An Iteratively Weighted MMSE Approach to Distributed Sum-Utility Maximization for a MIMO Interfering Broadcast Channel

Qingjiang Shi, Meisam Razaviyayn, Zhi-Quan Luo, *Fellow, IEEE*, and Chen He, *Member, IEEE*

***Abstract*—Consider the multiple-input multiple-output (MIMO) interfering broadcast channel whereby multiple base stations in a cellular network simultaneously transmit signals to a group of users in their own cells while causing interference to each other. The basic problem is to design linear beamformers that can maximize the system throughput. In this paper, we propose a linear transceiver design algorithm for weighted sum-rate maximization that is based on iterative minimization of weighted mean-square error (MSE). The proposed algorithm only needs local channel knowledge and converges to a stationary point of the weighted sum-rate maximization problem. Furthermore, the algorithm and its convergence can be extended to a general class of sum-utility maximization problem. The effectiveness of the proposed algorithm is validated by numerical experiments.**

***Index Terms*—Linear beamformer, MIMO interfering broadcast channel, sum-utility maximization, weighted MMSE, weighted sum-rate maximization.**

I. Introduction

CONSIDER a multiple-input multiple-output (MIMO) interfering broadcast channel (IBC) in which a number of transmitters, each equipped with multiple antennas, wish to simultaneously send independent data streams to their intended receivers. As a generic model for multiuser downlink communication, MIMO-IBC can be used in the study of many practical systems such as digital subscriber lines (DSLs), cognitive radio systems, ad hoc wireless networks, wireless cellular communication, to name just a few. Unfortunately, despite the importance and years of intensive research, the search for optimal transmit/receive strategies that can maximize the weighted sum-rate of all users in a MIMO-IBC remains rather elusive. In fact, even for the simpler case of MIMO interference channel, the optimal strategy is still unknown. This lack of understanding of the capacity region has motivated a pragmatic approach whereby we simply treat interference as noise and maximize the weighted sum-rate by searching within the class of linear transmit/receive strategies.

Weighted sum-rate maximization for an interference channel (IFC), which is a special case of IBC, has been a topic of intensive research in recent years. From the optimization's perspective, this problem is nonconvex and NP-hard even in the single-antenna case [1]. Thus, most current research efforts have been focused on finding a high quality suboptimal solution efficiently. For instance, to facilitate distributed implementation, noncooperative game theoretic methods have been proposed for the weighted sum-rate maximization problem in an IFC. These methods include those for the two-user IFC [3]–[5], the single-input single-output (SISO) and multiple-input single-output (MISO) IFC [6], [7], [9], as well as distributed algorithms for the MIMO IFC with no multiplexing, i.e., one datastream for each user [10], [11]. In the game theoretic methods, users are considered as players in a game where they each greedily maximize their own utility/payoff functions. Since only local channel information is required at each user, these methods are ideal for distributed implementation, although their convergence is only towards a Nash equilibrium which may be far away from sum-rate maximum. In contrast, the so-called *interference-pricing* based methods (e.g., [6]) let each user maximize its own utility minus the interference cost determined by the interference prices. By choosing interference price as the marginal decrease in the sum-rate utility per unit increase in interference power, the interference pricing based methods can reach a stationary point of the overall utility maximization problem (rather than a Nash equilibrium). In particular, [6] has established the convergence of the interference pricing algorithm to a stationary point for a set of utility functions, which unfortunately does not include the standard Shannon rate function. Several extensions and variations of the interference pricing algorithm have been proposed [9] for the SISO and MISO IFC that can monotonically converge to a stationary point of the weighted sum-rate maximization problem. A similar algorithm for the MIMO interference channel was proposed in [10] without considering multiplexing (i.e., one data stream per user). All of these

Manuscript received November 26, 2010; revised April 11, 2011; accepted April 18, 2011. Date of publication April 29, 2011; date of current version August 10, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Ta-Sung Lee. This work was supported in part by the Army Research Office, Grant W911NF-09-1-0279 and in part by the National Science Foundation, Grant CMMI-0726336, as well as by a research gift from Huawei Technologies, Inc. The work of Q. Shi was performed while he was a visiting Ph.D. student at the University of Minnesota, with financial support from the China Scholarship Council and the Important National Science & Technology Specific Projects (2010ZX03003-002-03 and 2011ZX03003-001-03).

Q. Shi was with the Department of Electronic Engineering, Shanghai Jiao Tong University. He is now with the Research & Innovation Center, Alcatel-Lucent Shanghai Bell Company, Limited, Shanghai 201206, China (e-mail: Qingjiang.Shi@alcatel-sbell.com.cn).

M. Razaviyayn and Z.-Q. Luo are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: luozq@umn.edu; razav002@umn.edu).

C. He is with the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China, 200240 (e-mail: chenhe@sjtu.edu.cn).

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

Digital Object Identifier 10.1109/TSP.2011.2147784

algorithms allow only one user to update its beamformer at a time, which may lead to excessive communication overhead for price exchanges. A general distributed interference pricing algorithm that allows multiple users to update simultaneously was proposed in [11] for the MIMO interference channel with no multiplexing, although no convergence analysis is provided for the algorithm. (The DBA-RF algorithm in [11] is in essence a distributed pricing algorithm which adopts a different pricing strategy from [10].)

By fixing the receiver structure to any of the standard linear receivers (e.g., the MMSE or zero-forcing receivers), we can reduce the linear transceiver design to a transmit covariance matrix design problem. Reference [2] proposed an iterative algorithm based on the gradient projection method for the transmit covariance matrix design problem. The algorithm allows each user to update its own covariance matrix locally, provided that the channel state information and the covariance matrices of other users can be gathered. Based on a local linear approximation, reference [12] proposed a distributed algorithm which lets each user update its own covariance matrix by solving a convex optimization problem. This algorithm can be viewed as the MIMO extension version of the sequential distributed pricing algorithm in [9]. We henceforth unify the name of these algorithms as the iterative linear approximation (ILA) algorithm. Moreover, since these algorithms use a local tight concave lower bound approximation of the weighted sum-rate objective function, they ensure that the rates increase monotonically and that the transmit covariance matrices converge to a stationary point of the original objective function (i.e., the weighted sum-rate) [9], [21].

A different sum-rate maximization approach was proposed in [13] for the MIMO broadcast downlink channel, where the weighted sum-rate maximization problem is transformed to an equivalent weighted sum MSE minimization (WMMSE) problem with some specially chosen weight matrices that depend on the optimal beamforming matrices. Since the weight matrices are generally unknown, the authors of [13] proposed an iterative algorithm that adaptively chooses the weight matrices and updates the linear transmit–receive beamformers at each iteration. A nonconvex cost function was constructed [13] and shown to monotonically decrease as the algorithm progresses. However, the convergence of the iterates to a stationary point (or the global minimum) of the cost function is not known. A similar algorithm has been proposed in [14] for the interference channel where each user only transmits one data stream.

Inspired by the work of [13] and [14] and utilizing the block coordinate descent technique [15], we propose a simple distributed linear transceiver design method, named the WMMSE algorithm, for general utility maximization in an interfering broadcast channel. This algorithm extends the existing algorithms of [13] and [14] in several directions. In particular, it can handle fairly general utility functions (which includes weighted sum-rate utility function as a special case), and works for general MIMO interfering broadcast channel (which includes MIMO broadcast channel [13] and MISO interference channel [14] as special cases). Moreover, the proposed WMMSE algorithm can be extended to accommodate channel estimation error to achieve robust weighted sum-rate maximization, although this extension will be reported elsewhere. Theoretically, we show that the sequence of iterates generated by the WMMSE algorithm converges to at least a local optima of the utility maximization problem, and does so with low communication and computational complexity.

Throughout this paper, we adopt the following notations. We use the superscript ^{H} to denote the Hermitian transpose of a matrix and overline the complex conjugate. Furthermore, we use capital bold face letters for matrices while keeping the small bold for vectors and small normal face for scalars. The identity matrix is denoted by $\mathbf{I}$, and $\mathbb{C}^{m\times n}$ denotes an m by n dimensional complex space. The complex and real normal distribution are represented by $\mathcal{CN}(\cdot,\cdot)$ and $\mathcal{N}(\cdot,\cdot)$, respectively. The notations $\mathbb{E}(\cdot)$, $\mathrm{Tr}(\cdot)$, and $\det(\cdot)$ represent the expectation, trace, and determinant operator. $\nabla f(\cdot)$ is used to denote the gradient of the function $f(\cdot)$. For symmetric matrices $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{A}\succ\mathbf{B}(\mathbf{A}\succeq\mathbf{B})$ signifies that $\mathbf{A}-\mathbf{B}$ is positive definite (positive semidefinite).

