Beamforming Design for Full-Duplex MIMO Interference Channels-QoS and Energy-Efficiency Considerations

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2016.2606411, IEEE Transactions on Communications

Beamforming Design for Full-Duplex MIMO Interference Channels–QoS and Energy-Efficiency Considerations

Ali Cagatay Cirik, *Member, IEEE*, Sudip Biswas, *Student Member, IEEE*, Satyanarayana Vuppala, *Member, IEEE*, and Tharmalingam Ratnarajah, *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. Two transmit beamforming design problems are solved, i) sum-power minimization problem subject to rate constraints, and ii) energy-efficiency maximization problem subject to individual power constraints. To tackle the sum-power minimization problem, we first generalize the well-known relationship between weighted-sum-rate (WSR) and weighted minimum-mean-squared-error (WMMSE) problems, originally used to solve the sum-rate maximization problems, and then propose a low complexity centralized algorithm which converges to a stationary point. To decrease the exchange of a huge amount of data and excessive signaling traffic among the nodes, a distributed algorithm is also proposed. For the energy-efficiency maximization problem, the original fractional form optimization problem is first transformed into an equivalent subtractive-form optimization problem by exploiting the properties of fractional programming, and then perform a dual-layer optimization scheme. In the outer layer, the energy-efficiency parameter is searched using a simple one-dimensional search, and in the inner layer, the relationship between WSR and WMMSE is exploited to solve the subtractive form optimization problem. Since the proposed algorithms require perfect channel-state-information (CSI), which is difficult to acquire in practice, we also propose a robust design, by taking the imperfect channel knowledge into consideration. It is shown in the simulations that the sum-power achieved in FD mode depends heavily on the transmitter/receiver distortion. Also the energy-efficiency of FD systems is lower than that of half-duplex (HD) systems, as FD nodes need to overcome self-interference and increased inter-user interference which leads to high power consumption.**

***Keywords*—Keywords-Energy-efficiency, full-duplex, MIMO interference channel, QoS, self interference, transceiver designs.**

This work is supported by the Seventh Framework Programme for Research of the European Commission under grant number ADEL-619647. T. Ratnarajah is supported by the U.K. Engineering and Physical Sciences Research Council (EPSRC) under grant number EP/L025299/1.

A. C. Cirik was with the Institute for Digital Communications, School of Engineering, University of Edinburgh, Edinburgh EH9 3JL, United Kingdom. He is now with the Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC V6T 1Z4, Canada (email: cirik@ece.ubc.ca)

S. Biswas, S. Vuppala, and T. Ratnarajah are with the Institute for Digital Communications, School of Engineering, University of Edinburgh, Edinburgh EH9 3JL, United Kingdom (email: {sudip.biswas, s.vuppala, t.ratnarajah}@ed.ac.uk). Correspondent author: S. Biswas.

I. Introduction

MOBILE data traffic has explosively grown in recent years, leading to an ever-growing demand for much higher capacity, lower latency and energy-efficiency in wireless networks. It has culminated in the development of the fifth generation (5G) wireless communication systems, expected to be deployed by the year 2020, with key goals of data rates in the range of Gbps, billions of connected devices, lower latency and low-cost, energy-efficient and environment-friendly operation. Half-duplex (HD) 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 (FD) wireless communication systems on the other hand enable simultaneous transmission and reception at the same time in the same frequency band, and can be a promising technique to potentially double the link capacity, and meet the projected increase in spectral efficiency. Such systems have recently gained considerable attention in academia [1]-[8].

The limiting factor on the performance of FD systems is the strong self-interference at the front-end of the receiver created by the signal leakage from the transmitter antennas of a FD node to its own receiver antennas. Promising results from experimental research that demonstrate the feasibility of FD transmission using the off-the-shelf hardware are available in [1]-[3], although RF front-end interference cancellation is still an on-going research topic. 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. The residual self-interference can still significantly affect the performance of the system. Depending on the strength of the residual self-interference, optimal transmit strategies for a HD system can be far from optimal for the FD system. If the residual self-interference is not well handled, it can still prevent us from exploiting the benefits of FD wireless communications. Therefore, optimization problems (power allocation, transceiver beamforming, etc.) related to FD systems under this residual self-interference were considered in [4]-[8].

Of late, the multiple-input multiple-output (MIMO) interference channels have attracted significant attention, since it is the fundamental model behind many practical problems [9]. The performance of cellular communication systems (open spectrum, multi-cell systems, etc.), where each cell causes

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2016.2606411, IEEE Transactions on Communications

interference to other cells can be carried out by focusing on MIMO interference channels [10]. MIMO interference channel for FD systems has been studied in [7] for sum-rate maximization and [8] for sum mean-squared-error (MSE) minimization problems. To the best of our knowledge, sum-power minimization and energy-efficiency maximization problems for FD MIMO interference-channels have not been studied so far. The authors in [4]-[7] have focused on the maximization of the achievable rate and have not addressed the issue of Quality-of-Service (QoS). However, in most practical cases, each user has a desired data rate and likes to achieve it within the available power. Thus, it is also an important problem to guarantee all the users' desired data rates in a system while consuming minimum power [11]-[14].

Motivated by above, we provide an algorithm that minimizes the sum-power in the system while guaranteeing the data rate required by each user. In particular, we compute the minimum sum-power consumed at a K-link FD MIMO interference channel subject to rate constraints at each node of the system. Each link has two FD nodes exchanging information simultaneously, and thus the nodes in each pair suffer from self-interference due to operating in FD mode, and co-channel-interference (CCI) due to simultaneous transmission at all links. The method minimizes the sum-power consumed in the system resulting in reduced amount of self-interference at the users and CCI between users that are using the same frequency band at the same time. In [15] and [16], an interesting solution for the design of the transmit matrices for the weighted-sum-rate (WSR) problem in a MIMO broadcast and interfering broadcast channel is developed, respectively. Instead of directly focusing on the WSR problem, a relation is established to the weighted minimum-mean-squared (WMMSE) problem, which is simpler to solve. The original WSR problem is then solved through the WMMSE problem. This solution was generalized to solve the sum-power minimization problem under the individual rate constraints for MIMO interference networks in [14]. Here, we adopt this approach to solve the QoS problem in FD MIMO interference channels, and develop a centralized low complexity algorithm to find a stationary point.

The proposed centralized algorithm has a fast convergence speed for small scale networks. However, it is not suitable for large scale networks because the implementation of the centralized algorithm requires a central scheduler to coordinate the calibration of channel matrices, collect all channel matrices, and then compute and distribute the transmitter covariance matrices of all links. Hence, we propose a distributed dual-layer iterative algorithm, in which the computation of the whole problem is decomposed into many smaller subproblems, and is solved at each node with local channel-state-information (CSI) and limited information exchange, so the computational overhead is controllable as network grows.

Traditionally, the efficiency of a communication system has been measured in terms of spectral efficiency, which is directly related to the channel capacity in bit/s. This metric enables us to identify how efficiently a limited frequency spectrum is utilized. However, it fails to provide any insight on the efficiency of the energy consumption of the system. Energy-efficiency is currently one of the primary design goals of any wireless communication system because of the increasing gap between power consumption of signal processing circuits and battery capacity. Improved energy-efficiency involves maximizing a "throughput per Joule" metric [17]. Hence, to obtain an insight on the efficiency of energy consumption of communication systems, it is imperative to incorporate an energy-efficiency metric in the performance evaluation framework.

Since a node transmits and receives simultaneously in the FD system, it requires more energy than an HD node to mitigate the self-interference and CCI [18], [19]. Hence, if the transmit beamforming matrices are not designed properly, the energy-efficiency of the FD system can be outperformed by that of the HD system [6]. In this paper we study the transmit beamforming design for energy-efficiency maximization in FD MIMO interference channels. Since the original problem is non-convex due to the coupling between variables and its fractional form, we first transform the problem into an equivalent subtractive form by exploiting the fractional programming [20]-[21], and then propose a two-layer approach. In the outer layer, we apply one dimension search to compute the energy-efficiency parameter, while in the inner layer, an efficient beamforming optimization problem is solved by exploiting the relationship between WSR and WMMSE problems.

Since perfect CSI is assumed to be available at the transmitters for the aforementioned algorithms, which is practically impossible due to the inaccurate channel estimation, robust transceiver designs that take into account imperfect channel knowledge are of interest. Therefore, we propose a robust precoder scheme to maximize the energy-efficiency of the network subject to power constraints at the users under norm-bounded channel estimation errors. We adopt an alternating iterative approach to solve this non-convex optimization problem which is proven to converge, wherein a convex sub-problem is solved at each step.

Notations: 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.

II. System Model

In this section, we describe the system model of a FD MIMO interference channel consisting of K pairs as seen in Fig. 1. The signals mentioned below are defined in complex baseband. Each pair is equipped with multiple antennas and exchanges information simultaneously in a two way communication. We assume that the FD nodes in the 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 [4], which

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2016.2606411, IEEE Transactions on Communications

has also been commonly used in [7], [8]. 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, 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 [22] and [23] 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 desired 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 at the node $i^{(a)}$. $\mathbf{H}_{ij}^{(ac)} \in \mathbb{C}^{M_i \times N_j}$ denotes the CCI 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, $(i,j) \in \{1,\ldots,K\},\ j \neq i$ and $(a,c) \in \{1,2\}$. Note that unlike self-interference, to the best of our knowledge, no cancellation algorithm has been implemented for the CCI so far. However, to handle the CCI, there are some methods proposed in literature for FD cellular systems, which are out of scope of our paper:

1) Avoid scheduling proximate users that cause serious CCI to each other. For example, when a user sees strong CCI from the other user, scheduler can schedule them in orthogonal resources [24].
2) Use a subcarrier assignment that allocates each subcarrier to a pair of uplink and downlink nodes that have lower CCI gain compared to the UL channel gain (CCI gain should be smaller than the uplink channel gain to prevent the excessive CCI) [25].
3) Focus on the scenario where a FD base-station communicates to FD users instead of HD users, and the users are allocated exclusive subcarriers used for downlink and uplink communications to avoid the CCI [26].

The transmitted data vector 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 is assumed to be complex, zero mean, independent and identically distributed (i.i.d.) with $\mathbb{E}\left\{\mathbf{d}_i^{(a)}\left(\mathbf{d}_i^{(a)}\right)^H\right\} = \mathbf{I}_{d_i}$ The $N_i \times 1$ signal vector transmitted by node $i^{(a)}$ is given by

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

where $\mathbf{V}_i^{(a)} \in \mathbb{C}^{N_i \times d_i}$ represents the precoding matrix.

The received signal at node $i^{(a)}$ is written as

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

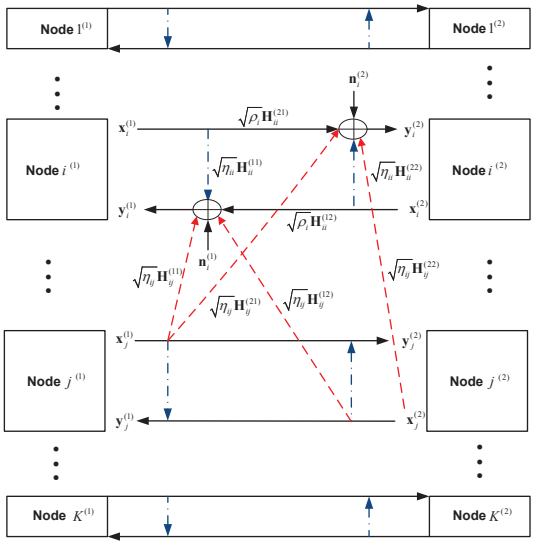

Fig. 1. An illustration of a full-duplex MIMO interference channel.

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 (2), 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 CCI channel between the nodes at the ith and jth pair.

Moreover, in (2), $\mathbf{c}_i^{(a)} \in \mathbb{C}^{N_i},\ i \in \{1,\ldots,K\},\ a \in \{1,2\}$ 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 κ ($\kappa \ll 1$) times the energy of the intended signal at each transmit antenna [4]. In particular $\mathbf{c}_i^{(a)}$ is modeled as

$$\mathbf{c}_i^{(a)} \sim \mathcal{CN}\left(\mathbf{0}, \kappa\,\mathrm{diag}\left(\mathbf{V}_i^{(a)}\left(\mathbf{V}_i^{(a)}\right)^H\right)\right),\ \mathbf{c}_i^{(a)} \perp \mathbf{x}_i^{(a)}, \tag{3}$$

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

Finally, in (2), $\mathbf{e}_i^{(a)} \in \mathbb{C}^{M_i},\ i \in \{1,\ldots,K\},\ a \in \{1,2\}$ 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 β ($\beta \ll 1$) times the energy of the undistorted received signal at each receive antenna [4]. In particular, $\mathbf{e}_i^{(a)}$ is modeled as

$$\mathbf{e}_i^{(a)} \sim \mathcal{CN}\left(\mathbf{0}, \beta\mathrm{diag}\left(\mathbf{\Phi}_i^{(a)}\right)\right),\ \mathbf{e}_i^{(a)} \perp \mathbf{u}_i^{(a)}, \tag{4}$$

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

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2016.2606411, IEEE Transactions on Communications

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 [4]. The self-interference canceled signal can then be written as

$$\begin{aligned}\tilde{\mathbf{y}}_i^{(a)} &= \mathbf{y}_i^{(a)} - \sqrt{\eta_{ii}}\,\mathbf{H}_{ii}^{(aa)}\mathbf{x}_i^{(a)} \\ &= \sqrt{\rho_i}\,\mathbf{H}_{ii}^{(ab)}\mathbf{x}_i^{(b)} + \mathbf{v}_i^{(a)}, \end{aligned} \quad (5)$$

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

$$\begin{aligned}\mathbf{v}_i^{(a)} = \sqrt{\rho_i}\,\mathbf{H}_{ii}^{(ab)}\mathbf{c}_i^{(b)} + \sqrt{\eta_{ii}}\,\mathbf{H}_{ii}^{(aa)}\mathbf{c}_i^{(a)} + \mathbf{e}_i^{(a)} + \mathbf{n}_i^{(a)} \\ + \sum_{j\neq i}^{K}\sum_{c=1}^{2}\sqrt{\eta_{ij}}\,\mathbf{H}_{ij}^{(ac)}\left(\mathbf{x}_j^{(c)} + \mathbf{c}_j^{(c)}\right). \end{aligned} \quad (6)$$

Using (3)-(4), similar to [4], $\mathbf{\Sigma}_i^{(a)}$, the covariance matrix of $\mathbf{v}_i^{(a)}$, is approximated as in (7) shown at the bottom of the following page [1]. The achievable rate of the node $i^{(a)}$, under Gaussian signaling, can be written as

$$\begin{aligned}&I_i^{(a)}(\mathbf{V}) \\ &= \log_2\left|\mathbf{I}_{M_i} + \rho_i\mathbf{H}_{ii}^{(ab)}\mathbf{V}_i^{(b)}\left(\mathbf{V}_i^{(b)}\right)^H\left(\mathbf{H}_{ii}^{(ab)}\right)^H\left(\mathbf{\Sigma}_i^{(a)}\right)^{-1}\right|, \end{aligned} \quad (8)$$

where $\mathbf{V} = \left\{\mathbf{V}_i^{(a)} : \forall(i,a)\right\}$ is the set of all precoding matrices.

