A distributed numerical approach to interference alignment and applications to wireless interference networks," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3309–3322, Jun. 2011.

[62] M. C. Bromberg, "Optimizing MIMO multipoint wireless networks assuming Gaussian other-user interference," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2352–2362, Oct. 2003.

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 the 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.

Rui Wang (M'14) received the B.S. degree from Anhui Normal University, Wuhu, China, in 2006, and the M.S. degree from Shanghai University, Shanghai, China, in 2009 and the Ph.D. degree from Shanghai Jiao Tong University, China, in 2013, all in electronic engineering. From August 2012 to February 2013, he was a visiting Ph.D. student at the Department of Electrical Engineering of University of California, Riverside. From October 2013 to October 2014, he was with the Institute of Network Coding, The Chinese University of Hong Kong as a postdoctoral research associate. Since October 2014, he has been with the College of Electronics and Information Engineering, Tongji University as an Assistant Professor. His research interests include wireless cooperative communications, MIMO technique, network coding, and OFDM.

Yingbo Hua (S'86–M'88–SM'92–F'02) received the B.S. degree from Southeast University, Nanjing, China, in 1982 and the Ph.D. degree from Syracuse University, Syracuse, NY, in 1988. He held a faculty position with the University of Melbourne, Australia, where he was promoted to the rank of Reader and Associate Professor from 1996. He was a Visiting Professor with Hong Kong University of Science and Technology (1999–2000), and a Consultant with Microsoft Research, WA (summer 2000). Since 2001, he has been a Professor with the University of California at Riverside.

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 on technical and/or organizing committees for over 50 international conferences and workshops. He has authored/coauthored over 300 articles and coedited three volumes of books, with more than 8000 citations in the fields of sensing, signal processing, and communications. He is a Fellow of IEEE and AAAS.

Matti Latva-aho (S'96–M'98–SM'06) was born in Kuivaniemi, Finland, in 1968. He received the M.Sc., Lic.Tech., and Dr.Tech (Hons.) degrees in electrical engineering from the University of Oulu, Finland, in 1992, 1996, and 1998, respectively. From 1992 to 1993, he was a Research Engineer at Nokia Mobile Phones, Oulu, Finland. During the years 1994–1998, he was a Research Scientist at Telecommunication Laboratory and Centre for Wireless Communications at the University of Oulu. Currently, he is the Department Chair Professor of Digital Transmission Techniques at the University of Oulu, Department for Communications Engineering. He was Director of Centre for Wireless Communications at the University of Oulu during 1998–2006 and Head of Department until 2014. His research interests are related to mobile broadband communication systems and currently his group focuses on 5G systems research. Dr. Latva-aho has published 300 conference or journal papers in the field of wireless communications. He has been TPC Chairman for PIMRC'06, TPC Co-Chairman for ChinaCom'07 and General Chairman for WPMC'08, CROWNCOM'14 and 5GU'14. He acted as the Chairman and vice-chairman of IEEE Communications Finland Chapter in 2000–2003.