II. System Model and Problem Formulation

Consider a K cell interfering broadcast channel where the base station k, $k=1,2,\ldots,K$ is equipped with M_k transmit antennas and serves I_k users in cell k. Let us define i_k to be the ith user in cell k and N_{i_k} be the number of receive antennas at receiver i_k. Let us also define $\mathcal{I}$ to be the set of all receivers, i.e.,

$$\mathcal{I}=\{i_k \mid k\in\{1,2,\ldots,K\}, i\in\{1,2,\ldots,I_k\}\}.$$

Let $\mathbf{V}_{i_k}\in\mathbb{C}^{M_k\times d_{i_k}}$ denote the beamformer that base station k uses to transmit the signal $\mathbf{s}_{i_k}\in\mathbb{C}^{d_{i_k}\times 1}$ to receiver i_k, $i=1,2,\ldots,I_k$, i.e.,

$$\mathbf{x}_k=\sum_{i=1}^{I_k}\mathbf{V}_{i_k}\mathbf{s}_{i_k}$$

where we assume $\mathbb{E}\left[\mathbf{s}_{i_k}\mathbf{s}_{i_k}^H\right]=\mathbf{I}$. Assuming a linear channel model, the received signal $\mathbf{y}_{i_k}\in\mathbb{C}^{N_{i_k}\times 1}$ at receiver i_k can be written as

$$\mathbf{y}_{i_k}=\underbrace{\mathbf{H}_{i_kk}\mathbf{V}_{i_k}\mathbf{s}_{i_k}}_{\text{desired signal}}+\underbrace{\sum_{m=1,m\neq i}^{I_k}\mathbf{H}_{i_kk}\mathbf{V}_{m_k}\mathbf{s}_{m_k}}_{\text{intracell interference}}+\underbrace{\sum_{j\neq k,j=1}^{K}\sum_{\ell=1}^{I_j}\mathbf{H}_{i_kj}\mathbf{V}_{\ell_j}\mathbf{s}_{\ell_j}+\mathbf{n}_{i_k}}_{\text{intercell interference plus noise}},\ \forall i_k\in\mathcal{I}$$

where matrix $\mathbf{H}_{i_kj}\in\mathbb{C}^{N_{i_k}\times M_j}$ represents the channel from the transmitter j to receiver i_k, while $\mathbf{n}_{i_k}\in\mathbb{C}^{N_k\times 1}$ denotes the additive white Gaussian noise with distribution $\mathcal{CN}(0,\sigma_{i_k}^2\mathbf{I})$. We assume that the signals for different users are independent from each other and from receiver noises. In this paper, we treat interference as noise and consider linear receive beamforming strategy so that the estimated signal is given by

$$\hat{\mathbf{s}}_{i_k}=\mathbf{U}_{i_k}^H\mathbf{y}_{i_k},\quad\forall\, i_k\in\mathcal{I}.$$

Then, the problem of interest is to find the transmit and receive beamformers[1] $\{\mathbf{V},\mathbf{U}\}$ such that a certain utility of the system is maximized, while the power budget of each transmitter is respected:

$$\sum_{i=1}^{I_k}\mathrm{Tr}(\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H)\le P_k$$

where P_k denotes the power budget of transmitter k.

In what follows, we first consider the popular sum-rate utility function and show that the weighted sum-rate maximization problem can be related to a matrix-weighted sum-MSE minimization problem. Then, we generalize the results to a wider class of utility functions.

A. Weighted Sum-Rate Maximization and a Matrix-Weighted Sum-MSE Minimization

A popular utility maximization problem is the weighted sum-rate maximization which can be written as

$$\begin{aligned}\max_{\mathbf{V}}\quad & \sum_{k=1}^{K}\sum_{i_k=1}^{I_k}\alpha_{i_k}R_{i_k}\\ \text{s.t.}\quad & \sum_{i=1}^{I_k}\mathrm{Tr}(\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H)\le P_k,\ \forall\, k=1,2,\ldots,K \end{aligned}\tag{1}$$

where R_{i_k} is the rate of user i_k which can be written as (2), shown at the bottom of the page; the weight α_{i_k} is used to represent the priority of user i_k in the system.

Another popular utility maximization problem for MIMO-IBC is sum-MSE minimization. Under the independence assumption of $\mathbf{s}_{i_k}$'s and $\mathbf{n}_{i_k}$'s, the MSE matrix $\mathbf{E}_{i_k}$ can be written as

$$\begin{aligned}\mathbf{E}_{i_k}&\triangleq\mathbb{E}_{\mathbf{s},\mathbf{n}}\left[(\hat{\mathbf{s}}_{i_k}-\mathbf{s}_{i_k})(\hat{\mathbf{s}}_{i_k}-\mathbf{s}_{i_k})^H\right]\\ &=(\mathbf{I}-\mathbf{U}_{i_k}^H\mathbf{H}_{i_kk}\mathbf{V}_{i_k})(\mathbf{I}-\mathbf{U}_{i_k}^H\mathbf{H}_{i_kk}\mathbf{V}_{i_k})^H\\ &\quad+\sum_{(\ell,j)\neq(i,k)}\mathbf{U}_{i_k}\mathbf{H}_{i_kj}\mathbf{V}_{\ell_j}\mathbf{V}_{\ell_j}^H\mathbf{H}_{i_kj}^H\mathbf{U}_{i_k}^H+\sigma_{i_k}^2\mathbf{U}_{i_k}^H\mathbf{U}_{i_k},\end{aligned}\tag{3}$$

[1]The notation $\mathbf{V}$ is short for $\{\mathbf{V}_{i_k}\}_{i_k\in\mathcal{I}}$, which denotes all variables $\mathbf{V}_{i_k}$ with $i_k\in\mathcal{I}$. The short notations $\mathbf{U}=\{\mathbf{U}_{i_k}\}_{i_k\in\mathcal{I}}$, $v=\{v_{i_k}\}_{i_k\in\mathcal{I}}$, $u=\{u_{i_k}\}_{i_k\in\mathcal{I}}$, $w=\{w_{i_k}\}_{i_k\in\mathcal{I}}$, $\mathbf{U}^*=\{\mathbf{U}_{i_k}^*\}_{i_k\in\mathcal{I}}$, $\mathbf{V}^*=\{\mathbf{V}_{i_k}^*\}_{i_k\in\mathcal{I}}$, $\mathbf{W}^*=\{\mathbf{W}_{i_k}^*\}_{i_k\in\mathcal{I}}$ are defined similarly. Moreover, unless otherwise specified, $\mathbf{W}$ is short for $\{\mathbf{W}_{i_k}\}_{i_k\in\mathcal{I}}$.

and the sum-MSE minimization problem for the MIMO-IBC can be written as

$$\begin{aligned}\min_{\mathbf{U},\mathbf{V}}\quad & \sum_{k=1}^{K}\sum_{i=1}^{I_k}\mathrm{Tr}(\mathbf{E}_{i_k})\\ \text{s.t.}\quad & \sum_{i=1}^{I_k}\mathrm{Tr}\left(\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H\right)\le P_k,\quad k=1,2,\ldots,K.\end{aligned}\tag{4}$$

Fixing all the transmit beamformers $\mathbf{V}$ and minimizing (weighted) sum-MSE lead to the well-known MMSE receiver:

$$\mathbf{U}_{i_k}^{\mathrm{mmse}}=\mathbf{J}_{i_k}^{-1}\mathbf{H}_{i_kk}\mathbf{V}_{i_k}\tag{5}$$

where $\mathbf{J}_{i_k}\triangleq\sum_{j=1}^{K}\sum_{\ell=1}^{I_j}\mathbf{H}_{i_kj}\mathbf{V}_{\ell_j}\mathbf{V}_{\ell_j}^H\mathbf{H}_{i_kj}^H+\sigma_{i_k}^2\mathbf{I}$ is the covariance matrix of the total received signal at receiver i_k. Using this MMSE receiver, the corresponding MSE matrix is given by

$$\mathbf{E}_{i_k}^{\mathrm{mmse}}=\mathbf{I}-\mathbf{V}_{i_k}^H\mathbf{H}_{i_kk}^H\mathbf{J}_{i_k}^{-1}\mathbf{H}_{i_kk}\mathbf{V}_{i_k}.\tag{6}$$

The following result establishes the equivalence between the weighted sum-rate maximization problem and a matrix-weighted sum-MSE minimization problem (7).

Theorem 1: Let $\mathbf{W}_{i_k}\succeq\mathbf{0}$ be a weight matrix for receiver i_k. The problem

$$\begin{aligned}\min_{\mathbf{W},\mathbf{U},\mathbf{V}}\quad & \sum_{k=1}^{K}\sum_{i=1}^{I_k}\alpha_{i_k}\left(\mathrm{Tr}(\mathbf{W}_{i_k}\mathbf{E}_{i_k})-\log\det(\mathbf{W}_{i_k})\right)\\ \text{s.t.}\quad & \sum_{i=1}^{I_k}\mathrm{Tr}(\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H)\le P_k,k=1,2,\ldots,K\end{aligned}\tag{7}$$

is equivalent to the weighted sum-rate maximization problem (1), in the sense that the global optimal solution $\mathbf{V}$ for the two problems are identical.