Transceiver design in FD MIMO interference channels has also been studied in [7] and [8] under the same transmitter/receiver distortion model that we adopt in this paper. However, unlike the sum-rate maximization problem in [7] and sum-MSE minimization problem in [8], in this paper we consider two different objective functions, i.e., sum-power minimization and energy-efficiency maximization. The reasons for the choice of these metrics are:

- The solution of the sum rate maximization and MSE minimization problems favor the users in good channel conditions, and cannot ensure that all the users in the system are served with an acceptable QoS. In most practical cases, each user has a desired data rate and likes to achieve it within the available power. In this paper, we will therefore study the problem of minimizing the sum transmit power subject to achievable rate constraints for all users.
- Green communications have drawn increasing attention recently not only because of the rapid traffic increase with the popularity of smart phones but also the limited energy supply with ever increasing prices. With the promotion of green transmission in 5G networks, which emphasizes on incorporating energy awareness in communication systems, it is preferable to minimize the total transmission power consumed in the system or to maximize the energy-efficiency compared to the sum-rate maximization [7] and MSE minimization problems [8].
- Since the transmission and reception in the FD system operate simultaneously in an active mode, the FD system requires more energy than a HD one. Hence, if the transceivers are not designed properly, the power/energy consumption of the FD system can be much higher than that of the HD system. This motivates us to study the transceiver design for the sum-power minimization and energy-maximization problems in FD MIMO interference channels.

In [7], the well-known WSR and WMMSE relationship has been exploited to solve the sum-rate maximization problem, and in [8] second-order-cone programming (SOCP) method has been applied to solve the MSE-based transceiver design problems. In Section III, based on [14], we extend the relationship in [7] to solve the sum-power minimization problem. However, unlike [7], where this relationship results in a distributed solution, in the sum-power minimization problem, this relationship results in a centralized solution, and thus to decrease the exchange of a huge amount of data and excessive signaling traffic among the nodes in the centralized solution, in Section III-B, we propose a dual-decomposition based distributed algorithm [27]. As for the energy-efficiency maximization problem, in Section IV, we first propose to use a parametric approach known as the Dinkelbach algorithm [20], which allows us to express the fractional energy-efficiency objective function in a parametric programming problem, and then apply the well-known WSR and WMMSE relationship as used in [17]. Finally, the proposed transceiver designs are based on the assumption that perfect CSI is available at the nodes, which is not realistic in practice. Hence, in Section IV-B, we extend our proposed algorithms to robust transceiver designs under the assumption that the knowledge of CSI is imperfect, and propose semidefinite programming (SDP)-based method to solve these optimization problems.

III. Sum-Power Minimization

The sum-power minimization problem subject to individual rate constraints is formulated as follows:

$$\min_{\mathbf{V}} \quad \sum_{i=1}^{K}\sum_{a=1}^{2}\mathrm{tr}\left\{\mathbf{V}_i^{(a)}\left(\mathbf{V}_i^{(a)}\right)^H\right\} \quad (9)$$

$$\text{s.t.} \quad I_i^{(a)}(\mathbf{V}) \geq R_i^{(a)},\ i=1,\ldots,K,\ a=1,2, \quad (10)$$

where $R_i^{(a)}$ is the desired data rate at the node $i^{(a)}$ in bits/s/Hz.

In this section, adopting the approach in [14], we generalize the method in [15], [16] to find a stationary point for the sum-power minimization problem (9)-(10). To that end, we first need to establish the relationship between the achievable rate and the MSE matrix in the FD MIMO interference channels. Following the same steps as in [7], we can show the relationship between the achievable rate and the minimum-mean-squared-error (MMSE)-matrix as

$$I_i^{(a)}(\mathbf{V}) = \log_2\left|\left(\mathbf{E}_i^{(a)}(\mathbf{V})\right)^{-1}\right|, \quad (11)$$

[1]Note that in (7), the terms including the multiplication of κ and β are negligible, and have been ignored in the approximation, since in practice $\kappa \ll 1$ and $\beta \ll 1$ as discussed in [4].

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2016.2606411, IEEE Transactions on Communications

where MMSE matrix $\mathbf{E}_i^{(a)}(\mathbf{V})$ is defined as

$$\mathbf{E}_i^{(a)}(\mathbf{V}) = \left(\mathbf{I}_{d_i} + \rho_i \left(\mathbf{V}_i^{(b)}\right)^H \left(\mathbf{H}_{ii}^{(ab)}\right)^H \left(\mathbf{\Sigma}_i^{(a)}\right)^{-1} \mathbf{H}_{ii}^{(ab)} \mathbf{V}_i^{(b)}\right)^{-1}. \quad (12)$$

Now consider a sum-power minimization problem under the individual weighted MSE constraints as

$$\min_{\mathbf{W},\mathbf{R},\mathbf{V}} \quad \sum_{i=1}^{K}\sum_{a=1}^{2} \text{tr}\left\{\mathbf{V}_i^{(a)}\left(\mathbf{V}_i^{(a)}\right)^H\right\} \quad (13)$$

$$\text{s.t.} \quad \text{tr}\left\{\mathbf{W}_i^{(a)}\mathbf{MSE}_i^{(a)}\right\} - \log_2\left|\ln 2\mathbf{W}_i^{(a)}\right| - \frac{d_i}{\ln 2} \le -R_i^{(a)},\ i=1,\ldots,K,\ a=1,2, \quad (14)$$

where $\mathbf{W}_i^{(a)} \in \mathbb{C}^{d_i \times d_i}$ denotes a weight matrix associated with the node $i^{(a)}$ and $\mathbf{W} = \left\{\mathbf{W}_i^{(a)} :\ \forall(i,a)\right\}$ is the set of all weight matrices. Here, $\mathbf{MSE}_i^{(a)}$ is the MSE matrix of the node $i^{(a)}$, and can be written as

$$\mathbf{MSE}_i^{(a)} = \left(\sqrt{\rho}_i\mathbf{R}_i^{(a)}\mathbf{H}_{ii}^{(ab)}\mathbf{V}_i^{(b)} - \mathbf{I}_{d_i}\right)\left(\sqrt{\rho}_i\mathbf{R}_i^{(a)}\mathbf{H}_{ii}^{(ab)}\mathbf{V}_i^{(b)} - \mathbf{I}_{d_i}\right)^H + \mathbf{R}_i^{(a)}\mathbf{\Sigma}_i^{(a)}\left(\mathbf{R}_i^{(a)}\right)^H, \quad (15)$$

where $\mathbf{R}_i^{(a)} \in \mathbb{C}^{d_i \times M_i}$ is the linear receiver applied at node $i^{(a)}$ to estimate the signal transmitted from node $i^{(b)}$ and $\mathbf{R} = \left\{\mathbf{R}_i^{(a)} :\ \forall(i,a)\right\}$ is the set of all receiver matrices. The optimal receiver at the node $i^{(a)}$ for the problem is MMSE receiver filter, and can be expressed as

$$\bar{\mathbf{R}}_i^{(a)} = \sqrt{\rho_i}\left(\mathbf{V}_i^{(b)}\right)^H\left(\mathbf{H}_{ii}^{(ab)}\right)^H \times \left(\rho_i\mathbf{H}_{ii}^{(ab)}\mathbf{V}_i^{(b)}\left(\mathbf{V}_i^{(b)}\right)^H\left(\mathbf{H}_{ii}^{(ab)}\right)^H + \mathbf{\Sigma}_i^{(a)}\right)^{-1}. \quad (16)$$

Similar to [15], [16], it can be shown that the gradient of the problem (9)-(10) and the gradient of the problem (13)-(14) are equal if the MSE-weights $\mathbf{W}_i^{(a)}$ are chosen as:

$$\mathbf{W}_i^{(a)} = \frac{1}{\ln 2}\left(\mathbf{E}_i^{(a)}(\mathbf{V})\right)^{-1}. \quad (17)$$

Since the Karush-Kuhn-Tucker (KKT) conditions of the (9)-(10) and (13)-(14) problems can be satisfied simultaneously with the choice of MSE-weights (17), we can solve the problem (9)-(10) through solving the problem (13)-(14). In other words, if $\bar{\mathbf{V}}$, $\bar{\mathbf{W}}$, $\bar{\mathbf{R}}$ denote the optimal solution of the problem (13)-(14), then $\bar{\mathbf{V}}$ is also the optimal solution of the problem (9)-(10). Hence we only need to solve problem (13)-(14), which is convex in each of the optimization variables $\mathbf{V}$, $\mathbf{W}$, $\mathbf{R}$. We can use the block coordinate decent method to solve (13)-(14). First, for fixed $\mathbf{V},\mathbf{R}$, the optimal $\bar{\mathbf{W}}$ is computed from (17). Secondly, for fixed $\mathbf{V},\mathbf{W}$, the optimal $\bar{\mathbf{R}}$ is the MMSE receiver given in (16). Finally, for fixed $\mathbf{R},\mathbf{W}$, the optimal transmit beamforming matrix, $\bar{\mathbf{V}}$ can be obtained by solving (13)-(14) using the Lagrange dual method as explained below.

The Lagrange function of the problem (13)-(14) is written as

$$\mathcal{L}(\boldsymbol{\lambda},\mathbf{V},\mathbf{W},\mathbf{R}) = \sum_{i=1}^{K}\sum_{a=1}^{2}\text{tr}\left\{\mathbf{V}_i^{(a)}\left(\mathbf{V}_i^{(a)}\right)^H\right\} + \sum_{i=1}^{K}\sum_{a=1}^{2}\lambda_i^{(a)}\left(\text{tr}\left\{\mathbf{W}_i^{(a)}\mathbf{MSE}_i^{(a)}\right\} - \log_2\left|\ln 2\mathbf{W}_i^{(a)}\right| - \frac{d_i}{\ln 2} + R_i^{(a)}\right), \quad (18)$$

where $\boldsymbol{\lambda} = \left\{\lambda_i^{(a)} :\ \forall(i,a)\right\}$ is the Lagrange multiplier vector. The dual function of the problem (13)-(14) with fixed $\mathbf{R},\mathbf{W}$ is

$$\mathcal{J}(\boldsymbol{\lambda}) = \min_{\mathbf{V}}\ \mathcal{L}(\boldsymbol{\lambda},\mathbf{V},\mathbf{W},\mathbf{R}) = \min_{\mathbf{V}}\ \sum_{i=1}^{K}\sum_{a=1}^{2}\text{tr}\left\{\mathbf{V}_i^{(a)}\left(\mathbf{V}_i^{(a)}\right)^H\right\} + \sum_{i=1}^{K}\sum_{a=1}^{2}\lambda_i^{(a)}\text{tr}\left\{\mathbf{W}_i^{(a)}\mathbf{MSE}_i^{(a)}\right\}. \quad (19)$$

$$\begin{aligned}\mathbf{\Sigma}_i^{(a)} \approx{}& \rho_i\kappa\mathbf{H}_{ii}^{(ab)}\text{diag}\left(\mathbf{V}_i^{(b)}\left(\mathbf{V}_i^{(b)}\right)^H\right)\left(\mathbf{H}_{ii}^{(ab)}\right)^H + \eta_{ii}\kappa\mathbf{H}_{ii}^{(aa)}\text{diag}\left(\mathbf{V}_i^{(a)}\left(\mathbf{V}_i^{(a)}\right)^H\right)\left(\mathbf{H}_{ii}^{(aa)}\right)^H \\ &+ \beta\rho_i\text{diag}\left(\mathbf{H}_{ii}^{(ab)}\mathbf{V}_i^{(b)}\left(\mathbf{V}_i^{(b)}\right)^H\left(\mathbf{H}_{ii}^{(ab)}\right)^H\right) + \beta\eta_{ii}\text{diag}\left(\mathbf{H}_{ii}^{(aa)}\mathbf{V}_i^{(a)}\left(\mathbf{V}_i^{(a)}\right)^H\left(\mathbf{H}_{ii}^{(aa)}\right)^H\right) \\ &+ \sum_{j\neq i}^{K}\sum_{c=1}^{2}\eta_{ij}\left[\mathbf{H}_{ij}^{(ac)}\left(\mathbf{V}_j^{(c)}\left(\mathbf{V}_j^{(c)}\right)^H + \kappa\text{diag}\left(\mathbf{V}_j^{(c)}\left(\mathbf{V}_j^{(c)}\right)^H\right)\right)\left(\mathbf{H}_{ij}^{(ac)}\right)^H\right] \\ &+ \sum_{j\neq i}^{K}\sum_{c=1}^{2}\beta\eta_{ij}\text{diag}\left(\mathbf{H}_{ij}^{(ac)}\mathbf{V}_j^{(c)}\left(\mathbf{V}_j^{(c)}\right)^H\left(\mathbf{H}_{ij}^{(ac)}\right)^H\right) + \mathbf{I}_{M_i} \\ ={}& \tilde{\mathbf{\Sigma}}_i^{(a)} + \mathbf{I}_{M_i}. \end{aligned} \quad (7)$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2016.2606411, IEEE Transactions on Communications

For fixed $\boldsymbol{\lambda}$, the optimal transmit beamforming matrix is computed by taking the partial derivative of the function $\mathcal{J}(\boldsymbol{\lambda})$ with respect to $\mathbf{V}_i^{(b)}$, and is written as

$$\bar{\mathbf{V}}_i^{(b)}(\boldsymbol{\lambda}) = \sqrt{\rho_i}\lambda_i^{(a)}\left(\mathbf{I}_{N_i} + \mathbf{X}_i^{(b)}\right)^{-1}\left(\mathbf{R}_i^{(a)}\mathbf{H}_{ii}^{(ab)}\right)^H \mathbf{W}_i^{(a)}, \quad (20)$$

where $\mathbf{X}_i^{(b)}$ is shown in (21) at the bottom of the following page. Since the problem (13)-(14) is convex under fixed $\mathbf{R}, \mathbf{W}$, the optimal solution of the transmit beamforming is given by $\bar{\mathbf{V}}_i^{(b)}(\boldsymbol{\lambda}^*)$, where $\boldsymbol{\lambda}^*$ is the solution of the dual problem

$$\max_{\boldsymbol{\lambda}} \quad \mathcal{J}(\boldsymbol{\lambda}), \text{ s.t.} \quad \boldsymbol{\lambda} \geq \mathbf{0}. \quad (23)$$

Since the dual problem $\mathcal{J}(\boldsymbol{\lambda})$ in (23) is concave, a standard subgradient algorithm can be used to solve the dual problem in (23). The subgradient of $\mathcal{J}(\boldsymbol{\lambda})$ is given as

$$\text{tr}\left\{\mathbf{W}_i^{(a)}\mathbf{MSE}_i^{(a)}(\boldsymbol{\lambda})\right\} - \log_2\left|\ln 2\mathbf{W}_i^{(a)}\right| - \frac{d_i}{\ln 2} + R_i^{(a)}, \ i = 1, \ldots, K, \ a = 1, 2, \quad (24)$$

where $\mathbf{MSE}_i^{(a)}(\boldsymbol{\lambda})$ is obtained from (15) by replacing $\mathbf{V}_i^{(b)}$ with $\bar{\mathbf{V}}_i^{(b)}(\boldsymbol{\lambda})$.