The proof of Theorem 1 is relegated to the Appendix A. To gain some insight, let us consider the special case of a SISO interference channel where the above equivalence relation can be seen more directly. In particular, all the channel matrices $\mathbf{H}_{i_kj}$ in this case become scalars which we denote by h_{i_kj}. Then, the sum-rate maximization problem (1) can be simplified as

$$\begin{aligned}\max_{v}\quad & \sum_{k=1}^{K}\sum_{i=1}^{I_k}\log\left(1+\frac{|h_{i_kk}|^2|v_{i_k}|^2}{\sum\limits_{(j,\ell)\neq(k,i)}|h_{i_kj}|^2|v_{\ell_j}|^2+\sigma_{i_k}^2}\right)\\ \text{s.t.}\quad & \sum_{i=1}^{I_k}|v_{i_k}|^2\le P_k,k=1,2,\ldots,K.\end{aligned}\tag{8}$$

$$R_{i_k}\triangleq\log\det\left(\mathbf{I}+\mathbf{H}_{i_kk}\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H\mathbf{H}_{i_kk}^H\left(\sum_{(\ell,j)\neq(i,k)}\mathbf{H}_{i_kj}\mathbf{V}_{\ell_j}\mathbf{V}_{\ell_j}^H\mathbf{H}_{i_kj}^H+\sigma_{i_k}^2\mathbf{I}\right)^{-1}\right)\tag{2}$$

This sum-rate maximization problem is equivalent to the following weighted sum-MSE minimization problem:

$$\begin{aligned}\min_{w,u,v} \quad & \sum_{k=1}^{K}\sum_{i=1}^{I_k}(w_{i_k}e_{i_k} - \log w_{i_k}) \\ \text{s.t.} \quad & \sum_{i=1}^{I_k}|v_{i_k}|^2 \le P_k, k=1,2,\ldots,K \end{aligned} \tag{9}$$

where w_{i_k} is a positive weight variable, and e_{i_k} is the mean-square estimation error

$$e_{i_k} \triangleq |u_{i_k}h_{i_k k}v_{i_k} - 1|^2 + \sum_{(j,\ell)\neq(k,i)} |u_{i_k}h_{i_k j}v_{\ell_j}|^2 + \sigma_{i_k}^2|u_{i_k}|^2.$$

To see the equivalence, we can check the first optimality condition to find the optimal w_{i_k} and u_{i_k}

$$\begin{aligned} u_{i_k}^{\text{opt}} &= \frac{h_{i_k k}v_{i_k}}{\sum_{j=1}^{K}\sum_{\ell=1}^{I_j}|h_{i_k j}|^2|v_{\ell_j}|^2 + \sigma_{i_k}^2}, \\ w_{i_k}^{\text{opt}} &= e_{i_k}^{-1}, \quad \forall i_k \in \mathcal{I}. \end{aligned}$$

Plugging these optimal values in e_{i_k} and simplifying (9) gives the following equivalent optimization problem:

$$\begin{aligned}\max_{v} \quad & \sum_{k=1}^{K}\sum_{i=1}^{I_k}\log\left(1 - \frac{|h_{i_k k}|^2|v_{i_k}|^2}{\sum_{j=1}^{K}\sum_{\ell=1}^{I_j}|h_{i_k j}|^2|v_{\ell_j}|^2 + \sigma_k^2}\right)^{-1} \\ \text{s.t.} \quad & \sum_{i=1}^{I_k}|v_{i_k}|^2 \le P_k, k=1,2,\ldots,K \end{aligned}$$

which is further equivalent to (8).

The equivalence relation implies that maximizing sum-rate can be accomplished via weighted MSE minimization (9). The latter problem is in the space of (u,v,w) and is easier to handle since optimizing each variable while holding others fixed is convex and easy (e.g., closed form). This property will be exploited in Section III when we design the WMMSE algorithm. In contrast, the original sum-rate maximization problem (8) is nonconvex in the design variable v, which makes the direct optimization of (8) difficult.

B. General Utility Maximization

The equivalence of the sum-rate maximization problem (1) and the matrix-weighted MSE minimization problem (7) can be extended to other system utility functions reflecting user fairness.

Let $u_{i_k}(\cdot)$ be an increasing utility function of the receiver i_k's data rate (denoted by R_{i_k}, cf. (2)). We consider the sum-utility maximization problem

$$\begin{aligned}\max_{\mathbf{V}} \quad & \sum_{k=1}^{K}\sum_{i_k=1}^{I_k}u_{i_k}(R_{i_k}) \\ \text{s.t.} \quad & \sum_{i=1}^{I_k}\operatorname{Tr}(\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H) \le P_k, k=1,2,\ldots,K. \end{aligned} \tag{10}$$

Recall the well-known relation between the MSE covariance matrix $\mathbf{E}_{i_k}^{\text{mmse}}$ and the rate R_{i_k} (see (21) in Appendix A)

$$R_{i_k} = \log\det\left(\left(\mathbf{E}_{i_k}^{\text{mmse}}\right)^{-1}\right).$$

Thus, the sum-utility maximization problem (10) can also be written as the following sum-MSE cost minimization problem:

$$\begin{aligned}\min_{\mathbf{V},\mathbf{U}} \quad & \sum_{k=1}^{K}\sum_{i=1}^{I_k}c_{i_k}(\mathbf{E}_{i_k}) \\ \text{s.t.} \quad & \sum_{i=1}^{I_k}\operatorname{Tr}(\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H) \le P_k, k=1,2,\ldots,K, \end{aligned} \tag{11}$$

and $\mathbf{E}_{i_k}$ is given by (3) where $\mathbf{E}_{i_k}$ and $c_{i_k}(\mathbf{E}_{i_k}) = -u_{i_k}(-\log\det(\mathbf{E}_{i_k}))$ are the MSE covariance matrix and the cost function of receiver i_k respectively. Similar to Theorem 1, we introduce auxiliary weight matrix variables $\{\mathbf{W}_{i_k}\}_{i_k\in\mathcal{I}}$ and define the following matrix-weighted sum-MSE minimization problem:

$$\begin{aligned}\min_{\mathbf{V},\mathbf{U},\mathbf{W}} \quad & \sum_{k=1}^{K}\sum_{i=1}^{I_k}\Big(\operatorname{Tr}\left(\mathbf{W}_{i_k}^H\mathbf{E}_{i_k}\right) + c_{i_k}\left(\boldsymbol{\gamma}_{i_k}(\mathbf{W}_{i_k})\right) \\ & - .\operatorname{Tr}\left(\mathbf{W}_{i_k}^H\boldsymbol{\gamma}_{i_k}(\mathbf{W}_{i_k})\right)\Big) \\ \text{s.t.} \quad & \sum_{i=1}^{I_k}\operatorname{Tr}(\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H) \le P_k, k=1,2,\ldots,K \end{aligned} \tag{12}$$

where $\boldsymbol{\gamma}_{i_k}(\cdot): \mathbb{R}^{d_{i_k}\times d_{i_k}} \mapsto \mathbb{R}^{d_{i_k}\times d_{i_k}}$ is the inverse mapping of the gradient map $\nabla c_{i_k}(\mathbf{E}_{i_k})$. The following theorem provides a sufficient condition for the equivalence of (12) and (11).

Theorem 2: Suppose $c_{i_k}(\cdot) = -u_{i_k}(-\log\det(\cdot))$ is a strictly concave function for all i_k. Then the inverse gradient map $\boldsymbol{\gamma}_{i_k}(\cdot)$ is well-defined, and for any fixed transmit/receive beamformers $\{\mathbf{V},\mathbf{U}\}$, the optimal weight matrix $\mathbf{W}_{i_k}$ for (12) is given by $\mathbf{W}_{i_k}^{\text{opt}} = \nabla c_{i_k}(\mathbf{E}_{i_k})$. Moreover, the sum-utility maximization problem (10) is equivalent to the matrix-weighted sum-MSE minimization problem (12) in the sense that they have the same global optimal solution.

The proof of Theorem 2 is relegated to Appendix B. Weighted sum-rate maximization problem, weighted sum SINR maximization, geometric mean maximization of one plus rates are examples of utility functions that satisfy the conditions of Theorem 2. In particular, for the sum-rate utility, the proportional fairness utility and the harmonic mean rate utility, we respectively have $u_{i_k}(R_{i_k}) = \alpha_{i_k}R_{i_k}$, $u_{i_k}(R_{i_k}) = \log R_{i_k}$ and $u_{i_k}(R_{i_k}) = -R_{i_k}^{-1}$. It can be checked that in each of these cases, the resulting cost function $c_{i_k}(\cdot) = -u_{i_k}(-\log\det(\cdot))$ is strictly concave, so Theorem 2 is applicable.

III. WMMSE Algorithm for Sum-Utility Maximization

In this section, we exploit the equivalence relation developed in Section II (Theorems 1 and 2) to design a simple distributed WMMSE algorithm for the sum-utility maximization problem (10). To minimize notations, we first focus our presentation for the sum-rate maximization problem (1); the extensions to general sum-utility maximization will be described later.