A. Remarks

1) *Convergence:* The iterative alternating algorithm for solving the sum-power minimization problem (9)-(10) through the minimization problem (13)-(14) is given in Algorithm 1. Since the problem (13)-(14) is not jointly convex over the 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 [28], several reasonable choices such as right singular matrices, random matrices and interference alignment (IA) initialization have been proposed. It follows from the general optimization theory [29] that the proposed algorithm, which is the block coordinate descent method applied to (13)-(14), converges to a stationary point of (13)-(14). It remains to verify that $\bar{\mathbf{V}}$ is a stationary point of (9)-(10) if and only if $\bar{\mathbf{V}}, \bar{\mathbf{R}}, \bar{\mathbf{W}}$ be a stationary point of (13)-(14), which can be easily proved by following a similar analysis as in [16, Appendix C] and [14].

2) *Implementation:* At each time slot, the receivers feedback their direct and cross links to the central node, and based on the global CSI obtained from all the receivers, the central node first performs the precoding optimization to compute the optimal $\bar{\mathbf{V}}$, and then distributes the optimal transmit beamforming matrices to the transmitters for MIMO transmission. In an interference channel (or an ad-hoc network), the central node can reside at any node in the network. In a dynamic environment, the scheduler can be adaptively elected among the eligible nodes in the network [30]-[32]. The election can be done based on the capacity of a node, the status of a node, and the location of a node, etc. The research of the scheduler election issues is important but beyond the scope of this paper. We assume that a scheduler is available for the network within the time scale of interest. As will be detailed in Section IV-A, the complexity of the proposed precoding optimization algorithm is polynomial with respect to the number of users and antennas at each node.

Algorithm 1 Sum-Power Minimization Algorithm.

1: Set the iteration number $n = 0$ and initialize the transmit beamforming matrices $\mathbf{V}_i^{(a),[0]}, \ \forall (i,a)$.
2: **repeat**
3: $n \leftarrow n + 1$.
4: Update the receive filter $\mathbf{R}_i^{(a),[n]}, \ \forall (i,a)$ using (16).
5: Update the weighting matrix $\mathbf{W}_i^{(a),[n]}, \ \forall (i,a)$ using (17).
6: Update the optimal transmit beamforming matrix $\mathbf{V}_i^{(b),[n]} = \bar{\mathbf{V}}_i^{(b),[n]}(\boldsymbol{\lambda}^*), \ \forall (i,b)$ using (20), where λ^* is the optimal solution of (23).
7: **until** convergence or maximum number of iterations is reached.

B. Distributed Algorithm

The centralized algorithm proposed in the previous section requires a central node to aggregate all CSI and perform the optimization, which would incur heavy signaling overhead and limit the network scalability. Therefore, we propose a distributed algorithm that lowers the communication overhead and explicit signaling mechanisms. Defining $\mathbf{Q}_i^{(a)} = \mathbf{V}_i^{(a)}\left(\mathbf{V}_i^{(a)}\right)^H, \ \forall (i,a)$ as the source-covariance matrix at the user $i^{(a)}$, the achievable rate of the node $i^{(a)}$ in (8) can be rewritten as

$$I_i^{(a)}(\mathbf{Q}) = \log_2 \frac{\left|\mathbf{I}_{M_i} + \mathbf{H}_{ii}^{(ab)}\mathbf{Q}_i^{(b)}\left(\mathbf{H}_{ii}^{(ab)}\right)^H + \tilde{\mathbf{\Sigma}}_i^{(a)}\right|}{\left|\tilde{\mathbf{\Sigma}}_i^{(a)} + \mathbf{I}_{M_i}\right|}, \quad (25)$$

where the interference covariance matrix $\tilde{\mathbf{\Sigma}}_i^{(a)}$ is defined in (7), and $\mathbf{Q} = \left\{\mathbf{Q}_i^{(a)} : \ \forall (i,a)\right\}$ is the set of all source-covariance matrices.

To facilitate the distributed algorithm design, we first transform the sum power minimization problem (9)-(10) into the following equivalent form:

$$\min_{\mathbf{Q}, \tilde{\mathbf{\Sigma}}} \quad \sum_{i=1}^{K}\sum_{a=1}^{2} \text{tr}\left\{\mathbf{Q}_i^{(a)}\right\} \quad (26)$$

$$\text{s.t.} \quad I_i^{(a)}(\mathbf{Q}) \geq R_i^{(a)}, \ \forall i \in \mathcal{K}, \ a = 1, 2, \quad (27)$$

$$\mathcal{R}_i^{(a)} = \tilde{\mathbf{\Sigma}}_i^{(a)}, \ \forall i \in \mathcal{K}, \ a = 1, 2, \quad (28)$$

$$\mathbf{Q}_i^{(a)} \succeq \mathbf{0}, \ \forall i \in \mathcal{K}, \ a = 1, 2, \quad (29)$$

where $\tilde{\mathbf{\Sigma}} = \left\{\tilde{\mathbf{\Sigma}}_i^{(a)} : \ \forall (i,a)\right\}$ is the set of all interference-covariance matrices. The introduction of the auxiliary variable $\mathcal{R}_i^{(a)}$ helps to decompose the joint optimization problem into subproblems and facilitates the distributed implementation of the algorithm, as will be shown later.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2016.2606411, IEEE Transactions on Communications

Due to the non-convex rate constraint (27), the problem (26)-(29) is non-convex. Here we apply successive convex approximation (SCA) adopted in [33]-[34] to approximate (27) by a series of convex problems, so that a suboptimal solution of the problem (26)-(29) can be obtained by successively solving these convex problems. Due to the concavity of the $\log\det$ function, the first-order approximation of the $\log\det$ function at $\mathbf{X}_0$ is written as $\log|\mathbf{I}+\mathbf{X}| \leq \log|\mathbf{I}+\mathbf{X}_0| + \mathrm{tr}\left\{(\mathbf{I}+\mathbf{X}_0)^{-1}(\mathbf{X}-\mathbf{X}_0)\right\}$, which is true for $\mathbf{X} \succeq \mathbf{0}$. Let $\mathbf{Q}_i^{(a),[n-1]}$, $\tilde{\mathbf{\Sigma}}_i^{(a),[n-1]}$, $\forall\,(i,a)$ denote the feasible point of the problem (26)-(29) at the $(n-1)$th iteration, applying the first-order approximation with $\mathbf{X} = \tilde{\mathbf{\Sigma}}_i^{(a)}$, and $\mathbf{X}_0 = \tilde{\mathbf{\Sigma}}_i^{(a),[n-1]}$, the constraint (27) can be written as

$$
\begin{aligned}
&\log_2 \frac{\left|\mathbf{I}_{M_i} + \mathbf{H}_{ii}^{(ab)}\mathbf{Q}_i^{(b)}\left(\mathbf{H}_{ii}^{(ab)}\right)^H + \tilde{\mathbf{\Sigma}}_i^{(a)}\right|}{\left|\tilde{\mathbf{\Sigma}}_i^{(a)} + \mathbf{I}_{M_i}\right|} \\
&= \left(\log_2\left|\mathbf{I}_{M_i} + \mathbf{H}_{ii}^{(ab)}\mathbf{Q}_i^{(b)}\left(\mathbf{H}_{ii}^{(ab)}\right)^H + \tilde{\mathbf{\Sigma}}_i^{(a)}\right| \right.\\
&\qquad \left. - \log_2\left|\tilde{\mathbf{\Sigma}}_i^{(a)} + \mathbf{I}_{M_i}\right|\right) \\
&\leq \left(\log_2\left|\mathbf{I}_{M_i} + \mathbf{H}_{ii}^{(ab)}\mathbf{Q}_i^{(b)}\left(\mathbf{H}_{ii}^{(ab)}\right)^H + \tilde{\mathbf{\Sigma}}_i^{(a)}\right| \right.\\
&\quad - \log_2\left|\mathbf{I}_{M_i} + \tilde{\mathbf{\Sigma}}_i^{(a),[n-1]}\right| \\
&\quad \left. - \mathrm{tr}\left\{\left(\mathbf{I}_{M_i} + \tilde{\mathbf{\Sigma}}_i^{(a),[n-1]}\right)^{-1}\left(\tilde{\mathbf{\Sigma}}_i^{(a)} - \tilde{\mathbf{\Sigma}}_i^{(a),[n-1]}\right)\right\}\right) \\
&\geq R_i^{(a)},\ \forall i \in \mathcal{K},\ a = 1,2. \qquad (30)
\end{aligned}
$$

With this approximation, at the nth iteration, the following problem is solved

$$
\begin{aligned}
\min_{\mathbf{Q},\tilde{\mathbf{\Sigma}}} \quad & \sum_{i=1}^{K}\sum_{a=1}^{2}\mathrm{tr}\left\{\mathbf{Q}_i^{(a)}\right\} && (31)\\
\text{s.t.} \quad & (28),\ (29),\ (30). && (32)
\end{aligned}
$$

Since the primal problem in (31) is not strictly convex, the dual problem may not be differentiable at every point. To tackle the difficulty due to the lack of strict convexity of (31), we use the proximal optimization algorithm [35, pp. 232], in which quadratic terms are added to the objective function (31), so that we can use a duality based approach. The resulting regularized problem is given by

$$
\begin{aligned}
\min_{\mathbf{Q},\tilde{\mathbf{\Sigma}},\mathbf{M},\mathbf{N}} \quad & \sum_{i=1}^{K}\sum_{a=1}^{2}\left(\mathrm{tr}\left\{\mathbf{Q}_i^{(a)}\right\} + \frac{c_i^{(a)}}{2}\left(\left\|\mathbf{Q}_i^{(a)} - \mathbf{M}_i^{(a)}\right\|_F^2 \right.\right. \\
& \qquad \left.\left. + \left\|\tilde{\mathbf{\Sigma}}_i^{(a)} - \mathbf{N}_i^{(a)}\right\|_F^2\right)\right) && (33)\\
\text{s.t.} \quad & (28),\ (29),\ (30), && (34)
\end{aligned}
$$

where $\mathbf{M}_i^{(a)}$ and $\mathbf{N}_i^{(a)}$ are the auxiliary variables associated with the original variables $\mathbf{Q}_i^{(a)}$ and $\tilde{\mathbf{\Sigma}}_i^{(a)}$, respectively, and $c_i^{(a)}/2$ is the weight of those quadratic terms. Here, the newly added optimization variables are defined as $\mathbf{M}(\mathbf{N}) = \left\{\mathbf{M}_i^{(a)}\left(\mathbf{N}_i^{(a)}\right) :\ \forall(i,a)\right\}$.

It can be easily verified that the optimal solution of (33)-(34) is also optimal to (31)-(32). Since (33) is strictly convex, dual decomposition [36] can be applied to deal with the coupled constraints (28). The partial Lagrangian of (33) is given by

$$
\begin{aligned}
&\mathcal{L}\left(\mathbf{Q},\tilde{\mathbf{\Sigma}},\mathbf{M},\mathbf{N},\mathbf{Z}\right) && (35)\\
&= \sum_{i=1}^{K}\sum_{a=1}^{2}\mathcal{L}_i^{(a)}\left(\mathbf{Q}_i^{(a)},\tilde{\mathbf{\Sigma}}_i^{(a)},\mathbf{M}_i^{(a)},\mathbf{N}_i^{(a)},\mathbf{Z}\right),
\end{aligned}
$$

where $\mathbf{Z}_i^{(a)}$ is the Lagrangian multiplier associated with the constraint (28), and $\mathbf{Z} = \left\{\mathbf{Z}_i^{(a)} :\ \forall(i,a)\right\}$. Here $\mathcal{L}_i^{(a)}(\cdot)$ is defined as

$$
\begin{aligned}
&\mathcal{L}_i^{(a)}\left(\mathbf{Q}_i^{(a)},\tilde{\mathbf{\Sigma}}_i^{(a)},\mathbf{M}_i^{(a)},\mathbf{N}_i^{(a)},\mathbf{Z}\right) \\
&= \mathrm{tr}\left\{\mathbf{Q}_i^{(a)}\right\} - \mathrm{tr}\left\{\left(\mathbf{Z}_i^{(a)}\right)^H\tilde{\mathbf{\Sigma}}_i^{(a)}\right\} \\
&+ \sum_{l\neq i}^{K}\sum_{m=1}^{2}\mathrm{tr}\left\{\left(\mathbf{Z}_l^{(m)}\right)^H\mathbf{H}_{li}^{(ma)}\mathbf{Q}_i^{(a)}\left(\mathbf{H}_{li}^{(ma)}\right)^H\right\} \\
&+ \sum_{l=1}^{K}\sum_{m=1}^{2}\mathrm{tr}\left\{\left(\mathbf{Z}_l^{(m)}\right)^H\left[\kappa\mathbf{H}_{li}^{(ma)}\mathrm{diag}\left(\mathbf{Q}_i^{(a)}\right)\right.\right. \\
&\left.\left.\times\left(\mathbf{H}_{li}^{(ma)}\right)^H + \beta\,\mathrm{diag}\left(\mathbf{H}_{li}^{(ma)}\mathbf{Q}_i^{(a)}\left(\mathbf{H}_{li}^{(ma)}\right)^H\right)\right]\right\} \\
&+ \frac{c_i^{(a)}}{2}\left(\left\|\mathbf{Q}_i^{(a)} - \mathbf{M}_i^{(a)}\right\|_F^2 + \left\|\tilde{\mathbf{\Sigma}}_i^{(a)} - \mathbf{N}_i^{(a)}\right\|_F^2\right). \qquad (36)
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{X}_i^{(b)} &= \lambda_i^{(a)}\rho_i\left(\left(\mathbf{H}_{ii}^{(ab)}\right)^H\mathbf{A}_i^{(a)}\mathbf{H}_{ii}^{(ab)} + \kappa\mathrm{diag}\left(\left(\mathbf{H}_{ii}^{(ab)}\right)^H\mathbf{A}_i^{(a)}\mathbf{H}_{ii}^{(ab)}\right) + \beta\left(\mathbf{H}_{ii}^{(ab)}\right)^H\mathrm{diag}\left(\mathbf{A}_i^{(a)}\right)\mathbf{H}_{ii}^{(ab)}\right) \\
&\quad + \lambda_i^{(b)}\eta_{ii}\left(\kappa\mathrm{diag}\left(\left(\mathbf{H}_{ii}^{(bb)}\right)^H\mathbf{A}_i^{(b)}\mathbf{H}_{ii}^{(bb)}\right) + \beta\left(\mathbf{H}_{ii}^{(bb)}\right)^H\mathrm{diag}\left(\mathbf{A}_i^{(b)}\right)\mathbf{H}_{ii}^{(bb)}\right) \\
&\quad + \sum_{j\neq i}^{K}\sum_{c=1}^{2}\lambda_j^{(c)}\eta_{ji}\left(\left(\mathbf{H}_{ji}^{(cb)}\right)^H\mathbf{A}_j^{(c)}\mathbf{H}_{ji}^{(cb)} + \kappa\mathrm{diag}\left(\left(\mathbf{H}_{ji}^{(cb)}\right)^H\mathbf{A}_j^{(c)}\mathbf{H}_{ji}^{(cb)}\right) + \beta\left(\mathbf{H}_{ji}^{(cb)}\right)^H\mathrm{diag}\left(\mathbf{A}_j^{(c)}\right)\mathbf{H}_{ji}^{(cb)}\right), \qquad (21)
\end{aligned}
$$

$$
\mathbf{A}_i^{(b)} = \left(\mathbf{R}_i^{(b)}\right)^H\mathbf{W}_i^{(b)}\mathbf{R}_i^{(b)},\ i = 1,\ldots,K,\ b = 1,2. \qquad (22)
$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2016.2606411, IEEE Transactions on Communications

With the above decomposition, an iterative dual-layer distributed algorithm can be derived to solve (33)-(34) based on the standard proximal point method [27]. In the inner layer, the original variables $\mathbf{Q}_i^{(a)}$ and $\tilde{\mathbf{\Sigma}}_i^{(a)}$, $\forall\,(i,a)$ are optimized mainly by iteratively solving the following problem (with the auxiliary variables temporarily fixed) using the subgradient method [36]

$$\begin{aligned}\max_{\mathbf{Z}} \quad & \min_{\{\mathbf{Q},\tilde{\mathbf{\Sigma}}\}\in\mathcal{D}} \ \mathcal{L}\left(\mathbf{Q},\tilde{\mathbf{\Sigma}},\mathbf{M},\mathbf{N},\mathbf{Z}\right),\\ \triangleq \max_{\mathbf{Z}} \quad & \sum_{i=1}^{K}\sum_{a=1}^{2} \min_{\left\{\mathbf{Q}_i^{(a)},\tilde{\mathbf{\Sigma}}_i^{(a)}\right\}\in\mathcal{D}_i^{(a)}} \\ & \qquad \mathcal{L}_i^{(a)}\left(\mathbf{Q}_i^{(a)},\tilde{\mathbf{\Sigma}}_i^{(a)},\mathbf{M}_i^{(a)},\mathbf{N}_i^{(a)},\mathbf{Z}\right), \quad (37)\end{aligned}$$

where $\mathcal{D}_i^{(a)}$ is the convex constraint set for $\mathbf{Q}_i^{(a)}$ and $\tilde{\mathbf{\Sigma}}_i^{(a)}$ defined by the uncoupled constraints (29) and (30). Here $\mathcal{D} = \mathcal{D}_1^{(1)} \times \mathcal{D}_1^{(2)} \times \ldots \times \mathcal{D}_K^{(1)} \times \mathcal{D}_K^{(2)}$. Note that the uncoupled constraints in (37) are due to the introduction of the auxiliary variable $\mathcal{R}_i^{(a)}$ in (28). In each outer iteration, the inner minimization in (37) is solved to obtain the optimal solution for the original variables. Then with the obtained original variables, in the inner iterations, the dual variables $\mathbf{Z}_i^{(a)}$ is iteratively updated by the subgradient method with the subgradient $\mathcal{R}_i^{(a)} - \tilde{\mathbf{\Sigma}}_i^{(a)}$ (cf. (28)) until their convergence. The algorithm ends when the auxiliary variables converge.

Note that in the proposed algorithm all computations can be carried out based on local information, and hence can be easily distributed. More precisely, in the inner minimization problem in (37) (the dual objective function), we have decomposed the original problem into $2K$ separate subproblems for each user $i = 1,\ldots,K,\ a = 1,2$. Given $\mathbf{Z}$, each subproblem can now be solved independently.

IV. Energy-Efficiency Maximization

While sum-rate maximization is one of the most popular design criteria in wireless communication systems, the energy-efficiency maximization is another design criteria that has drawn much attention recently due to increasing interest in green wireless networks. Thus, it is of importance to study the energy-efficiency of the FD MIMO interference channels.

The energy-efficiency metric is defined as the ratio of weighted sum rate to the total power consumption, given by

$$\begin{aligned} EE\left(\mathbf{V}\right) &= \frac{f_1\left(\mathbf{V}\right)}{f_2\left(\mathbf{V}\right)} \quad (38)\\ &= \frac{\sum_{i=1}^{K}\sum_{a=1}^{2}\mu_i^{(a)} I_i^{(a)}\left(\mathbf{V}\right)}{\sum_{i=1}^{K}\sum_{a=1}^{2}\left(\vartheta \mathrm{tr}\left\{\mathbf{V}_i^{(a)}\left(\mathbf{V}_i^{(a)}\right)^H\right\} + N_i P_c + P_0\right)}, \end{aligned}$$

where $\mu_i^{(a)}$ is the weight used to represent the priority of node $i^{(a)}$ in the system, $\vartheta \geq 1$ is the power-amplifier inefficiency which depends on the design and implementation of the power amplifier, P_c is the dynamic circuit power consumption per antenna corresponding to the power radiation of all circuit blocks in the transmit filter, i.e., mixer, frequency synthesizer, and DAC; and P_0 is the static circuit power consumed at the transmitter's cooling system, power supply, etc., which is independent of the number of transmit antennas. This linear power model is adopted from [37], in which not only the data-dependent transmit power, but also the circuit power consumption in all components used for signal processing such as mixer, filter, ADC, DAC, low-noise amplifier (LNA), etc. also plays an important role in energy-efficiency performance.

The energy-efficiency maximization problem to compute the optimal transmit beamforming matrices can be formulated as

$$\begin{aligned} \max_{\mathbf{V}} \quad & EE\left(\mathbf{V}\right) \quad (39)\\ \text{s.t.} \quad & \mathrm{tr}\left\{\mathbf{V}_i^{(a)}\left(\mathbf{V}_i^{(a)}\right)^H\right\} \leq P_i^{(a)},\ \forall (i,a), \quad (40)\end{aligned}$$

where $P_i^{(a)}$ is the power constraint at node $i^{(a)}$, $i = 1,\ldots,K,\ a = 1,2$. The problem (39)-(40) is non-convex due to the coupling of optimization variables and the fractional form of the objective function (39). Therefore convex optimization tools cannot be applied to solve this challenging problem. To this end, similar to [17], we will first transform the original fractional problem into an equivalent non-fractional problem by exploiting the relationship between fractional and parametric programming problems [20]-[21], and then the equivalent non-fractional problem is solved by exploiting the relationship between WSR and WMMSE. We will develop a two-layer approach to solve these two steps.

Theorem 1: The optimal precoding matrix $\bar{\mathbf{V}}$ achieves the maximum energy-efficiency q^*, defined as $q^* = \frac{f_1(\bar{\mathbf{V}})}{f_2(\bar{\mathbf{V}})} = \max_{\mathbf{V}} \frac{f_1(\mathbf{V})}{f_2(\mathbf{V})}$, if and only if

$$\begin{aligned} \max_{\mathbf{V}} \quad & f_1\left(\mathbf{V}\right) - q^* f_2\left(\mathbf{V}\right)\\ = \quad & f_1\left(\bar{\mathbf{V}}\right) - q^* f_2\left(\bar{\mathbf{V}}\right) = 0, \quad (41)\end{aligned}$$

for $f_1\left(\mathbf{V}\right) \geq 0$ and $f_2\left(\mathbf{V}\right) > 0$.

Proof: Theorem 1 can be proved by following a similar approach as in [17], [20]. The optimal energy-efficiency can be expressed as

$$\begin{aligned} q^* \quad = \quad & \frac{f_1\left(\bar{\mathbf{V}}\right)}{f_2\left(\bar{\mathbf{V}}\right)} \geq \frac{f_1\left(\mathbf{V}\right)}{f_2\left(\mathbf{V}\right)}\\ \Longrightarrow \quad & f_1\left(\mathbf{V}\right) - q^* f_2\left(\mathbf{V}\right) \leq 0, \ \text{ and}\\ & f_1\left(\bar{\mathbf{V}}\right) - q^* f_2\left(\bar{\mathbf{V}}\right) = 0.\end{aligned}$$

Therefore, we conclude that $\max_{\mathbf{V}}\ f_1\left(\mathbf{V}\right) - q^* f_2\left(\mathbf{V}\right) = 0$ and it is achievable by the precoding matrix $\bar{\mathbf{V}}$. The converse can be proved by reversing the steps of the proof. ■

From Theorem 1 we can conclude that for any objective function in fractional form, there exists an equivalent objective function in subtractive form, which shares the same objective and constraint values. As a result, to find the optimal precoding matrix with a given q, the optimization problem (39)-(40) in fractional form can be solved by focusing on the problem

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2016.2606411, IEEE Transactions on Communications

below, which is in a tractable subtractive form.

$$\max_{\mathbf{V}} \quad \sum_{i=1}^{K}\sum_{a=1}^{2}\mu_i^{(a)} I_i^{(a)}\left(\mathbf{V}\right) - q f_2\left(\mathbf{V}\right) \tag{42}$$

$$\text{s.t.} \quad \text{tr}\left\{\mathbf{V}_i^{(a)}\left(\mathbf{V}_i^{(a)}\right)^H\right\} \le P_i^{(a)},\ \forall\left(i,a\right). \tag{43}$$

It can be considered from Theorem 1 that the solution of the problem (39)-(40) is also optimal if we can find a parameter q such that the optimal value of problem (42)-(43) is zero. The value of q in (42) can be treated as the penalty to the energy-efficiency due to exceedingly high power consumption. Assume that there is no penalty in using exceedingly high power, i.e., $q = 0$, then the transformed objective function reduces to the weighted sum-rate maximization problem.

Although we have tackled the fractional form problem of the original energy-efficiency problem, the non-fractional form problem (42)-(43) is still non-convex, since its optimization variables are coupled. To tackle the coupling problem of (42)-(43), we exploit the relationship between WSR and WMMSE [15], [16] to further reformulate it into a tractable form. Introducing the receive filters $\mathbf{R}$ in (16) and MSE-weights $\mathbf{W}$ in (17), the problem (42)-(43) can be reformulated as

$$\max_{\mathbf{V},\mathbf{R},\mathbf{W}} \quad \bar{f}_1\left(\mathbf{V},\mathbf{R},\mathbf{W}\right) - q f_2\left(\mathbf{V}\right) \tag{44}$$

$$\text{s.t.} \quad \text{tr}\left\{\mathbf{V}_i^{(a)}\left(\mathbf{V}_i^{(a)}\right)^H\right\} \le P_i^{(a)},\ \ \forall\left(i,a\right), \tag{45}$$

where $\bar{f}_1\left(\mathbf{V},\mathbf{R},\mathbf{W}\right)$ is expressed as

$$\bar{f}_1\left(\mathbf{V},\mathbf{R},\mathbf{W}\right) = \sum_{i=1}^{K}\sum_{a=1}^{2}\left(-\text{tr}\left\{\mathbf{W}_i^{(a)}\mathbf{MSE}_i^{(a)}\right\} + \mu_i^{(a)}\log_2\left|\frac{\ln 2}{\mu_i^{(a)}}\mathbf{W}_i^{(a)}\right| + \frac{d_i\mu_i^{(a)}}{\ln 2}\right). \tag{46}$$

Proposition 2: The problems (42)-(43) and (44)-(45) share the same optimal transmit beamforming matrix $\bar{\mathbf{V}}$.

Proof: Substituting the optimal receive beamforming matrices $\bar{\mathbf{R}}$ in (16) and MSE-weights $\mathbf{W}$ in (47) in the objective function (44), and using the relation $I_i^{(a)}\left(\mathbf{V}\right) = \log_2\left|\left(\mathbf{E}_i^{(a)}\left(\mathbf{V}\right)\right)^{-1}\right|$ in (11), the objective function (42) is obtained. ∎

Since the problem (44)-(45) is tractable and convex in each of the optimization variable $\mathbf{V},\mathbf{R},\mathbf{W}$, we can apply the block coordinate ascent method to solve the problem (44)-(45). Under fixed $\mathbf{V},\mathbf{W}$, the optimal receive filters $\bar{\mathbf{R}}$ is given in (16), and the optimal MSE-weights under fixed $\mathbf{V},\mathbf{R}$ is given as

$$\bar{\mathbf{W}}_i^{(a)} = \frac{\mu_i^{(a)}}{\ln 2}\left(\mathbf{E}_i^{(a)}\left(\mathbf{V}\right)\right)^{-1}, \tag{47}$$

where $\mathbf{E}_i^{(a)}\left(\mathbf{V}\right)$ is given in (12). Under fixed $\mathbf{R},\mathbf{W}$, the problem to solve the optimal $\bar{\mathbf{V}}$ is expressed as

$$\max_{\mathbf{V}} \quad \sum_{i=1}^{K}\sum_{a=1}^{2}\left(-\text{tr}\left\{\mathbf{W}_i^{(a)}\mathbf{MSE}_i^{(a)}\right\} - q\vartheta\text{tr}\left\{\mathbf{V}_i^{(a)}\left(\mathbf{V}_i^{(a)}\right)^H\right\}\right) \tag{48}$$

$$\text{s.t.} \quad \text{tr}\left\{\mathbf{V}_i^{(a)}\left(\mathbf{V}_i^{(a)}\right)^H\right\} \le P_i^{(a)},\ \forall(i,a). \tag{49}$$

The Lagrange function of the problem (48)-(49) can be written as

$$\mathcal{L}\left(\mathbf{V},\boldsymbol{\lambda}\right) = \sum_{i=1}^{K}\sum_{a=1}^{2}\left(-\text{tr}\left\{\mathbf{W}_i^{(a)}\mathbf{MSE}_i^{(a)}\right\} - q\vartheta\text{tr}\left\{\mathbf{V}_i^{(a)}\left(\mathbf{V}_i^{(a)}\right)^H\right\}\right) - \sum_{i=1}^{K}\sum_{a=1}^{2}\lambda_i^{(a)}\left(\text{tr}\left\{\mathbf{V}_i^{(a)}\left(\mathbf{V}_i^{(a)}\right)^H\right\} - P_i^{(a)}\right), \tag{50}$$

where $\lambda_i^{(a)}$ is the Lagrange multiplier associated with the power constraint of transmitter $i^{(a)}$. For fixed $\boldsymbol{\lambda}$, by taking the partial derivative of $\mathcal{L}\left(\mathbf{V},\boldsymbol{\lambda}\right)$ with respect to $\mathbf{V}_i^{(b)}$, the closed-form solution is expressed as

$$\bar{\mathbf{V}}_i^{(b)}\left(\lambda_i^{(b)}\right) = \sqrt{\rho_i}\left(\lambda_i^{(b)}\mathbf{I}_{N_i} + q\vartheta\mathbf{I}_{N_i} + \mathbf{Y}_i^{(b)}\right)^{-1} \times \left(\mathbf{R}_i^{(a)}\mathbf{H}_{ii}^{(ab)}\right)^H\mathbf{W}_i^{(a)}, \tag{51}$$

where $\mathbf{Y}_i^{(b)}$ is shown in (52) at the bottom of the following page.