By Theorem 1, we only need to solve the equivalent sum-MSE minimization problem (7). Since the cost function of (7) is convex in each of the optimization variables $\mathbf{U}$, $\mathbf{V}$, $\mathbf{W}$, we propose to use the block coordinate descent method to solve (7). In particular, we minimize the weighted sum-MSE cost function by sequentially fixing two of the three variables $\mathbf{U}$, $\mathbf{V}$, $\mathbf{W}$ and updating the third. The update of the weight matrix variable $\mathbf{W}_{i_k}$ is in closed form (see (19) in Appendix A and (3)) which is given by

$$\mathbf{W}_{i_k}^{\mathrm{opt}} = \mathbf{E}_{i_k}^{-1} \tag{13}$$

while the update of receive beamformer $\mathbf{U}_{i_k}$ is given by the MMSE solution (5). The update of transmit beamformers $\mathbf{V}_{i_k}$ for all i_k can also be decoupled across transmitters, resulting in the following optimization problem:

$$\begin{aligned} \min_{\{\mathbf{V}_{i_k}\}_{i=1}^{I_k}} & \sum_{i=1}^{I_k} \mathrm{Tr}\left(\alpha_{i_k}\mathbf{W}_{i_k}(\mathbf{I} - \mathbf{U}_{i_k}^H\mathbf{H}_{i_k k}\mathbf{V}_{i_k}) \right. \\ & \left. \times (\mathbf{I} - \mathbf{U}_{i_k}^H\mathbf{H}_{i_k k}\mathbf{V}_{i_k})^H\right) \\ & + \sum_{i=1}^{I_k}\sum_{(\ell,j)\neq(i,k)} \mathrm{Tr}(\alpha_{\ell_j}\mathbf{W}_{\ell_j}\mathbf{U}_{\ell_j}^H\mathbf{H}_{\ell_j k}\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H\mathbf{H}_{\ell_j k}^H\mathbf{U}_{\ell_j}^H) \\ \text{s.t.} & \sum_{i=1}^{I_k} \mathrm{Tr}(\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H) \le P_k. \end{aligned} \tag{14}$$

This is a convex quadratic optimization problem which can be solved by using standard convex optimization algorithms. In fact, this problem also has a closed form solution using the Lagrange multipliers method. Specifically, attaching a Lagrange multiplier μ_k to the power budget constraint of transmitter k, we get the following Lagrange function:

$$\begin{aligned} & L(\{\mathbf{V}_{i_k}\}_{i=1}^{I_k}, \mu_k) \\ & \triangleq \sum_{i=1}^{I_k} \mathrm{Tr}\left(\alpha_{i_k}\mathbf{W}_{i_k}(\mathbf{I} - \mathbf{U}_{i_k}^H\mathbf{H}_{i_k k}\mathbf{V}_{i_k})(\mathbf{I} - \mathbf{U}_{i_k}^H\mathbf{H}_{i_k k}\mathbf{V}_{i_k})^H\right) \\ & \quad + \sum_{i=1}^{I_k}\sum_{(\ell,j)\neq(i,k)} \mathrm{Tr}\left(\alpha_{\ell j}\mathbf{W}_{\ell_j}\mathbf{U}_{\ell_j}^H\mathbf{H}_{\ell_j k}\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H\mathbf{H}_{\ell_j k}^H\mathbf{U}_{\ell_j}^H\right) \\ & \quad + \mu_k\Big(\sum_{i=1}^{I_k}\mathrm{Tr}\left(\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H\right) - P_k\Big). \end{aligned}$$

The first-order optimality condition of $L(\{\mathbf{V}_{i_k}\}_{i=1}^{I_k},\ \mu_k)$ with respect to each $\mathbf{V}_{i_k}$ yields

$$\begin{aligned} \mathbf{V}_{i_k}^{\mathrm{opt}} = & \left(\sum_{j=1}^{K}\sum_{\ell=1}^{I_j}\alpha_{\ell j}\mathbf{H}_{\ell_j k}^H\mathbf{U}_{\ell_j}\mathbf{W}_{\ell_j}\mathbf{U}_{\ell_j}^H\mathbf{H}_{\ell_j k} + \mu_k\mathbf{I}\right)^{-1} \\ & \times\alpha_{i_k}\mathbf{H}_{i_k k}^H\mathbf{U}_{i_k}\mathbf{W}_{i_k}, i = 1,\ldots,I_k \end{aligned} \tag{15}$$

where $\mu_k \ge 0$ should be chosen such that the complementarity slackness condition of the power budget constraint is satisfied. Let $\mathbf{V}_{i_k}(\mu_k)$ denote the right-hand side of (15). When the matrix $\sum_{j=1}^{K}\sum_{\ell=1}^{I_j}\alpha_{\ell_j}\mathbf{H}_{\ell_j k}^H\mathbf{U}_{\ell_j}\mathbf{W}_{\ell_j}\mathbf{U}_{\ell_j}^H\mathbf{H}_{\ell_j k}$ is invertible and $\sum_{i=1}^{I_k}\mathrm{Tr}(\mathbf{V}_{i_k}(0)\mathbf{V}_{i_k}(0)^H) \le P_k$, then $\mathbf{V}_{i_k}^{\mathrm{opt}} = \mathbf{V}_{i_k}(0)$, otherwise we must have

$$\sum_{i=1}^{I_k}\mathrm{Tr}\left(\mathbf{V}_{i_k}(\mu_k)\mathbf{V}_{i_k}(\mu_k)^H\right) = P_k \tag{16}$$

TABLE I
PSEUDO CODE OF THE PROPOSED WMMSE ALGORITHM FOR MIMO-IBC

1 Initialize $\mathbf{V}_{i_k}$'s such that $\mathrm{Tr}\left(\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H\right) = \frac{p_k}{I_k}$
2 **repeat**
3 $\quad\mathbf{W}'_{i_k} \leftarrow \mathbf{W}_{i_k}, \quad \forall\, i_k \in \mathcal{I}$
4 $\quad\mathbf{U}_{i_k} \leftarrow \left(\sum_{(j,\ell)}\mathbf{H}_{i_k j}\mathbf{V}_{\ell_j}\mathbf{V}_{\ell_j}^H\mathbf{H}_{i_k j}^H + \sigma_{i_k}^2\mathbf{I}\right)^{-1}\mathbf{H}_{i_k k}\mathbf{V}_k,\ \forall\, i_k \in \mathcal{I}$
5 $\quad\mathbf{W}_{i_k} \leftarrow \left(\mathbf{I} - \mathbf{U}_{i_k}^H\mathbf{H}_{i_k k}\mathbf{V}_{i_k}\right)^{-1},\ \forall\, i_k \in \mathcal{I}$
6 $\quad\mathbf{V}_{i_k} \leftarrow \alpha_{i_k}\left(\sum_{(j,\ell)}\alpha_{\ell_j}\mathbf{H}_{\ell_j k}^H\mathbf{U}_{\ell_j}\mathbf{W}_{\ell_j}\mathbf{U}_{\ell_j}^H\mathbf{H}_{\ell_j k} + \mu_k^*\mathbf{I}\right)^{-1}\mathbf{H}_{i_k k}^H\mathbf{U}_{i_k}\mathbf{W}_{i_k},\ \forall\, i_k$
7 **until** $\left|\sum_{(j,\ell)}\log\det\left(\mathbf{W}_{\ell_j}\right) - \sum_{(j,\ell)}\log\det\left(\mathbf{W}'_{\ell_j}\right)\right| \le \epsilon$

which is equivalent to

$$\mathrm{Tr}\left((\mathbf{\Lambda} + \mu_k\mathbf{I})^{-2}\mathbf{\Phi}\right) = P_k \tag{17}$$

where $\mathbf{D}\mathbf{\Lambda}\mathbf{D}^H$ is the eigendecomposition of $\sum_{j=1}^{K}\sum_{\ell=1}^{I_j}\mathbf{H}_{\ell_j k}^H\mathbf{U}_{\ell_j}\bar{\mathbf{W}}_{\ell_j}\mathbf{U}_{\ell_j}^H\mathbf{H}_{\ell_j k}$ and $\mathbf{\Phi} = \mathbf{D}^H\Big(\sum_{i=1}^{I_k} \times \mathbf{H}_{i_k k}^H\mathbf{U}_{i_k}\bar{\mathbf{W}}_{i_k}^2\mathbf{U}_{i_k}^H\mathbf{H}_{i_k k}\Big)\mathbf{D}$. Let $[\mathbf{X}]_{mm}$ denote the mth diagonal element of $\mathbf{X}$, then (17) can be simplified as

$$\sum_{m=1}^{M_k}\frac{[\mathbf{\Phi}]_{mm}}{([\mathbf{\Lambda}]_{mm} + \mu_k)^2} = P_k. \tag{18}$$

Note that the optimum μ_k (denoted by μ_k^*) must be positive in this case and the left-hand side of (18) is a decreasing function in μ_k for $\mu_k > 0$. Hence, (18) can be easily solved using one dimensional search techniques (e.g., bisection method). Finally, by plugging μ_k^* in (15), we get the solution for $\mathbf{V}_{i_k}(\mu_k^*)$, for all $i = 1,\ldots,I_k$.

The WMMSE algorithm for the MIMO-IBC is summarized in Table I. The next result shows that the WMMSE algorithm is guaranteed to converge to a stationary point of (1).

Theorem 3: Any limit point $(\mathbf{W}^*, \mathbf{U}^*, \mathbf{V}^*)$ of the iterates generated by the WMMSE algorithm is a stationary point of (7), and the corresponding $\mathbf{V}^*$ is a stationary point of (1). Conversely, if $\mathbf{V}^*$ is a stationary point of (1), then the point $(\mathbf{W}^*, \mathbf{U}^*, \mathbf{V}^*)$, with $\mathbf{W}_{i_k}^*$ and $\mathbf{U}_{i_k}^*$ defined by (13) and (5) respectively, is a stationary point of (7).