The values of the Lagrange multiplier $\lambda_i^{(b)}$ in (51) are calculated by taking the eigenvalue decomposition of $q\vartheta\mathbf{I}_{N_i} + \mathbf{Y}_i^{(b)} = \mathbf{U}_i^{(b)}\boldsymbol{\Delta}_i^{(b)}\left(\mathbf{U}_i^{(b)}\right)^H$ and writing the power constraint in (49), after simple steps, as

$$\text{tr}\left\{\bar{\mathbf{V}}_i^{(b)}\left(\lambda_i^{(b)}\right)\left(\bar{\mathbf{V}}_i^{(b)}\left(\lambda_i^{(b)}\right)\right)^H\right\} = \rho_i\sum_{k=1}^{N_i}\frac{g_{ik}^{(b)}}{\left(\lambda_i^{(b)} + \Delta_{ik}^{(b)}\right)^2} = P_i^{(b)}, \tag{53}$$

where $g_{ik}^{(b)}$ denotes the kth row and kth column element of $\left(\mathbf{U}_i^{(b)}\right)^H\left(\mathbf{H}_{ii}^{(ab)}\right)^H\left(\mathbf{R}_i^{(a)}\right)^H\mathbf{W}_i^{(a)}\left(\mathbf{W}_i^{(a)}\right)^H\mathbf{R}_i^{(a)}\mathbf{H}_{ii}^{(ab)}\mathbf{U}_i^{(b)}$ and $\Delta_{ik}^{(b)}$ denotes the kth row and kth column element of the matrix $\boldsymbol{\Delta}_i^{(b)}$. We can compute $\lambda_i^{(b)}$ from (53) numerically. If the values of the Lagrange multipliers $\lambda_i^{(b)}$ are negative, we assign $\lambda_i^{(b)}$ as zeros.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2016.2606411, IEEE Transactions on Communications

Algorithm 2 Energy-Efficiency Maximization Algorithm.

1: Set the energy-efficiency factor $q = 0$, and initialize all the transmit beamforming matrices $\mathbf{V}_i^{(a)}$, $\forall (i,a)$ such that $\text{tr}\left\{\mathbf{V}_i^{(a)}\left(\mathbf{V}_i^{(a)}\right)^H\right\} \leq P_i^{(a)}$.
2: **repeat**
3: Update the receive filter $\mathbf{R}_i^{(a)}$, $\forall (i,a)$ using (16).
4: Update the weighting matrix $\mathbf{W}_i^{(a)}$, $\forall (i,a)$ using (47).
5: Update the transmit beamforming $\mathbf{V}_i^{(b)}$ using (51).
6: **until** convergence or maximum number of iterations is reached.
7: **if** $\bar{f}_1(\mathbf{V},\mathbf{R},\mathbf{W}) - q f_2(\mathbf{V}) \leq \epsilon$ **then**
8: Stop the iterations
9: **else**
10: $q = \frac{\bar{f}_1(\mathbf{V},\mathbf{R},\mathbf{W})}{f_2(\mathbf{V})}$, and go to Step 3.
11: **end if**

A. Remarks

1) *Convergence:* The iterative alternating algorithm for solving the energy maximization problem (39)-(40) through the maximization problem (44)-(45) is given in Algorithm 2. Since the updates at step 3, step 4, and step 5 all maximize the objective function at each iteration, the iterations in Algorithm 2 lead to monotone increase of the objective function (44). Since the objective function under the practical power constraints is bounded, the convergence of the alternating maximization algorithm can be guaranteed with the monotonic convergence theorem [38]. Since the objective function (44) is differentiable, it follows from the general optimization theory [29], [38] that a block coordinate ascent method converges to a stationary point of problem (44)-(45). Based on these analyses, the convergence of the Algorithm 2 can be guaranteed with the fractional theorem obtained in [20].

2) *Complexity:* We count the floating point operations (flops) required per iteration, where both a scalar complex multiplication and a scalar complex addition count as one flop [39]. The complexity of some basic matrix calculations are approximately counted as follows: Multiplication of two $M \times N$ and $N \times L$ matrices involves $2MNL - ML$, multiplication of $\mathbf{A}^H\mathbf{A}$ where $\mathbf{A} \in \mathbb{C}^{M\times N}$ involves $MN^2 + N\left(M - \frac{N+1}{2}\right)$, and inverse of a $N \times N$ matrix involves $N^3 + N^2 + N$ flops. Based on these references, the computation required for each step in Algorithm 2 is as follows:

The computation of the receive beamforming matrix in (16):

- The terms inside the inverse:

$$2\sum_{j=1}^{K} N_j \left(d_j N_j + d_j - \frac{N_j+1}{2} + 4N_jM_i + 4M_i^2 - 2M_i\right) - 2M_i^2.$$

- The inverse function: $M_i^3 + M_i^2 + M_i$.
- The matrix multiplications outside of the inverse:

$$2d_iM_i\left(N_i + M_i - 1\right).$$

The computation of the weighting matrix in (47):

- For the inverse of $\mathbf{\Sigma}_i^{(a)}$: $M_i^3 + M_i^2 + M_i$.
- The matrix multiplications in $\mathbf{E}_i^{(a)}(\mathbf{V})$ given in (12):

$$d_i\left(2M_i^2 - M_i + 2M_id_i - d_i\right).$$

The computation of the optimal transmit beamforming matrix in (51):

- For the computation of $\mathbf{Y}_i^{(b)}$:

$$2\sum_{j=1}^{K}\left(M_jd_j\left(2d_j + 2M_j - 1 - \frac{M_j}{d_j}\right) + 2N_iM_j\left(2M_j + 2N_i - 1 - \frac{N_i}{M_j}\right)\right).$$

- For the inverse function: $N_i^3 + N_i^2 + N_i$.
- The matrix multiplications outside of the inverse:

$$N_i\left(2N_iM_i - M_i - 2d_i + 2M_id_i + 2d_i^2\right).$$

Assuming the same number of transmit and receive antennas at each node, i.e., $M = M_i$, $N = N_i$, $i = 1,\ldots,K$, using the flop computations above, the total computational complexity of the proposed algorithm is in the order of $\mathcal{O}\left(\phi_1\phi_2K^2\left(N^2M + M^2N + N^3\right) + KM^3\right)$, where ϕ_1 and ϕ_2 denote the number of iterations for the inner and outer loops in Algorithm 2, respectively. Note that the computational complexity of the sum-power minimization algorithm (Algorithm 1) has the same order of computational complexity as Algorithm 2, i.e., $\mathcal{O}\left(\phi_1K^2\left(N^2M + M^2N + N^3\right) + KM^3\right)$. Although Algorithm 2 may require a larger number of iterations than Algorithm 1, due to the additional search for the energy efficiency factor q, in our simulations we have observed that the search for the optimal q generally takes a few iterations, therefore it will not significantly increase the complexity.

$$\begin{aligned}\mathbf{Y}_i^{(b)} = {} & \rho_i\left(\mathbf{H}_{ii}^{(ab)}\right)^H\mathbf{A}_i^{(a)}\mathbf{H}_{ii}^{(ab)} + \rho_i\kappa\text{diag}\left(\left(\mathbf{H}_{ii}^{(ab)}\right)^H\mathbf{A}_i^{(a)}\mathbf{H}_{ii}^{(ab)}\right) + \rho_i\beta\left(\mathbf{H}_{ii}^{(ab)}\right)^H\text{diag}\left(\mathbf{A}_i^{(a)}\right)\mathbf{H}_{ii}^{(ab)} \\ & + \eta_{ii}\kappa\text{diag}\left(\left(\mathbf{H}_{ii}^{(bb)}\right)^H\mathbf{A}_i^{(b)}\mathbf{H}_{ii}^{(bb)}\right) + \eta_{ii}\beta\left(\mathbf{H}_{ii}^{(bb)}\right)^H\text{diag}\left(\mathbf{A}_i^{(b)}\right)\mathbf{H}_{ii}^{(bb)} \\ & + \sum_{j\neq i}^{K}\sum_{c=1}^{2}\eta_{ji}\left(\mathbf{H}_{ji}^{(cb)}\right)^H\mathbf{A}_j^{(c)}\mathbf{H}_{ji}^{(cb)} + \eta_{ji}\kappa\text{diag}\left(\left(\mathbf{H}_{ji}^{(cb)}\right)^H\mathbf{A}_j^{(c)}\mathbf{H}_{ji}^{(cb)}\right) + \eta_{ji}\beta\left(\mathbf{H}_{ji}^{(cb)}\right)^H\text{diag}\left(\mathbf{A}_j^{(c)}\right)\mathbf{H}_{ji}^{(cb)}. \quad (52)\end{aligned}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2016.2606411, IEEE Transactions on Communications

3) *Implementation:* The proposed algorithm can be implemented in a distributed manner, due to the fact that the updating of $\mathbf{V}, \mathbf{W}, \mathbf{R}$ all have independent closed-form expressions. In other words, our scheme requires only the *local* CSI (i.e., each transmitter needs to know only the CSI of the links originating from itself) [16], [40], [41]. 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 [40], [42]. Since only *local* CSI is required at each user, the proposed method fits for distributed implementation.

B. Robust Beamforming Design

Given the presence of channel uncertainty at the transmitters in practice, a more relevant and difficult problem of interest is a robust counterpart of (39)-(40) in the presence of bounded channel errors. In this section, the CSI of the channels are assumed to be imperfectly known. The imperfect CSI is modeled using deterministic norm-bounded error model [43]-[45], which is expressed as

$$\mathbf{H}_{ij}^{(ab)} \in \mathcal{H}_{ij}^{(ab)} = \left\{ \tilde{\mathbf{H}}_{ij}^{(ab)} + \mathbf{\Lambda}_i^{(a)} : \|\mathbf{\Lambda}_i^{(a)}\|_F \leq \delta_i^{(a)}, \right. \\ \left. j = 1, \ldots, K, \ b = 1, 2 \right\},$$

where $\tilde{\mathbf{H}}_{ij}^{(ab)}$, $\mathbf{\Lambda}_i^{(a)}$ and $\delta_i^{(a)} \geq 0$ denote the nominal value of the CSI known to the transmitters, the corresponding error matrix, and the uncertainty bounds, respectively.

With the imperfect CSI, the worst-case optimization problem under channel uncertainty can be formulated as

$$\max_{\mathbf{V}} \min_{\forall \mathbf{H}_{ij}^{(ab)} \in \mathcal{H}_{ij}^{(ab)}} \quad EE\left(\mathbf{V}\right) \tag{54}$$

$$\text{s.t.} \quad \text{tr}\left\{ \mathbf{V}_i^{(a)} \left(\mathbf{V}_i^{(a)} \right)^H \right\} \leq P_i^{(a)}, \ \forall (i,a). \tag{55}$$

By using the achievable rate and MSE relation used in the problem (44)-(45), the robust optimization problem can be written equivalently as

$$\max_{\mathbf{V}} \min_{\forall \mathbf{H}_{ij}^{(ab)} \in \mathcal{H}_{ij}^{(ab)}} \max_{\mathbf{R}, \mathbf{W}} \quad \frac{\bar{f}_1\left(\mathbf{V}, \mathbf{R}, \mathbf{W}\right)}{f_2\left(\mathbf{V}\right)} \tag{56}$$

$$\text{s.t.} \quad \text{tr}\left\{ \mathbf{V}_i^{(a)} \left(\mathbf{V}_i^{(a)} \right)^H \right\} \leq P_i^{(a)}, \ \forall (i,a), \tag{57}$$

where $\bar{f}_1\left(\mathbf{V}, \mathbf{R}, \mathbf{W}\right)$ is given in (46). Unfortunately, this formulation does not directly lead to a useful algorithm. Therefore, we look at the max-min version of the inner min-max problem in (56)-(57), which gives us the lower-bound. This formulation is written as

$$\max_{\mathbf{V}, \mathbf{R}, \mathbf{W}} \min_{\forall \mathbf{H}_{ij}^{(ab)} \in \mathcal{H}_{ij}^{(ab)}} \quad \frac{\bar{f}_1\left(\mathbf{V}, \mathbf{R}, \mathbf{W}\right)}{f_2\left(\mathbf{V}\right)} \tag{58}$$

$$\text{s.t.} \quad \text{tr}\left\{ \mathbf{V}_i^{(a)} \left(\mathbf{V}_i^{(a)} \right)^H \right\} \leq P_i^{(a)}, \ \forall (i,a). \tag{59}$$

With the use of Theorem 1, the problem (58)-(59) can be written as under fixed q

$$\max_{\mathbf{V}, \mathbf{R}, \mathbf{W}} \min_{\forall \mathbf{H}_{ij}^{(ab)} \in \mathcal{H}_{ij}^{(ab)}} \quad \bar{f}_1\left(\mathbf{V}, \mathbf{R}, \mathbf{W}\right) - q f_2\left(\mathbf{V}\right) \tag{60}$$

$$\text{s.t.} \quad \text{tr}\left\{ \mathbf{V}_i^{(a)} \left(\mathbf{V}_i^{(a)} \right)^H \right\} \leq P_i^{(a)}, \ \forall (i,a). \tag{61}$$

By plugging $\bar{f}_1\left(\mathbf{V}, \mathbf{R}, \mathbf{W}\right)$ in (46) into (60) and using epigraph form with the introduction of slack variables $\varsigma_i^{(a)}$, the problem (60)-(61) can be equivalently written as

$$\max_{\mathbf{V}, \mathbf{R}, \mathbf{W}, \boldsymbol{\varsigma}} \quad \sum_{i=1}^{K} \sum_{a=1}^{2} \left(-\varsigma_i^{(a)} + \mu_i^{(a)} \log_2 \left| \frac{\ln 2}{\mu_i^{(a)}} \mathbf{W}_i^{(a)} \right| + \frac{d_i \mu_i^{(a)}}{\ln 2} \right) \\ - q f_2\left(\mathbf{V}\right) \tag{62}$$

$$\text{s.t.} \quad \text{tr}\left\{ \mathbf{W}_i^{(a)} \mathbf{MSE}_i^{(a)} \right\} \leq \varsigma_i^{(a)}, \|\mathbf{\Lambda}_i^{(a)}\|_F \leq \delta_i^{(a)}, \forall (i,a) \tag{63}$$

$$\text{tr}\left\{ \mathbf{V}_i^{(a)} \left(\mathbf{V}_i^{(a)} \right)^H \right\} \leq P_i^{(a)}, \ \forall (i,a), \tag{64}$$

where $\boldsymbol{\varsigma} = \left\{ \varsigma_i^{(a)} : \forall \left(i,a\right) \right\}$.

To solve the optimization problem (62)-(64), we write $\text{tr}\left\{ \mathbf{W}_i^{(a)} \mathbf{MSE}_i^{(a)} \right\}$ in a vector form as $\text{tr}\left\{ \mathbf{W}_i^{(a)} \mathbf{MSE}_i^{(a)} \right\} \triangleq \left\| \mathbf{d}_i^{(a)} + \mathbf{D}_i^{(a)} vec\left(\mathbf{\Lambda}_i^{(a)} \right) \right\|_2^2$, where $\mathbf{d}_i^{(a)}$ and $\mathbf{D}_i^{(a)}$ are shown in (65) and (66), respectively given at the bottom of the following page[2]. In (65) and (66), $\mathbf{W}_i^{(a)}$ is decomposed as $\mathbf{W}_i^{(a)} = \mathbf{B}_i^{(a)} \left(\mathbf{B}_i^{(a)} \right)^H$. With the vector forms in hand, Schur complement lemma can be used to express the constraint $\left\| \mathbf{d}_i^{(a)} + \mathbf{D}_i^{(a)} vec\left(\mathbf{\Lambda}_i^{(a)} \right) \right\|_2^2 \leq \varsigma_i^{(a)}$ in (63) in linear matrix inequalities (LMI) form:

$$\begin{bmatrix} \varsigma_i^{(a)} & \left(\mathbf{d}_i^{(a)} \right)^H \\ \mathbf{d}_i^{(a)} & \mathbf{I}_{A_i^{(a)}} \end{bmatrix} \\ + \begin{bmatrix} 0 & vec\left(\mathbf{\Lambda}_i^{(a)} \right)^H \left(\mathbf{D}_i^{(a)} \right)^H \\ \mathbf{D}_i^{(a)} vec\left(\mathbf{\Lambda}_i^{(a)} \right) & \mathbf{0}_{A_i^{(a)} \times A_i^{(a)}} \end{bmatrix} \succeq 0, \tag{67}$$

where $A_i^{(a)}$, the dimension of the identity matrix, is written as

$$A_i^{(a)} = 2 d_i \left(\sum_{j=1, j \neq i}^{K} d_j + KN + \frac{M_i + d_i}{2} \right) + 2 M_i \sum_{j=1}^{K} d_j.$$

To further simplify this constraint, we use the following lemma:

Lemma 1 ([46]): Given matrices $\mathbf{P}, \ \mathbf{Q}, \ \mathbf{A}$ with $\mathbf{A} = \mathbf{A}^H$, the semi-infinite LMI of the form of

$$\mathbf{A} \succeq \mathbf{P}^H \mathbf{X} \mathbf{Q} + \mathbf{Q}^H \mathbf{X}^H \mathbf{P}, \qquad \forall \mathbf{X} : \|\mathbf{X}\|_F \leq \rho,$$

[2]To simplify the presentation, from now on we will assume the number of transmit antennas at the users is equal, i.e. $N = N_i, \ i = 1, \ldots, K$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2016.2606411, IEEE Transactions on Communications

holds if and only if $\exists \epsilon \geq 0$ such that

$$\left[\begin{array}{cc} \mathbf{A} - \epsilon \mathbf{Q}^H \mathbf{Q} & -\rho \mathbf{P}^H \\ -\rho \mathbf{P} & \epsilon \mathbf{I} \end{array} \right] \succeq 0.$$

By choosing

$$\mathbf{A} = \left[\begin{array}{cc} \varsigma_i^{(a)} & \left(\mathbf{d}_i^{(a)}\right)^H \\ \mathbf{d}_i^{(a)} & \mathbf{I}_{A_i^{(a)}} \end{array} \right], \ \mathbf{P} = \left[\mathbf{0}_{NM_i \times 1}, \ \left(\mathbf{D}_i^{(a)}\right)^H \right],$$
$$\mathbf{X} = vec\left(\mathbf{\Lambda}_i^{(a)}\right), \ \mathbf{Q} = \left[-1, \mathbf{0}_{1 \times A_i^{(a)}} \right],$$

$$\mathbf{d}_i^{(a)} = \left[\begin{array}{c} \sqrt{\rho_i} \left[\left(\mathbf{V}_i^{(b)}\right)^T \otimes \left(\left(\mathbf{B}_i^{(a)}\right)^H \mathbf{R}_i^{(a)} \right) \right] vec\left(\tilde{\mathbf{H}}_{ii}^{(ab)}\right) - vec\left(\left(\mathbf{B}_i^{(a)}\right)^H \right) \\ \sqrt{\rho_i \kappa} \left[\left(\mathrm{diag}\left(\mathbf{V}_i^{(b)} \left(\mathbf{V}_i^{(b)}\right)^H \right) \right)^{1/2} \otimes \left(\left(\mathbf{B}_i^{(a)}\right)^H \mathbf{R}_i^{(a)} \right) \right] vec\left(\tilde{\mathbf{H}}_{ii}^{(ab)}\right) \\ \sqrt{\beta \rho_i} \left[\left(\mathbf{V}_i^{(b)}\right)^T \otimes \left(\left(\mathrm{diag}\left(\left(\mathbf{R}_i^{(a)}\right)^H \mathbf{B}_i^{(a)} \left(\mathbf{B}_i^{(a)}\right)^H \mathbf{R}_i^{(a)} \right) \right)^{1/2} \right) \right] vec\left(\tilde{\mathbf{H}}_{ii}^{(ab)}\right) \\ \sqrt{\eta_{ii} \kappa} \left[\left(\mathrm{diag}\left(\mathbf{V}_i^{(a)} \left(\mathbf{V}_i^{(a)}\right)^H \right) \right)^{1/2} \otimes \left(\left(\mathbf{B}_i^{(a)}\right)^H \mathbf{R}_i^{(a)} \right) \right] vec\left(\tilde{\mathbf{H}}_{ii}^{(aa)}\right) \\ \sqrt{\beta \eta_{ii}} \left[\left(\mathbf{V}_i^{(a)}\right)^T \otimes \left(\left(\mathrm{diag}\left(\left(\mathbf{R}_i^{(a)}\right)^H \mathbf{B}_i^{(a)} \left(\mathbf{B}_i^{(a)}\right)^H \mathbf{R}_i^{(a)} \right) \right)^{1/2} \right) \right] vec\left(\tilde{\mathbf{H}}_{ii}^{(aa)}\right) \\ \left. \sqrt{\eta_{ij}} \left[\left(\mathbf{V}_j^{(c)}\right)^T \otimes \left(\left(\mathbf{B}_i^{(a)}\right)^H \mathbf{R}_i^{(a)} \right) \right] vec\left(\tilde{\mathbf{H}}_{ij}^{(ac)}\right) \right|_{j=1,\ldots,K, \ j \neq i, \ c=1,2} \\ \left. \sqrt{\eta_{ij} \kappa} \left[\left(\mathrm{diag}\left(\mathbf{V}_j^{(c)} \left(\mathbf{V}_j^{(c)}\right)^H \right) \right)^{1/2} \otimes \left(\left(\mathbf{B}_i^{(a)}\right)^H \mathbf{R}_i^{(a)} \right) \right] vec\left(\tilde{\mathbf{H}}_{ij}^{(ac)}\right) \right|_{j=1,\ldots,K, \ j \neq i, \ c=1,2} \\ \left. \sqrt{\beta \eta_{ij}} \left[\left(\mathbf{V}_j^{(c)}\right)^T \otimes \left(\left(\mathrm{diag}\left(\left(\mathbf{R}_i^{(a)}\right)^H \mathbf{B}_i^{(a)} \left(\mathbf{B}_i^{(a)}\right)^H \mathbf{R}_i^{(a)} \right) \right)^{1/2} \right) \right] vec\left(\tilde{\mathbf{H}}_{ij}^{(ac)}\right) \right|_{j=1,\ldots,K, \ j \neq i, \ c=1,2} \\ vec\left(\left(\mathbf{B}_i^{(a)}\right)^H \mathbf{R}_i^{(a)} \right) \end{array} \right] \tag{65}$$

$$\mathbf{D}_i^{(a)} = \left[\begin{array}{c} \sqrt{\rho_i} \left[\left(\mathbf{V}_i^{(b)}\right)^T \otimes \left(\left(\mathbf{B}_i^{(a)}\right)^H \mathbf{R}_i^{(a)} \right) \right] \\ \sqrt{\rho_i \kappa} \left[\left(\mathrm{diag}\left(\mathbf{V}_i^{(b)} \left(\mathbf{V}_i^{(b)}\right)^H \right) \right)^{1/2} \otimes \left(\left(\mathbf{B}_i^{(a)}\right)^H \mathbf{R}_i^{(a)} \right) \right] \\ \sqrt{\beta \rho_i} \left[\left(\mathbf{V}_i^{(b)}\right)^T \otimes \left(\left(\mathrm{diag}\left(\left(\mathbf{R}_i^{(a)}\right)^H \mathbf{B}_i^{(a)} \left(\mathbf{B}_i^{(a)}\right)^H \mathbf{R}_i^{(a)} \right) \right)^{1/2} \right) \right] \\ \sqrt{\eta_{ii} \kappa} \left[\left(\mathrm{diag}\left(\mathbf{V}_i^{(a)} \left(\mathbf{V}_i^{(a)}\right)^H \right) \right)^{1/2} \otimes \left(\left(\mathbf{B}_i^{(a)}\right)^H \mathbf{R}_i^{(a)} \right) \right] \\ \sqrt{\beta \eta_{ii}} \left[\left(\mathbf{V}_i^{(a)}\right)^T \otimes \left(\left(\mathrm{diag}\left(\left(\mathbf{R}_i^{(a)}\right)^H \mathbf{B}_i^{(a)} \left(\mathbf{B}_i^{(a)}\right)^H \mathbf{R}_i^{(a)} \right) \right)^{1/2} \right) \right] \\ \left. \sqrt{\eta_{ij}} \left[\left(\mathbf{V}_j^{(c)}\right)^T \otimes \left(\left(\mathbf{B}_i^{(a)}\right)^H \mathbf{R}_i^{(a)} \right) \right] \right|_{j=1,\ldots,K, \ j \neq i, \ c=1,2} \\ \left. \sqrt{\eta_{ij} \kappa} \left[\left(\mathrm{diag}\left(\mathbf{V}_j^{(c)} \left(\mathbf{V}_j^{(c)}\right)^H \right) \right)^{1/2} \otimes \left(\left(\mathbf{B}_i^{(a)}\right)^H \mathbf{R}_i^{(a)} \right) \right] \right|_{j=1,\ldots,K, \ j \neq i, \ c=1,2} \\ \left. \sqrt{\beta \eta_{ij}} \left[\left(\mathbf{V}_j^{(c)}\right)^T \otimes \left(\left(\mathrm{diag}\left(\left(\mathbf{R}_i^{(a)}\right)^H \mathbf{B}_i^{(a)} \left(\mathbf{B}_i^{(a)}\right)^H \mathbf{R}_i^{(a)} \right) \right)^{1/2} \right) \right] \right|_{j=1,\ldots,K, \ j \neq i, \ c=1,2} \\ \mathbf{0}_{d_i M_i \times N M_i} \end{array} \right] \tag{66}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2016.2606411, IEEE Transactions on Communications

we can apply Lemma 1 in (67), and the resulting overall optimization problem is formulated as

$$\max_{\mathbf{V},\mathbf{R},\mathbf{B},\boldsymbol{\varsigma},\boldsymbol{\epsilon}} \quad \sum_{i=1}^{K}\sum_{a=1}^{2}\left(-\varsigma_i^{(a)}+\mu_i^{(a)}\log_2\left|\frac{\ln 2}{\mu_i^{(a)}}\mathbf{B}_i^{(a)}\left(\mathbf{B}_i^{(a)}\right)^H\right| + \frac{d_i\mu_i^{(a)}}{\ln 2}\right) - qf_2(\mathbf{V}) \tag{68}$$

$$\text{s.t.} \quad \text{tr}\left\{\mathbf{V}_i^{(a)}\left(\mathbf{V}_i^{(a)}\right)^H\right\} \le P_i^{(a)},\ \forall(i,a), \tag{69}$$

$$\begin{bmatrix} \varsigma_i^{(a)}-\epsilon_i^{(a)} & \left(\mathbf{d}_i^{(a)}\right)^H & \mathbf{0}_{1\times NM_i} \\ \mathbf{d}_i^{(a)} & \mathbf{I}_{A_i^{(a)}} & -\delta_i^{(a)}\mathbf{D}_i^{(a)} \\ \mathbf{0}_{NM_i\times 1} & -\delta_i^{(a)}\left(\mathbf{D}_i^{(a)}\right)^H & \epsilon_i^{(a)}\mathbf{I}_{NM_i} \end{bmatrix} \succeq 0, \forall(i,a), \tag{70}$$

$$\epsilon_i^{(a)} \ge 0,\ \forall(i,a), \tag{71}$$

where $\boldsymbol{\epsilon} = \left\{\epsilon_i^{(a)} : \forall(i,a)\right\}$, and $\mathbf{B} = \left\{\mathbf{B}_i^{(a)} : \forall(i,a)\right\}$. Although the problem (68)-(71) is non-convex, it becomes a convex function of each optimization variable when the other two are fixed. Therefore we can apply the coordinate ascend method to update the transceiver matrices iteratively. In particular, when $\mathbf{V}$ and $\mathbf{R}$ are fixed, the optimum $\bar{\mathbf{B}}$ can be solved using MAX-DET algorithm [47], when $\mathbf{B}$ and $\mathbf{R}$ are fixed, the optimum $\bar{\mathbf{V}}$ can be computed by solving the resulting semidefinite programming (SDP) problem [48], and when $\mathbf{B}$ and $\mathbf{V}$ are fixed, the optimum $\bar{\mathbf{R}}$ can be computed by solving the resulting SDP problem. Since the alternating iterative updates lead to monotonic increase of the objective function in (68), and the fact that it is bounded above guarantees the convergence of the proposed algorithm.

V. Numerical Results

In this section, we numerically investigate the sum-power minimization and sum energy-efficiency maximization problems for a FD MIMO interference channel. For brevity, we set the same number of transmit and receive antennas at each node, i.e. $M_i = N_i = N,\ i = 1,\ldots,K$.[3]

All the direct and the CCI channel matrices are assumed to be standard Rayleigh fading of unit variance, i.e., the entries of each matrix are i.i.d. circular complex Gaussian variables with zero mean and unit variance. We adopt the Rician model in [1], in which the self-interference channel is distributed as $\tilde{\mathbf{H}}_0 \sim \mathcal{CN}\left(\sqrt{\frac{K_R}{1+K_R}}\hat{\mathbf{H}}_0, \frac{1}{1+K_R}\mathbf{I}_{N_0}\otimes\mathbf{I}_{M_0}\right)$, where K_R is the Rician factor, and $\hat{\mathbf{H}}_0$ is a deterministic matrix. Without loss of generality, we set $K_R = 1$ and $\hat{\mathbf{H}}_0$ to be the matrix of all ones for all experiments. We set the average power gain of the self-interference channel at the ith pair to be $\eta_{ii} = 1$ and the average power gain of the CCI channel between the nodes at the ith and jth pair to be $\eta_{ij} = 0.5$. The average power gain of any transmitter-receiver pair, i.e., ρ_i is set to 1. The initial values for the proposed iterative methods are generated randomly for each channel realization. The results are averaged over 100 independent channel realizations, and the convergence thresholds are 10^{-5}.