We relegate the proof of Theorem 3 to Appendix C. We remark that both the WMMSE algorithm and its convergence result (Theorem 3) can be extended to the general sum-utility maximization problem (10). In particular, we only need to replace line 5 of the WMMSE algorithm in Table I by "$\mathbf{W}_{i_k} \leftarrow \nabla c_{i_k}(\mathbf{E}_{i_k})$, $\forall i_k$" (see Theorem 2), while the resulting algorithm is guaranteed to converge to a stationary point of (12) and (10). It is important to note that the WMMSE algorithm for general sum-utility maximization does not require the explicit knowledge of the inverse gradient map $\boldsymbol{\gamma}_{i_k}(\cdot)$, even though (12) is defined in terms of $\boldsymbol{\gamma}_{i_k}(\cdot)$.

IV. DISTRIBUTED IMPLEMENTATION AND COMPLEXITY ANALYSIS

For the purpose of distributed implementation, we make two reasonable assumptions (similar to [11]). First, we assume that local channel state information is available for each user, namely, each transmitter k knows the local channel matrices

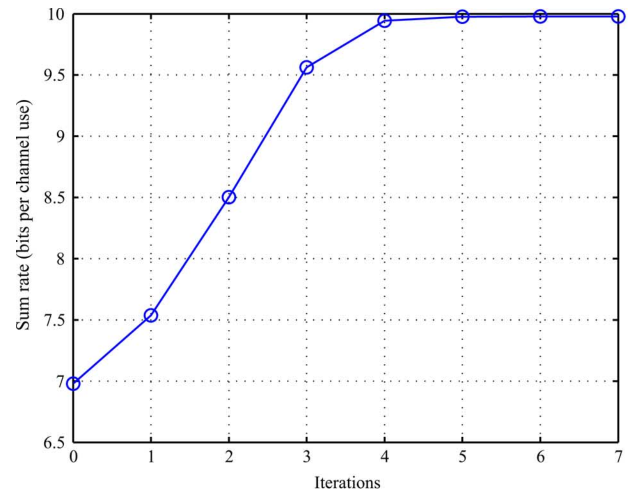
$\mathbf{H}_{\ell_j k}$ to all receivers ℓ_j. The second assumption is that each receiver has an additional channel to feedback information (e.g., the updated beamformers or equivalent information) to the transmitters. Under these two assumptions, the WMMSE algorithm can be implemented in a distributed fashion. More specifically, each receiver i_k locally estimates the received signal covariance matrix $\mathbf{J}_{i_k}$ and updates the matrices $\mathbf{U}_{i_k}$ and $\mathbf{W}_{i_k}$. Then, it feeds back the updated $\mathbf{W}_{i_k}$ and $\mathbf{U}_{i_k}$ to the transmitters. Note that, to reduce communication overhead, user i_k only needs to feedback either the upper triangular part of the matrix $\alpha_{i_k}\mathbf{U}_{i_k}\mathbf{W}_{i_k}\mathbf{U}_{i_k}^H$ or the decomposition $\hat{\mathbf{U}}_{i_k}$ where $\hat{\mathbf{U}}_{i_k}\hat{\mathbf{U}}_{i_k}^H = \alpha_{i_k}\mathbf{U}_{i_k}\mathbf{W}_{i_k}\mathbf{U}_{i_k}^H$ (depending on the relative size of N_{i_k} and d_{i_k}). It should be pointed out that the termination criterion in Table I may not be suitable for distributed implementation. In practice, we suggest setting a maximum number of iterations for the algorithm or simply just do one step of the algorithm within each packet.

Note that the ILA algorithm [9], [12] allows only one user to update its transmit covariance matrix at each iteration. When one user updates its variables, each user must compute $(K-1)$ prices[9] or gradient matrices[12] for other users and then broadcast them within the network. In contrast, the WMMSE algorithm allows simultaneous update among all users since the updating steps are decoupled across users when any of the two variables in $\{\mathbf{U}, \mathbf{V}, \mathbf{W}\}$ are fixed. Therefore, the WMMSE algorithm requires less CSI exchange within the network. For simplicity of complexity analysis, let $\kappa \triangleq |\mathcal{I}|$ be the total number of users in the system and T, R denote the number of antennas at each transmitter and receiver respectively. Also, since both the WMMSE algorithm and the ILA algorithm include a bisection step which generally takes few iterations, we ignore this bisection step in the complexity analysis. Under these assumptions, each iteration of the ILA algorithm involves only the computation of the price matrices in [12] (i.e., $\mathbf{A}_i$'s in (10) of [12]). To determine the price matrices in the ILA algorithm, we need to first calculate the covariance matrix of interference at all users and then compute their sum, yielding a complexity of $\mathcal{O}(\kappa^2)$ per user. As a result, the per-iteration complexity of the ILA algorithm is $\mathcal{O}(\kappa^3T^2R + \kappa^3R^2T + \kappa^2R^3)$. By a similar analysis, the per-iteration complexity of the WMMSE algorithm can be shown to be $\mathcal{O}(\kappa^2TR^2 + \kappa^2RT^2 + \kappa^2T^3 + \kappa R^3)$. Here, an iteration of the WMMSE or the ILA algorithm means one round of updating all users' beamformers or covariance matrices.

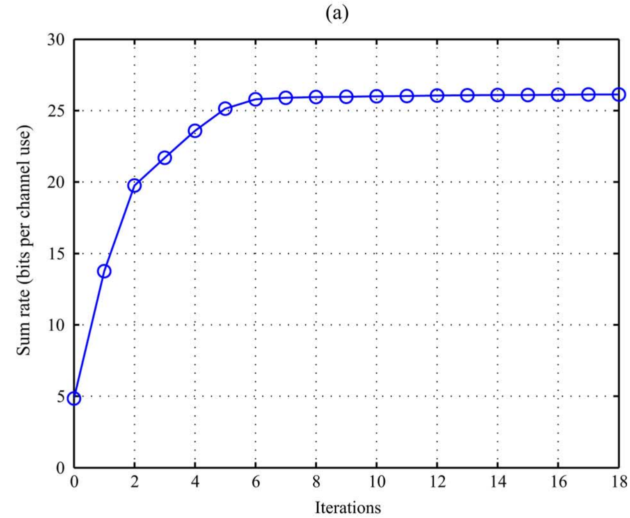
V. Simulation Results

In this section, we numerically evaluate the performance of the proposed WMMSE algorithm. For ease of comparison with existing algorithms, all simulations are conducted for MIMO interference channel (the degenerate MIMO-IBC case with one receiver per cell). The weights $\{\alpha_{i_k}\}$ and noise powers $\{\sigma_{i_k}^2\}$ are set equally for all users. The transmit power budget is set to P for all transmitters, where $P = 10^{\frac{\mathrm{SNR}}{10}}$. Moreover, all transmitters (or receivers) are assumed to have the same number of antennas, denoted by T (or R). We use uncorrelated fading channel model with channel coefficients generated from the complex Gaussian distribution $\mathcal{CN}(0,1)$.

Fig. 1(a) and (b) illustrates the convergence behavior of the WMMSE algorithm for the case of $\mathrm{SNR} = 25$ (dB). These plots show that the WMMSE algorithm converges in few steps and it does so monotonically.



(a)



(b)

Fig. 1. Convergence examples of the WMMSE algorithm: (a) SISO-IFC, $K = 3$, $\epsilon = 1e-3$; (b) MIMO-IFC, $K = 4$, $T = 3$, $R = 2$, $\epsilon = 1e-2$.

Fig. 2 plots the average sum-rate versus the SNR for the SISO interference channel case. Each curve is averaged over 100 random channel realizations. The term "WMMSE" represents running the WMMSE algorithm once while "WMMSE_10rand_int" means running the WMMSE algorithm ten times with different initialization and then keeping the best result. The terms "ILA" and "ILA_10rand_int" are similarly defined. It can be observed that the WMMSE algorithm and the ILA algorithm yield almost the same performance. The performance of the brute force search method (exponential complexity) is provided in the three users case as a benchmark. We can see that the gap between the performance of the WMMSE algorithm and the optimal performance is small and slowly increasing with SNR. However, repeating the WMMSE algorithm ten times can close this performance gap.

Similar observations can be made for the MIMO interference channel case, as Fig. 3 illustrates. As a comparison, we also provide the performance of the MMSE algorithm [21] which has been shown to perform better than the interference alignment

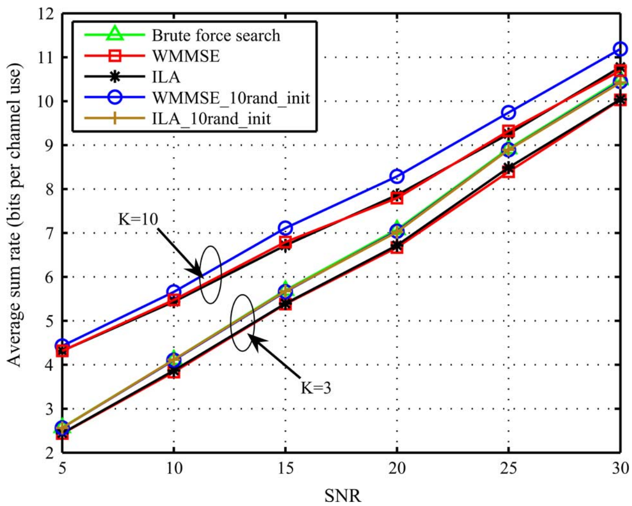


Fig. 2. Average sum-rate versus SNR in the SISO IFC case.