[3]Note that although the nodes in the ith link have $N_i + M_i$ antennas in total, similar to [4], 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.

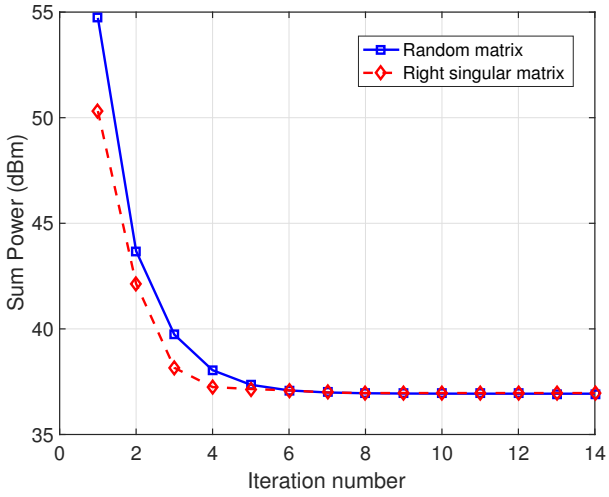

Fig. 2. Convergence behavior of the Sum-Power minimization algorithm.

We compare our proposed algorithms, in which all the pairs operate in FD mode, and transmit at the same time (in particular, we have both self-interference, and CCI from all the pairs in the system) with the HD baseline system. For the HD mode, we adopt the same algorithm proposed for the FD mode, by only omitting the self-interference channel and some part of the CCI channel. In particular, in HD mode, 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 CCI from only one side of nodes, and sum-rate is divided by 2.

A. Sum-Power Minimization

The rate constraint at each node is assumed to be the same, i.e., $R_i^{(a)} = R,\ \forall(i,a)$. Unless otherwise stated, the parameters we use: $K = 2$, $N = 4$, $\kappa = \beta = -70$dB, and $R = 2$bits/s/Hz.

Fig. 2 shows the convergence of the proposed sum-power minimization algorithm given in Algorithm 1. The monotonic decrease of the sum-power is verified for a random channel realization. Moreover, since iteration method's performance may rely on initialization state, for the sake of comparison, plots for two initialization methods are also shown in the figure. In particular,

1) Random matrices initialization: Initialize all the transmit filters with i.i.d. Gaussian random variables.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2016.2606411, IEEE Transactions on Communications

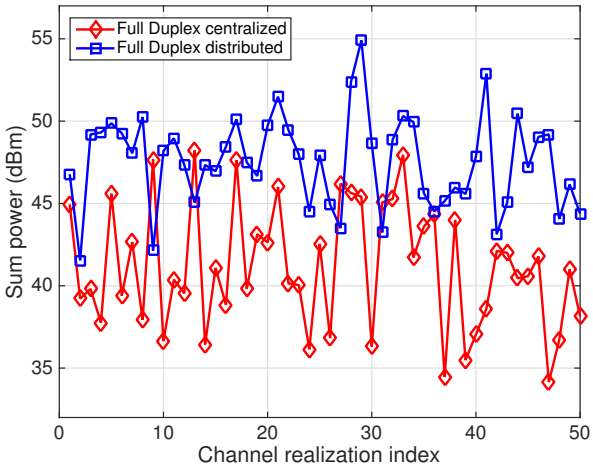

Fig. 3. Sum-Power comparison of FD systems for the centralized and distributed algorithms.

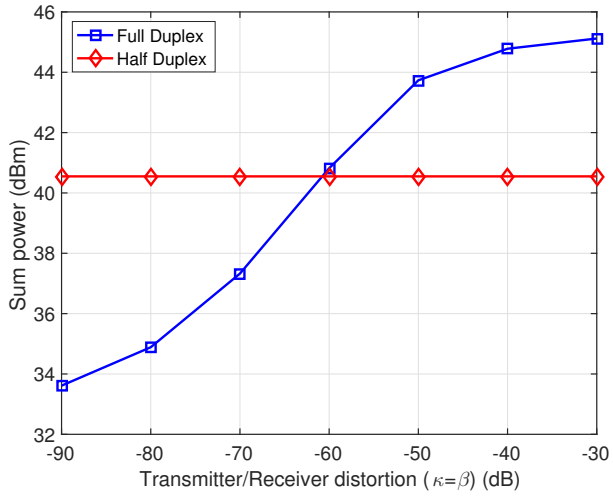

Fig. 4. Sum-Power comparison of FD and HD systems with respect to transmitter/receiver distortion, i.e., κ, β.

2) Right singular matrices initialization: Initialize the $i^{(a)}$th transmit filter $\mathbf{V}_i^{(a)}$ with the first d_i columns of the right singular matrix of $\mathbf{H}_{ii}^{(ba)}$. The initial transmit filters are normalized to satisfy the individual power constraints.

While solid lines correspond to random matrix initialization, dashed lines correspond to right singular matrices initialization. It can be seen from the figure that, both initialization techniques yield almost the same output, which suggests that the algorithm is insensitive to initialization methods, and the right singular matrices initialization converges slightly faster than random matrices one.

In our next example, we compare the centralized and distributed algorithms discussed in Section III for FD systems. Fig. 3 depicts the power consumption in 50 different channel realizations. Since both algorithms converge to locally optimal solutions, there is no guarantee that one will always outperform the other. However, it is seen that the centralized algorithm usually performs better than the distributed one. Although the centralized algorithm outperforms the distributed one, the distributed scheme requires only the local CSI (i.e., each transmitter needs to know only the CSI of the links originating from itself), whereas the centralized method requires the CSI for all links (global CSI). 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, since the computation is now distributed among all the nodes and not only the central node. This may be an important consideration when the network size grows. However, hereinafter in the following figures related to sum-power minimization problem, we will only consider the centralized algorithm.

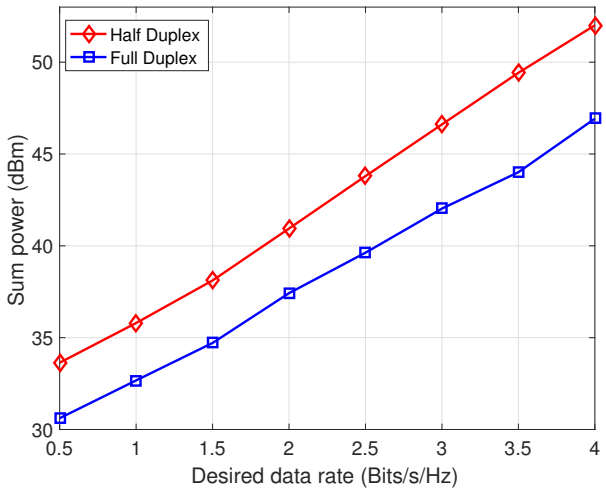

Fig. 5. Sum-Power comparison of FD and HD systems with respect to desired data rate, R.

In the following example, we will analyze the effect of the transmitter and receiver distortion on FD and HD systems. It can be seen from Fig. 4 that HD system is not affected with κ and β values. When the distortion at the transmitter/receiver is low, FD system requires less power to achieve the same throughput. However, this trend changes when the distortion values are high. The performance shift happens at around $\kappa = \beta = -60$dB.

In Fig. 5, we compare FD with HD systems in terms of sum-power consumption for different values of desired data rates. It is quite evident from the figure that greater demands of rate can be fulfilled by increasing the transmitted power. However, at low transmitter/receiver distortion (in this case -70dB), FD systems requires less power than HD system to provide a certain rate. Also can be seen that the gap in performance between FD and HD systems increases with the increase in data rates, which further emphasizes the superiority of FD

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2016.2606411, IEEE Transactions on Communications

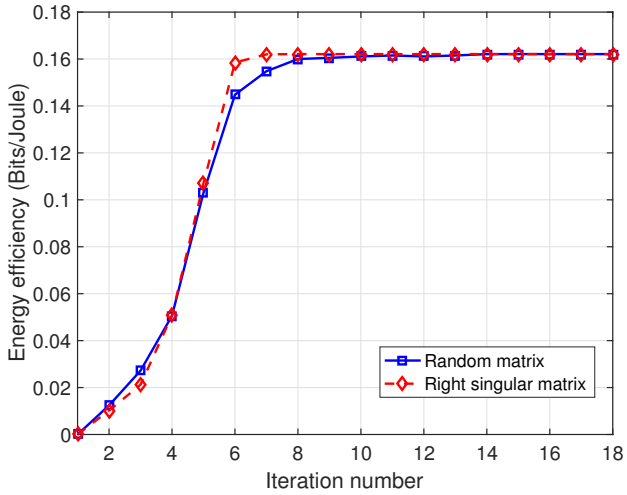

Fig. 6. Convergence behavior of the EE maximization algorithm.

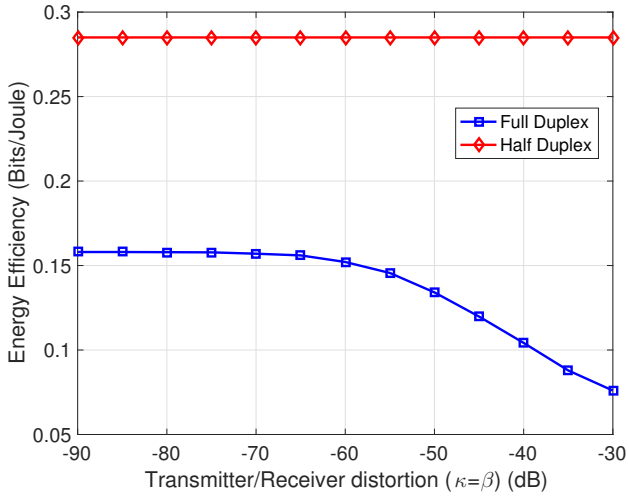

Fig. 7. EE comparison of FD and HD systems with respect to transmitter/receiver distortion, i.e., κ, β.

systems over HD systems.

B. Energy-efficiency Maximization

Hereinafter, we analyze the energy-efficiency maximization algorithm based on a certain transmit power constraint for each node in the system, i.e., $P_i^{(b)}$, $\forall (i,b)$. We also assumed the same weights, i.e., $\mu_i^{(a)} = \mu$, $\forall (i,a)$. Unless otherwise stated, the parameters we use: $K = 2$, $N = 4$, $P_i^{(b)} = 33$dBm, $\kappa = \beta = -70$dB, and $\vartheta = 1/0.32$.

Fig. 6 shows the evolution of the proposed algorithm given in Algorithm 2. We set the circuit power to $P_c = 38$dBm and the basic power consumed by the transmitters to $P_o = 27$dBm. It can be seen that the energy-efficiency maximization problem converges in a few steps and it does so monotonically. Moreover, similar to Fig. 2, we show that this algorithm is also insensitive to initialization techniques, where both random and right singular matrix initializations yield almost the same result.

In Fig. 7, the comparison between the energy-efficiency of a FD system to that of a HD system is shown. Based on the optimal energy-efficiency design, the FD and HD curves are plotted with respect to $\kappa = \beta$ values. The settings for the circuit power and the basic power consumed by the transmitters are maintained from the previous figure. As can be seen in the figure, the performance of HD system is not affected with κ and β values. However, HD system outperforms FD system in terms of energy-efficiency. This can be explained as follows. All the channels of the FD systems are always active, none being silent at any point of time as all the nodes transmit and receive at the same time-slot, and needs to overcome self-interference and extra CCI which leads to more power consumption than the HD transmission mode. However, in HD system, only the nodes of one side (left or right) transmit in one time slot. As a result the HD system requires less energy than the FD one for the same period of operation, which makes it more energy-efficient as compared to FD systems.

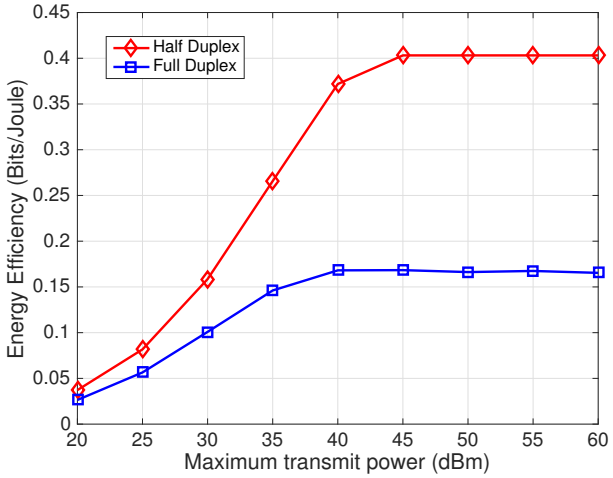

Fig. 8. EE comparison of FD and HD systems with respect to $P_i^{(a)}$.

In Fig. 8, we compare the energy-efficiency of a FD system to that of a HD system with respect to the maximum transmit power constraint. The circuit power is set to $P_c = 38$dBm and the basic power consumed by the transmitters to $P_o = 27$dBm. Similar to the previous figure, HD system is more energy-efficient than FD systems at high transmit power constraints. However, the gap in performance is less and somewhat equivalent at lower transmitted power. The reason is that increased transmission power is more beneficial for HD systems in term of sum-rate, because higher transmission power leads to higher self-interference power in FD systems.

In Fig. 9, we compare the energy-efficiency of a FD system with respect to two designs, namely i) energy-efficiency (EE) maximization and ii) spectral efficiency (SE) maximization. Note that the EE maximization algorithm proposed in Algorithm 2 can be modified to solve the WSR problem by dropping

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2016.2606411, IEEE Transactions on Communications

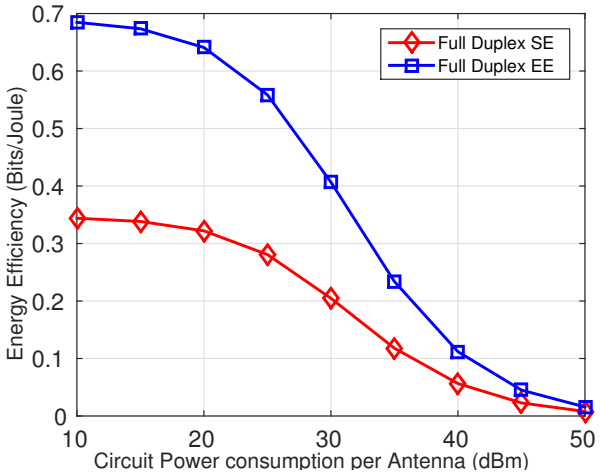

Fig. 9. Comparison between EE and SE maximization algorithms with respect to P_c.

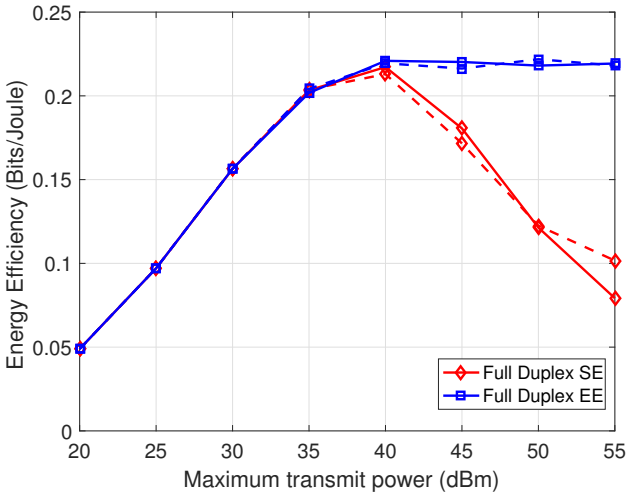

Fig. 10. Comparison between EE and SE maximization algorithms with respect to $P_i^{(a)}$.

the outer layer, where the energy-efficiency parameter q is updated. The basic power consumed by the transmitters is $P_o = 35$dBm and the transmit power constraint is 42dBm.

Based on the two transmit beamforming design techniques, the FD curves are plotted with respect to the power consumed by the circuit components in the system. It can be seen from the figure that the transmit beamforming design based on EE maximization performs better and at low circuit power significantly outperforms the other. This is due to the fact that the transmit beamforming design based on the SE maximization does not take into consideration other sources of power consumed. We however note that as the circuit power increases, the energy-efficiency of both the designs decreases and somewhat becomes equivalent. At high circuit power values, maximizing the energy-efficiency is equivalent to maximizing the spectral-efficiency, since the total consumed power depends mostly on the circuit power consumption.

To further compare the two designs, in Fig. 10 we evaluate the energy-efficiency performance of both the designs with respect to the maximum transmit power constraint. We set the circuit power to $P_c = 38$dBm and the basic power consumed by the transmitters to $P_o = 27$dBm. At low transmit power consumption region, both designs show equivalent energy-efficiency performance, and the energy-efficiency of both designs increases as the maximum transmit power constraint increases, since in this low transmit power consumption region, the total power consumption is primarily due to the circuit power consumption. However, as we increase the transmit power, the energy-efficiency of SE transmit beamforming design attains a maximum energy-efficiency for a certain power and its energy-efficiency performance decreases after that. On the other hand, the energy-efficiency of the EE transmit beamforming design does not decrease, and increases until it reaches to a peak value and remains constant after that. This is due to the fact that to maximize the spectral-efficiency, full power transmission is required in the SE maximization design, which leads to a linear increase in total power consumption in high transmit power regime. The spectral-efficiency however, increases on a logarithmic scale with respect to the transmitted power which becomes more implicit in the high transmit power regime. Hence, the energy-efficiency of the SE design reduces at high transmit power consumption region. The technique used for energy-efficiency maximization in the EE design, ensures that an optimal transmit power is found and if that optimal power is less than the maximum transmitted power, the EE-optimal design does not transmit at full power. As a result the energy-efficiency of this design remains constant after it reaches a certain peak power. Moreover, for the sake of comparison, plots for two initialization methods, as discussed before are also shown in this figure. While solid lines correspond to random matrix initialization, dashed lines correspond to right singular matrices initialization. As was seen before, here again both initialization techniques provide equivalent performance.

VI. Conclusion

In this work, we have addressed the transmit beamforming design for sum-power minimization and sum energy-efficiency maximization problems in a FD MIMO interference channel that suffer from self-interference and interference from other links under the limited DR at the transmitters and receivers. Since the globally optimal solution is difficult to obtain due to the non-convex nature of the problems, an alternating iterative algorithm to find a stationary point was proposed based on the relationship between WSR and WMMSE problems. Extensions to distributed and robust algorithms under imperfect CSI assumption are also provided. The numerical results show that in HD systems, the sum-power obtained is invariant to the transmitter/receiver distortion. However, the sum-power achieved in FD mode depends heavily on the transmitter/receiver distortion as can be expected. Also the energy-efficiency obtained is lower in FD than HD systems. In particular, there is a penalty in energy efficiency for FD operation, since additional power is consumed to overcome

the interference from introduced by FD operation, i.e., self-interference and extra CCI.

REFERENCES

[1] 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.

[2] 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.

[3] Y. Hua, Y. Ma, A. Gholian, Y. Li, A. C. Cirik, P. Liang, "Radio self-interference cancellation by transmit beamforming, all-analog cancellation and blind digital tuning," *Elsevier Signal Process.*, vol. 108, pp. 322-340, 2014.

[4] B. P. Day, A. R. Margetts, D. W. Bliss, and P. Schniter, "Full-duplex bidirectional MIMO: Achievable rates under limited dynamic range," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3702-3713, July 2012.

[5] S. Huberman, and T. Le-Ngoc, "MIMO full-duplex precoding: a joint beamforming and self-interference cancellation structure," *IEEE Trans. Wireless Commun.*, vol. 14, no. 4, pp. 2205-2217, April 2015.

[6] 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.

[7] A. C. Cirik, R. Wang, Y. Hua, and M. Latva-aho, "Weighted sum-rate maximization for full-duplex MIMO interference channels," *IEEE Trans. Commun.*, vol. 63, no. 3, pp. 801-815, March 2015.

[8] A. C. Cirik, R. Wang, Y. Rong, and Y. Hua,"MSE-based transceiver designs for full-duplex MIMO cognitive radios," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2056-2070, June 2015.

[9] 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.

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

[11] F. Rashid-Farrokhi, K. Liu, and L. Tassiulas, "Transmit beamforming and power control for cellular wireless systems," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1437-1450, Oct. 1998.

[12] C. Y. Wong, R. Cheng, K. Lataief, and R. Murch, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 10, pp. 1747-1758, Oct. 1999

[13] Y. Rong, and Y. Hua, "Space-time power schedule for distributed MIMO links–Fairness and QoS considerations," *IEEE J. Sel. Topics Signal Process.*, vol. 2, pp. 171-180, Apr. 2008.

[14] A. Liu, and V. K. N. Lau, "Mixed-timescale precoding and cache control in cached MIMO interference network," *IEEE Trans. Signal Process.*, vol. 61, no. 24, pp. 6320-6332, Dec. 2013.

[15] 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, pp. 4792-4799, Dec. 2008.

[16] 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 Proces.*, vol. 59, no. 9, pp. 4331-4340, Sep. 2011.

[17] G. Miao, S. He, Y. Huang, S. Jin, and L. Yang, "Coordinated beamforming for energy efficient transmission in multicell multiuser systems," *IEEE Trans. Commun.*, vol. 61, no. 12, pp. 4961-4971, Dec. 2013

[18] S. Goyal, P. Liu, S. S. Panwar, R. A. DiFazio, R. Yang, J. Li, and E. Bala, "Improving small cell capacity with common-carrier full duplex radios," *IEEE Intl. Conf. Commun.*, pp. 4987-4993, June 2014.

[19] H. Chen, G. Li, and J. Cai, "Spectral-energy efficiency tradeoff in full-duplex two-way relay networks," *IEEE Systems Journal*, in press.

[20] W. Dinkelbach, "On nonlinear fractional programming," *Management Science*, vol. 13, no. 7, pp. 492-498, Mar. 1967.

[21] J. Crouzeix, and J. Ferland, "Algorithms for generalized fractional programming," *Math. Programming*, vol. 52, pp. 191-207, Jun. 1991.

[22] 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, pp. 867-876, Aug. 2008.

[23] W. Namgoong, "Modeling and analysis of nonlinearities and mismatches in AC-coupled direct-conversion receiver," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 163-173, Jan. 2005.

[24] Y.-S. Choi and H. Shirani-Mehr, "Simultaneous transmission and reception: Algorithm, design and system level performance," IEEE Trans. Wireless Commun., vol. 12, no. 12, pp. 5992-6010, Dec. 2013.

[25] C. Nam, C. Joo, and S. Bahk, "Radio resource allocation with inter-node interference in full-duplex OFDMA networks," *IEEE ICC*, pp. 3885-3890, 2015.

[26] S. Goyal, P. Liu, S. Panwar, R. Difazio, R. Yang, and E. Bala, "Full duplex cellular systems: will doubling interference prevent doubling capacity?" *IEEE Commun. Mag.*, vol. 53, no. 5, pp. 121-127, May 2015.

[27] N. Parikh and S. Boyd, "Proximal algorithms," *Found. Trends Optim.*, vol. 1, no. 3, pp. 123-231, 2013.

[28] 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.

[29] M. V. Solodov, "On the convergence of constrained parallel variable distribution algorithm," *SIAM J. Optim.*, vol. 8, no. 1, pp. 187-196, Feb. 1998.

[30] Y. Rong and Y. Hua, "Optimal power schedule for distributed MIMO links," *IEEE Trans. Wireless Commun.*, vol. 7, no. 8, pp. 2896-2900, Aug. 2008.

[31] Y. Rong, Y. Hua, A. Swami, and A. L. Swindlehurst, "Space-time power schedule for distributed MIMO links without instantaneous channel state information at the transmitting nodes," *IEEE Trans. Signal Proccess.*, vol. 56, no. 2, pp. 686-700, Feb. 2008.

[32] Y. Rong and Y. Hua, "Space-time power scheduling of MIMO links-Fairness and QoS considerations," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 2, pp. 171-180, Apr. 2008.

[33] M. Razaviyayn, M. Sanjabi, and Z. Luo, "Linear transceiver design for interference alignment: Complexity and computation," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 2896-2910, May 2012.

[34] C. T. K. Ng and H. Huang, "Linear precoding in cooperative MIMO cellular networks with limited coordination clusters," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1446-1454, Dec. 2010.

[35] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*, Prentice-Hall, New Jersey, 1989.

[36] D. P. Palomar and M. Chiang, "A tutorial on decomposition methods for network utility maximization," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1439-1451, Aug. 2006.

[37] O. Arnold, F. Richter, G. Fettweis, and O. Blume, "Power consumption modeling of different base station types in heterogeneous cellular networks," in *Proc. 19th Future Netw. MobileSummit*, pp. 18, Jun. 2010.

[38] L. Grippo and M. Sciandrone, "On the convergence of the block nonlinear Gauss-Seidel method under convex constraints," *Operations Research Letters*, vol. 26, no. 3, pp. 1-10, 2000.

[39] R. Hunger, "Floating point operations in matrix-vector calculus," *Tech. Rep.*, Munich University of Technology, Sept. 2007.

[40] D. Nguyen and M. Krunz, "Power-efficient spatial multiplexing for multi-antenna MANETs," *IEEE Int. Conf. Commun. (ICC)*, pp. 4016-4020, June 2012.

[41] 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, 2011.

[42] 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.

[43] L. Zhang, Y.-C. Liang, Y. Xin, and H. V. Poor, "Robust cognitive beamforming with partial channel state information," *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 4143-4153, Aug. 2009.

[44] E. A. Gharavol, Y.-C. Liang, and K. Mouthaan, "Robust linear transceiver design in MIMO ad hoc cognitive radio networks with imperfect channel state information," *IEEE Trans. Wireless Commun.*, vol. 10, no. 5, pp. 1448-1457, May 2011.

[45] Y. Zhang, E. DallAnese, and G. B. Giannakis, "Distributed optimal beamformers for cognitive radios robust to channel uncertainties," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6495-6508, Dec. 2012.

[46] Y. C. Eldar, A. Ben-Tal, and A. Nemirovski, "Robust mean-squared-error estimation in the presence of model uncertainties," *IEEE Trans. Signal Process.*, vol. 53, pp. 161-176, Jan. 2005.

[47] L. Vandenberghe, S. Boyd, and S. Wu, "Determinant maximization with linear matrix inequality constraints," *SIAM journal on matrix analysis and applications*, vol. 19, pp. 499-533, 1998.

[48] S. Boyd, and L. Vandenberghe. "Convex optimization". Cambridge university press, 2004.

Ali Cagatay Cirik (S'13-M'14) received the B.S and M.S. degrees in telecommunications and electronics engineering from Sabanci University, Istanbul, Turkey, in 2007 and 2009, respectively, and Ph.D. degree in electrical engineering from University of California, Riverside in 2014. He held research fellow positions at Centre for Wireless Communications, Oulu, Finland and Institute for Digital Communications (IDCOM), University of Edinburgh, U.K between June 2014 and November 2015. His industry experience includes internships at Mitsubishi Electric Research Labs (MERL), Cambridge, MA, in 2012 and at Broadcom Corporation, Irvine, CA, in 2013. Currently, he is working as a research scientist at Sierra Wireless, Richmond, Canada. He is also affiliated with University of British Columbia, Vancouver, Canada. His primary research interests are full-duplex communication, 5G non-orthogonal multiple-access (NOMA), MIMO signal processing, and convex optimization.

Sudip Biswas (S'15) received the B.Tech. degree in electronics and communication engineering from the Sikkim Manipal Institute of Technology, Sikkim, India, in 2010, and the M.Sc. degree in signal processing and communications from the University of Edinburgh, Edinburgh, U.K., in 2013. He is currently pursuing the Ph.D. degree in digital communications at the University of Edinburgh's Institute for Digital Communications. His research interests include various topics in wireless communications and network information theory with particular focus on stochastic geometry and possible 5G technologies such as massive MIMO, mmWave, and full-duplex.

Satyanarayana Vuppala (S'12-M'16) received the B.Tech. degree with distinction in Computer Science and Engineering from JNTU Kakinada, India, in 2009, and the M.Tech. degree in Information Technology from the National Institute of Technology, Durgapur, India, in 2011. He received the Ph.D. degree in Electrical Engineering from Jacobs University Bremen in 2014. He is currently a post-doctoral researcher at IDCOM in University of Edinburgh. His main research interests are physical, access, and network layer aspects of wireless security. He also works on performance evaluation of mmWave systems. He is a recipient of MHRD, India scholarship during the period of 2009-2011.

Tharmalingam Ratnarajah (A'96-M'05-SM'05) is currently with the Institute for Digital Communications, University of Edinburgh, Edinburgh, UK, as a Professor in Digital Communications and Signal Processing. His research interests include signal processing and information theoretic aspects of 5G wireless networks, full-duplex radio, mmWave communications, random matrices theory, interference alignment, statistical and array signal processing and quantum information theory. He has published over 300 publications in these areas and holds four U.S. patents. He is currently the coordinator of the FP7 projects HARP (3.2M€) in the area of highly distributed MIMO and ADEL (3.7M€) in the area of licensed shared access. Previously, he was the coordinator of FP7 Future and Emerging Technologies project CROWN (2.3M€) in the area of cognitive radio networks and HIATUS (2.7M€) in the area of interference alignment. Dr Ratnarajah is a Fellow of Higher Education Academy (FHEA), U.K., and an associate editor of the IEEE Transactions on Signal Processing.