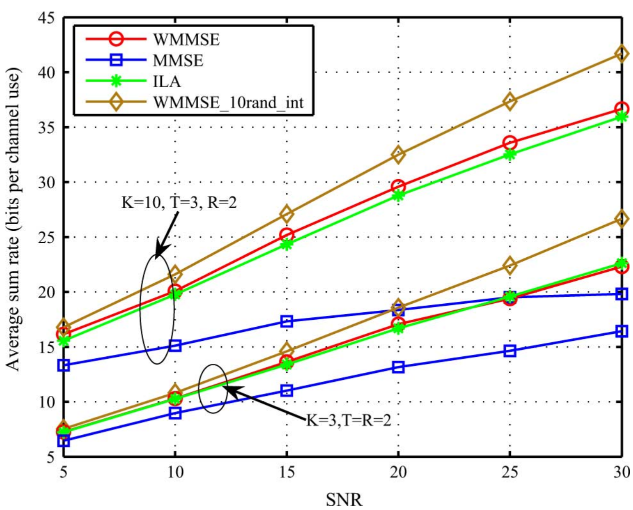


Fig. 3. Average sum-rate versus SNR in the MIMO IFC case.

method [22]. Obviously, the WMMSE algorithm significantly outperforms the MMSE algorithm in terms of the achieved sum-rate. This is due to the use of iterative weighting matrices $\mathbf{W}$.

Although the ILA algorithm yields almost the same performance as the WMMSE algorithm in terms of the sum rate, it has higher complexity. Fig. 4 represents the average CPU time comparison of the two algorithms under the same termination criterion. It can be observed that the WMMSE algorithm significantly outperform the ILA algorithm when the number of users is large.

VI. Concluding Remarks

Coordinated beamforming is an important signal processing technique for interference mitigation in a cellular communication system. Many of the linear transceiver design problems can be formulated as a sum-utility maximization problem. Introducing a weight matrix variable, we transform the sum-utility maximization problem for a MIMO-IBC to an equivalent sum-MSE cost minimization problem. The latter formulation is amenable to a simple distributed block coordinate descent

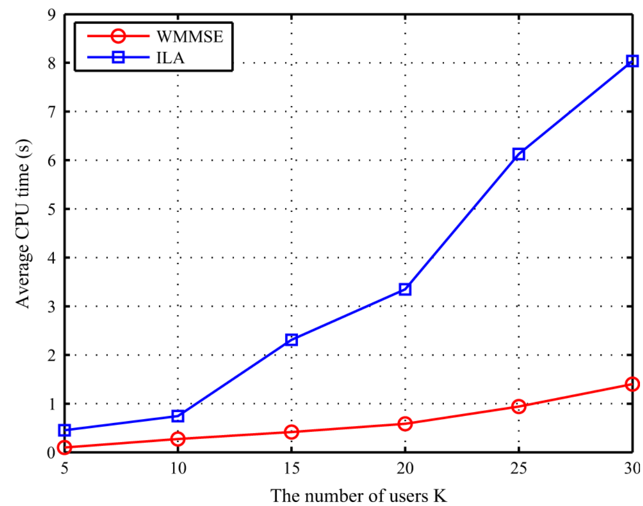


Fig. 4. Average CPU time versus the number of users in the MIMO IFC case ($T = 3$, $R = 2$).

minimization strategy, which we call the WMMSE method. This method has a low per-iteration complexity and is guaranteed to converge to a stationary point of the original sum-utility maximization problem. The WMMSE method is very versatile as it is applicable to a broad class of sum-utility maximization problems, including the sum-rate maximization, harmonic mean maximization of one plus rate, sum SINR maximization and sum-MSE minimization, just to name a few. In computer simulations, the WMMSE method shows a substantial gain in performance and computational efficiency over the existing approaches for coordinated beamforming and interference alignment. We expect the WMMSE method to be an effective tool in other application contexts such as dynamic spectrum management for DSL networks, or distributed power control in wireless networks.

Appendix A
Proof of Theorem 1

Proof: First of all, it can be seen that the optimum $\mathbf{U}_{i_k}$ for minimizing (7) is given by $\mathbf{U}_{i_k}^{\mathrm{mmse}}$ in (5). Furthermore, fixing the other variables, the objective function in (7) is convex with respect to $\mathbf{W}_{i_k}$. Therefore, by checking the first order optimality condition for $\mathbf{W}_{i_k}$, we get

$$\mathbf{W}_{i_k}^{\mathrm{opt}} = \mathbf{E}_{i_k}^{-1}. \tag{19}$$

Substituting the optimal $\mathbf{U}_{i_k}$ and $\mathbf{W}_{i_k}$, for all $i_k \in \mathcal{I}$, in (7), we have the following equivalent optimization problem:

$$\begin{aligned} \max_{\{\mathbf{V}\}} \quad & \sum_{k=1}^{K}\sum_{i=1}^{I_k} \alpha_{i_k} \log\det\left((\mathbf{E}_{i_k}^{\mathrm{mmse}})^{-1}\right) \\ \text{s.t.} \quad & \mathrm{Tr}(\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H) \le P_k, \ \forall i_k \in \mathcal{I}. \end{aligned} \tag{20}$$

Defining $\mathbf{\Upsilon}_{i_k} \triangleq \sum_{(\ell,j)\neq(i,k)} \mathbf{H}_{i_k j}\mathbf{V}_{\ell_j}\mathbf{V}_{\ell_j}^H\mathbf{H}_{i_k j}^H + \sigma_{i_k}^2\mathbf{I}$, we get

$$\begin{aligned} \log\det\left(\left(\mathbf{E}_{i_k}^{\mathrm{mmse}}\right)^{-1}\right) &= \log\det\left(\mathbf{I} + \mathbf{V}_{i_k}^H\mathbf{H}_{i_k k}^H\mathbf{\Upsilon}_{i_k}^{-1}\mathbf{H}_{i_k k}\mathbf{V}_{i_k}\right) \\ &= \log\det(\mathbf{I} + \mathbf{H}_{i_k k}\mathbf{Q}_{i_k}\mathbf{H}_{i_k k}^H\mathbf{\Upsilon}_{i_k}^{-1}) \end{aligned} \tag{21}$$

where the first equality follows from applying the Woodbury matrix identity to (6), while the second equality is due to the fact that $\det(\mathbf{I}+\mathbf{A}_1\mathbf{A}_2)=\det(\mathbf{I}+\mathbf{A}_2\mathbf{A}_1)$. Combining (21) with (20) completes the proof. ■

Appendix B
Proof of Theorem 2

Proof: To simplify notations, we drop the subscript i_k. Note, here the notation $\mathbf{W}$ is short for $\mathbf{W}_{i_k}$. We first argue that $\nabla c(\mathbf{E})$ is an invertible mapping, so its inverse $\boldsymbol{\gamma}(\cdot)$ is well-defined. Assume the contrary, i.e., for two different MSE values $\mathbf{E}_1$ and $\mathbf{E}_2$, $\nabla c(\mathbf{E}_1)=\nabla c(\mathbf{E}_2)$. Since $c(\cdot)$ is strictly concave, we have

$$c(\mathbf{E}_1)<c(\mathbf{E}_2)+\mathrm{Tr}\left((\nabla c(\mathbf{E}_2))^H(\mathbf{E}_1-\mathbf{E}_2)\right),$$
$$c(\mathbf{E}_2)<c(\mathbf{E}_1)+\mathrm{Tr}\left((\nabla c(\mathbf{E}_1))^H(\mathbf{E}_2-\mathbf{E}_1)\right).$$

Summing up these two inequalities immediately yields a contradiction. Next we prove that $g(\mathbf{W})\triangleq c(\boldsymbol{\gamma}(\mathbf{W}))-\mathrm{Tr}\left(\mathbf{W}^H\boldsymbol{\gamma}(\mathbf{W})\right)$ is strictly convex. To this end, we first calculate the gradient of $g(\cdot)$:

$$\begin{aligned}\nabla g(\mathbf{W})&=\sum_{i,j}\frac{\partial c}{\partial\boldsymbol{\gamma}_{i,j}}\nabla\boldsymbol{\gamma}_{i,j}(\mathbf{W})-\nabla_{\mathbf{W}}\left(\sum_{i,j}\mathbf{W}_{i,j}\boldsymbol{\gamma}_{i,j}(\mathbf{W})\right)\\&=\sum_{i,j}\frac{\partial c}{\partial\boldsymbol{\gamma}_{i,j}}\nabla\boldsymbol{\gamma}_{i,j}(\mathbf{W})-\sum_{i,j}\nabla(\mathbf{W}_{i,j})\boldsymbol{\gamma}_{i,j}(\mathbf{W})\\&\quad-\sum_{i,j}\mathbf{W}_{i,j}\nabla\boldsymbol{\gamma}_{i,j}(\mathbf{W})\\&=\sum_{i,j}\mathbf{W}_{i,j}\nabla\boldsymbol{\gamma}_{i,j}(\mathbf{W})-\sum_{i,j}\mathbf{1}_{i,j}\boldsymbol{\gamma}_{i,j}(\mathbf{W})\\&\quad-\sum_{i,j}\mathbf{W}_{i,j}\nabla\boldsymbol{\gamma}_{i,j}(\mathbf{W})\\&=-\boldsymbol{\gamma}(\mathbf{W})\end{aligned}\tag{22}$$

where $\mathbf{X}_{i,j}$ denotes the (i,j)th element of the matrix $\mathbf{X}$ and $\mathbf{1}_{i,j}$ is an all zero matrix except the (i,j)th element which is one. For any fixed point $\mathbf{W}$ and any feasible direction $\mathbf{Z}$, let us define $q(t;\mathbf{W},\mathbf{Z})=g(\mathbf{W}+t\mathbf{Z})$. In order to prove the strict convexity of $g(\cdot)$, it suffices to check the strict convexity of $q(t;\mathbf{W},\mathbf{Z})$ for any fixed $\mathbf{W}$ and $\mathbf{Z}$. Using the definition of $q(t;\mathbf{W},\mathbf{Z})$, we have (23), shown at the bottom of the page, where the inequality is due to strict concavity of $c(\cdot)$. The above overall inequality shows the strict convexity of $q(t;\mathbf{W},\mathbf{Z})$ in t. Since $q(\cdot;\mathbf{W},\mathbf{Z})$ is strictly convex for any fixed $\mathbf{W}$ and $\mathbf{Z}$, $g(\mathbf{W})$ is strictly convex. Therefore, the objective function of (12) is strictly convex with respect to $\mathbf{W}$. Using (22), we have $\mathbf{E}_{i_k}-\boldsymbol{\gamma}(\mathbf{W}_{i_k}^{\mathrm{opt}})=0$, which further implies $\mathbf{W}_{i_k}^{\mathrm{opt}}=\nabla c_{i_k}(\mathbf{E}_{i_k})$ for all i_k. By plugging back this optimum value of $\mathbf{W}$ in (12), we arrive at the equivalence of (12) and (11).

Appendix C
Proof of Theorem 3

Proof: The optimization problem (7) has a differentiable objective function and a constraint set that is separable in the variables $\mathbf{W}$, $\mathbf{U}$ and $\mathbf{V}$. It follows from the general optimization theory [23] that the WMMSE algorithm, which is the block coordinate descent method applied to (7), converges to a stationary point of (7). It remains to verify that $\mathbf{V}^*$ is a stationary point of (1) if and only if $(\mathbf{W}^*,\mathbf{U}^*,\mathbf{V}^*)$ be a stationary point of (7) for some $\mathbf{W}^*$ and $\mathbf{U}^*$. Let us define

$$\begin{aligned}&\psi_1(\mathbf{W},\mathbf{U},\mathbf{V})\\&\quad\triangleq\sum_{k=1}^{K}\sum_{i=1}^{I_k}\alpha_{i_k}\left(\mathrm{Tr}(\mathbf{W}_{i_k}\mathbf{E}_{i_k})-\log\det(\mathbf{W}_{i_k})\right)\\&\psi_2(\mathbf{V})\triangleq\sum_{k=1}^{K}\sum_{i=1}^{I_k}\alpha_{i_k}\log\det(\mathbf{E}_{i_k}^{\mathrm{mmse}}).\end{aligned}$$

Since $(\mathbf{W}^*,\mathbf{U}^*,\mathbf{V}^*)$ is a stationary point of (7) and the constraint of (7) are in the form of Cartesian product, we have

$$\mathrm{Tr}\left(\nabla_{\mathbf{U}_{i_k}}\psi_1(\mathbf{W}^*,\mathbf{U}^*,\mathbf{V}^*)^H(\mathbf{U}_{i_k}-\mathbf{U}_{i_k}^*)\right)\le 0,\ \ \forall\mathbf{U}_{i_k},\ \forall i_k\tag{24}$$

$$\mathrm{Tr}\left(\nabla_{\mathbf{W}_{i_k}}\psi_1(\mathbf{W}^*,\mathbf{U}^*,\mathbf{V}^*)^H(\mathbf{W}_{i_k}-\mathbf{W}_{i_k}^*)\right)\le 0,\ \ \forall\mathbf{W}_{i_k},\ \forall i_k\tag{25}$$

$$\mathrm{Tr}\left(\nabla_{\mathbf{V}}\psi_1(\mathbf{W}^*,\mathbf{U}^*,\mathbf{V}^*)^H(\mathbf{V}-\mathbf{V}^*)\right)\le 0,\quad\forall\mathbf{V}\in\mathbb{S}\tag{26}$$

$$\begin{aligned}q(t_2;\mathbf{W},\mathbf{Z})&=g(\mathbf{W}+t_2\mathbf{Z})\\&=c(\boldsymbol{\gamma}(\mathbf{W}+t_2\mathbf{Z}))-\mathrm{Tr}\left((\mathbf{W}+t_2\mathbf{Z})^H\boldsymbol{\gamma}(\mathbf{W}+t_2\mathbf{Z})\right)\\&>c(\boldsymbol{\gamma}(\mathbf{W}+t_1\mathbf{Z}))-\mathrm{Tr}\left((\nabla c(\boldsymbol{\gamma}(\mathbf{W}+t_2\mathbf{Z})))^H\left(\boldsymbol{\gamma}(\mathbf{W}+t_1\mathbf{Z})-\boldsymbol{\gamma}(\mathbf{W}+t_2\mathbf{Z})\right)\right)\\&\quad-\mathrm{Tr}\left((\mathbf{W}+t_2\mathbf{Z})^H\boldsymbol{\gamma}(\mathbf{W}+t_2\mathbf{Z})\right)\\&=c(\boldsymbol{\gamma}(\mathbf{W}+t_1\mathbf{Z}))-\mathrm{Tr}\left((\mathbf{W}+t_2\mathbf{Z})^H\left(\boldsymbol{\gamma}(\mathbf{W}+t_1\mathbf{Z})-\boldsymbol{\gamma}(\mathbf{W}+t_2\mathbf{Z})\right)\right)-\mathrm{Tr}\left((\mathbf{W}+t_2\mathbf{Z})^H\boldsymbol{\gamma}(\mathbf{W}+t_2\mathbf{Z})\right)\\&=c(\boldsymbol{\gamma}(\mathbf{W}+t_1\mathbf{Z}))-\mathrm{Tr}\left((\mathbf{W}+t_2\mathbf{Z})^H\boldsymbol{\gamma}(\mathbf{W}+t_1\mathbf{Z})\right)\\&=c(\boldsymbol{\gamma}(\mathbf{W}+t_1\mathbf{Z}))-\mathrm{Tr}\left((\mathbf{W}+t_1\mathbf{Z})^H\boldsymbol{\gamma}(\mathbf{W}+t_1\mathbf{Z})\right)-(t_2-t_1)\mathrm{Tr}\left(\mathbf{Z}^H\boldsymbol{\gamma}(\mathbf{W}+t_1\mathbf{Z})\right)\\&=g(\mathbf{W}+t_1\mathbf{Z})+\mathrm{Tr}\left(\mathbf{Z}^H\nabla g(\mathbf{W}+t_1\mathbf{Z})\right)(t_2-t_1)\\&=q(t_1;\mathbf{W},\mathbf{Z})+q'(t_1;\mathbf{W},\mathbf{Z})\cdot(t_2-t_1)\end{aligned}\tag{23}$$

where $\mathbb{S} = \{\mathbf{V} \mid \sum_{i=1}^{I_k} \mathrm{Tr}(\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H) \leq P_k, \ \forall k\}$ is the feasible set. Since (24) and (25) must hold for every $\mathbf{W}_{i_k}$ and $\mathbf{U}_{i_k}$ (unconstrained), we get

$$\mathbf{U}_{i_k}^* = \mathbf{U}_{i_k}^{\mathrm{mmse}} \ \text{ and } \ \mathbf{W}_{i_k}^* = (\mathbf{E}_{i_k}^{\mathrm{mmse}})^{-1}. \tag{27}$$

Let $v_{\ell_j,mn}$ be the (m,n)th entry of $\mathbf{V}_{\ell_j}$. Using chain rule, we get

$$\begin{aligned} &\frac{\partial \psi_1(\mathbf{V}^*, \mathbf{U}^*, \mathbf{W}^*)}{\partial v_{\ell_j,mn}} \\ &= \sum_{k=1}^{K}\sum_{i=1}^{I_k} \alpha_{i_k} \mathrm{Tr}\left(\mathbf{W}_{i_k}^* \frac{\partial \mathbf{E}_{i_k}(\mathbf{V}^*, \mathbf{U}^*)}{\partial v_{\ell_j,mn}}\right) \\ &= \sum_{k=1}^{K}\sum_{i=1}^{I_k} \alpha_{i_k} \mathrm{Tr}\left((\mathbf{E}_{i_k}^{\mathrm{mmse}})^{-1} \frac{\partial \mathbf{E}_{i_k}^{\mathrm{mmse}}(\mathbf{V}^*)}{\partial v_{\ell_j,mn}}\right) \\ &= \frac{\partial \psi_2(\mathbf{V}^*)}{\partial v_{\ell_j,mn}} \end{aligned}$$

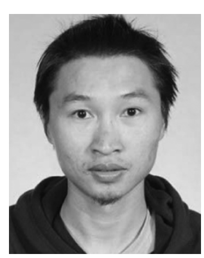
where the first and the last equality follows from the chain rule and the second equality is obtained from (27). Hence, it follows from (26) that

$$\begin{aligned} &\mathrm{Tr}\left(\nabla_{\mathbf{V}}\psi_2(\mathbf{V}^*)^H(\mathbf{V} - \mathbf{V}^*)\right) \\ &\qquad = \mathrm{Tr}\left(\nabla_{\mathbf{V}}\psi_1(\mathbf{W}^*, \mathbf{U}^*, \mathbf{V}^*)^H(\mathbf{V} - \mathbf{V}^*)\right) \leq 0 \end{aligned}$$

which is the stationarity condition of $\mathbf{V}^*$ for (1). The converse can be proved by reversing the steps of the proof. ∎

REFERENCES

[1] Z.-Q. Luo and S. Zhang, "Dynamic spectrum management: complexity and duality," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 57–73, Feb. 2008.
[2] S. Ye and R. S. Blum, "Optimized signaling for MIMO interference systems with feedback," *IEEE Trans. Signal Process.*, vol. 51, no. 11, pp. 2839–2848, Nov. 2003.
[3] E. Larsson and E. Jorswieck, "Competition versus cooperation on the MISO interference channel," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 7, pp. 1059–1069, Sep. 2008.
[4] E. A. Jorswieck and E. G. Larsson, "The MISO interference channel from a game-theoretic perspective: a combination of selfishness and altruism achieves Pareto optimality," presented at the IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP), Las Vegas, NV, Apr. 2008.
[5] C. Shi, R. A. Berry, and M. L. Honig, "Distributed interference pricing with MISO channels," presented at the Allerton Conf. Commun., Control, Comput., Urbana-Champaign, IL, Sep. 2008.
[6] J. Huang, R. A. Berry, and M. L. Honig, "Distributed interference compensation for wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 5, pp. 1074–1084, May 2006.
[7] Z.-Q. Luo and J. S. Pang, "Analysis of iterative waterfilling algorithm for multiuser power control in digital subscriber lines," *EURASIP J. Appl. Signal Process.*, vol. 2006, no. 1, pp. 1–10, Jan. 2006.
[8] G. Scutari, D. P. Palomar, and S. Barbarossa, "The MIMO iterative waterfilling algorithm," *IEEE Trans. Signal Process.*, vol. 57, no. 5, pp. 1917–1935, May 2009.
[9] C. Shi, R. A. Berry, and M. L. Honig, "Monotonic convergence of distributed interference pricing in wireless networks," presented at the IEEE Int. Symp. Inf. Theory (ISIT), Seoul, Korea, Jun. 2009.
[10] C. Shi, R. A. Berry, and M. L. Honig, "Local interference pricing for distributed beamforming in MIMO networks," presented at the IEEE MILCOM, Boston, MA, Oct. 2009.
[11] Z. K. M. Ho and D. Gesbert, "Balancing egoism and altruism on interference channel: The MIMO case," presented at the IEEE ICC, Cape Town, South Africa, May 2010.
[12] S. J. Kim and G. B. Giannakis, "Optimal resource allocation for MIMO ad hoc cognitive radio networks," presented at the Allerton Conf. Commun., Control, Comput., Urbana-Champaign, IL, Sep. 2008.
[13] S. S. Christensen, R. Argawal, E. de Carvalho, and J. M. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 1–7, Dec. 2008.
[14] D. Schmidt, C. Shi, R. A. Berry, M. L. Honig, and W. Utschick, "Minimum mean squared error interference alignment," presented at the Asilomar Conf. Signals, Syst., Comput., Pacific Grove, CA, Nov. 2009.
[15] D. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, MA: Athena Scientific, 1999.
[16] R. Tresh and M. Guillaud, "Cellular interference alignment with imperfect channel knowledge," presented at the IEEE Int. Conf. Commun. (ICC), Dresden, Germany, Jun. 2009.
[17] D. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 502–513, Mar. 2000.
[18] S. Serbetli and A. Yener, "MMSE transmitter design for correlated MIMO systems with imperfect channel estimates: power allocation trade-offs," *IEEE Trans. Wireless Commun.*, vol. 5, no. 8, pp. 2295–2304, Aug. 2006.
[19] M. Ding and S. D. Blostein, "MIMO minimum total MSE transceiver design with imperfect CSI at both ends," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 1141–1150, Mar. 2009.
[20] M. Medard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 933–946, May 2000.
[21] M. Razaviyayn, M. S. Boroujeni, and Z.-Q. Luo, "Linear transceiver design for interference alignment: Complexity and computation," presented at the IEEE Signal Process. Adv. in Wireless Commun. (SPAWC), Marrakech, Morocco, Jun. 2010.
[22] K. Gomadam, V. R. Cadambe, and S. A. Jafar, "Approaching the capacity of wireless networks through distributed interference alignment," presented at the IEEE GLOBECOM, Miami, FL, Dec. 2008.
[23] M. V. Solodov, "On the convergence of constrained parallel variable distribution algorithm," *SIAM J. Optim.*, vol. 8, no. 1, pp. 187–196, Feb. 1998.



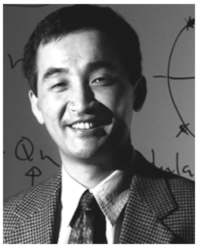
Qingjiang Shi received the B.S. degree in electronic engineering from the China University of Petroleum, Shangdong, China, in 2003 and the Ph.D. degree in communication engineering from Shanghai Jiao Tong University, Shanghai, China, in 2011.

From September 2009 to September 2010, he obtained a national scholarship and visited Prof. Z.-Q. (Tom) Luo's research group at the University of Minnesota, Twin Cities. He is currently a Research Scientist at the Research & Innovation Center (Bell Labs China), Alcatel-Lucent Shanghai Bell Company, Ltd., China. His research interests lie in algorithm design for signal processing and wireless communication.

Dr. Shi received the Best Paper Award from IEEE PIMRC 2009 conference.

Meisam Razaviyayn received the B.Sc. degree from Isfahan University of Technology, Isfahan, Iran, in 2008.

During summer 2010, he was working as a research intern at Huawei Technologies. He is currently working towards the Ph.D. degree in electrical engineering at the University of Minnesota. His research interests include the design and analysis of efficient optimization algorithms with application to data communication and signal processing.



Zhi-Quan (Tom) Luo (F'07) received the B.Sc. degree in applied mathematics from Peking University, China, in 1984 and the Ph.D. degree in operations research from the Massachusetts Institute of Technology, Cambridge, in 1989.

From 1989 to 2003, he was with the Department of Electrical and Computer Engineering, McMaster University, Canada, where he later served as the department head and held a senior Canada Research Chair in Information Processing. He is currently a Professor in the Department of Electrical and Computer Engineering at the University of Minnesota, Twin Cities, where he holds an endowed ADC Chair in digital technology. His research interests lie in the union of optimization algorithms, data communication, and signal processing.

Dr. Luo is a Fellow of SIAM. He was a recipient of the IEEE Signal Processing Society's Best Paper Award in 2004 and 2009, and the EURASIP Best Paper Award in 2011. He was awarded the 2010 Farkas Prize by the INFORMS Optimization Society. He currently chairs the IEEE Signal Processing Society's Technical Committee on Signal Processing for Communications and Networking (SPCOM). He has held editorial positions for several international journals, including the *Journal of Optimization Theory and Applications*, the *Mathematics of Computation*, the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the *SIAM Journal on Optimization*, *Management Sciences*, and the *Mathematics of Operations Research*.

Chen He (S'93–M'96) received the B.E. degree and M.E. degree in electronic engineering from Southeast University of China in 1982 and 1985, respectively, and the Ph.D. degree in electronics system from Tokushima University of Japan in 1994.

He worked in the Department of Electronic Engineering in Southeast University of China from 1985 to 1990. He joined the Department of Electronic Engineering in Shanghai Jiao Tong University of China in 1996. He visited Tokushima University of Japan as a foreign researcher from October 1990 to September 1991 and visited the Communication Research Laboratory of Japan from December 1999 to December 2000 as a research fellow. He is currently a Professor and Vice-Director of Advanced Communication Institute of Shanghai Jiao Tong University in China. His current research interests are 4G wireless communication systems, wireless sensor network, and signal processing. He has published over 200 journal papers and over 80 conference papers.

Dr. He received the Best Paper Award in the IEEE GLOBECOM 2007.