Pricing-based Distributed Downlink Beamforming in Multi-Cell OFDMA Networks

Weiqiang Xu, *Member, IEEE,* Xiaodong Wang, *Fellow, IEEE*

***Abstract*—We address the problem of downlink beamforming for mitigating the co-channel interference in multi-cell OFDMA networks. Based on the network utility maximization framework, we formulate the problem as a non-convex optimization problem subject to the per-cell power constraints, in which a general utility function of SINR is used to characterize the network performance. Some classical utility functions, such as the proportional fairness utility, the weighted sum-rate utility and the α-fairness utility, are subsumed as special cases of our formulation. To solve the problem in a distributed fashion, we devise an algorithm based on the non-cooperative game with pricing mechanism. We give a sufficient condition for the convergence of the algorithm to the Nash equilibrium (NE), and analyze the information exchange overhead among the base stations. Moreover, to speed up the optimization of the beamvectors at each cell, we derive an efficient algorithm to solve for the KKT conditions at each cell. We provide extensive simulation results to demonstrate that the proposed distributed multi-cell beamforming algorithm converges to an NE point in just a few iterations with low information exchange overhead. Moreover, it provides significant performance gains, especially under the strong interference scenario, in comparison with several existing multi-cell interference mitigation schemes, such as the distributed interference alignment method.**

***Index Terms*—Multi-cell, downlink beamforming, distributed algorithm, game theory, pricing mechanism, utility optimization, dual decomposition.**

I. Introduction

In multi-cell wireless networks, besides the intra-cell interference caused by spatial multiplexing within each cell, another impediment arises from inter-cell interference due to the ever-shrinking cell sizes. Alleviating the effects of inter-cell interference requires the base stations (BSs) to adjust their transmission schemes collectively. In fact, inter-cell interference mitigation has been identified as a key issue for future wireless networks. In particular, for downlink transmissions, if the inter-cell interference is mitigated via coordinated processing across multiple BSs, significant performance gains can be possibly obtained, especially for the users at the cell edges. Therefore, recently, there has been a rapidly growing interest in shifting the design paradigm from the conventional single-cell to the cooperative multi-cell networks [1]. Various methods, such as [2], [3], [4], have been proposed to provide network-wide, macroscopic cooperation among different BSs. In these studies, it is assumed that the BSs in a multi-cell network are connected via backhaul links to a central processing unit, which has the global knowledge of the transmitted data from all the users in the network and the downlink channels from each BS to all the users. Such a fully coordinated case is sometimes referred to as networked MIMO. However, for large and dense networks, networked MIMO obviously incurs a substantial infrastructural and computational overhead, which increases the system costs and hinders the practical implementations. This motivates the problem of constrained cooperation, taking into account many practical factors, e.g., limited backhaul capacity [5], local cooperation [6], processing complexity and delay [7], imperfect channel state information (CSI) [8] [9], and feedback errors [10].

On the other hand, future cellular networks are envisioned to be distributed systems with autonomous and self-coordinated cells. Each BS can make independent and rational decisions in a decentralized manner, with limited information exchange with the neighboring BSs. This motivates the study of distributed multi-cell interference mitigation, which requires only the local and neighboring CSI at each BS, without the need of a central controller, and is therefore much easier to implement. Based on a generalization of uplink-downlink duality to the multi-cell setting, an iterative algorithm is proposed in [11] to optimally solve the multicell downlink beamforming problem for minimizing either the total weighted transmit power or the maximum per-antenna power subject to the SINR constraints. An alternative to the transmit power minimization problem is the rate maximization problem subject to the power constraints, which is in general non-convex. An approach based on the concept of virtual SINR is proposed in [12]. In [13] an iterative algorithm is developed for solving the KKT conditions of the weighted sum-rate maximization problem subject to per-cell power constraints. However, the proof of convergence is still an open problem. Other related works include [14], which explores the relationship between the MISO interference channel and the cognitive radio MISO channel to devise rate-optimal strategies for decentralized multi-cell cooperative beamforming.

Game theory provides a systematic mathematical framework for the study of competition and cooperation among intelligent and rational decision makers.There has been a significant amount of recent research that applies game theory to resource allocation problems in wireless networks [15]- [19].In general, game models can be classified into two main categories: non-

Manuscript received October 30, 2010; accepted August 29, 2011. This work was supported in part by the National Science Foundation of China (NSFC) under grants 60702081, 60873020, 61002016, by the Joint Research Fund for Overseas Chinese, Hong Kong and Macao Young Scholars under grant 61028001, by the Key Project of Chinese Ministry of Education under grant 212066, and by the Zhejiang Provincial Science Foundation under grants Z1080702, Y1090980, Y12F020196.

W. Xu is with School of Information Science & Technology, Zhejiang Sci-Tech University, Hangzhou, 310018, P. R. China. (email: wqxu@zstu.edu.cn).

X. Wang is with the Department of Electrical Engineering, Columbia University, New York, NY, 10027, USA. (e-mail: wangx@ee.columbia.edu).

cooperative and cooperative games. Although non-cooperative game is a useful tool to devise totally distributed algorithms, the Nash equilibrium (NE) of the non-cooperative game may suffer a significant performance degradation compared with the optimal centralized solution. On the other hand, the cooperative game approach offers performance gain over the non-cooperative game, but it requires extensive message exchanges among all players, which implies a large communication overhead and poor scalability when applied to large networks. The pricing mechanism is another alternative to overcome the inefficiency of the non-cooperative game approach. In this paper, we develop a provably convergent distributed multi-cell beamforming technique based on the pricing-based non-cooperative game.

In particular, we formulate the problem as a general network utility maximization problem subject to per-cell power constraints, which is a non-convex problem. Examples of the utility functions include the weighted sum-rate utility, the proportional fairness utility, and the α-fairness utility [20], etc. We treat each cell in the multi-cell network as a player, and design an efficient distributed pricing mechanism to optimize the network performance through coordination among the players. We give a sufficient condition for the convergence of the proposed distributed multi-cell beamforming algorithm. Moreover, we derive an efficient algorithm based on the dual decomposition technique for solving the KKT conditions of the downlink beamforming problem at each BS. The proposed technique can converge rapidly to the NE point with a low information exchange overhead among the BSs. It provides significant performance gains, especially in the strong interference scenario, in comparison with several existing approaches, including the recently proposed distributed interference alignment method [21].

The remainder of this paper is organized as follows. In Section II we introduce the system model and the problem formulation. In Section III we develop the pricing-based multicell distributed downlink beamforming technique. In Section IV, we derive the beamforming optimization algorithm at each cell based on dual decomposition. Simulation results are given in Section V. Finally Section VI concludes the paper.

II. System Model and Problem Formation

We consider a downlink multi-cellular network where a set of BSs $\mathcal{M} = \{1, 2, \ldots, M\}$ simultaneously transmit on the orthogonal sub-channels[1] $\mathcal{N} = \{1, 2, \ldots, N\}$ during each scheduling interval. Each BS $m \in \mathcal{M}$ is equipped with T transmit antennas and space-division multiple-access (SDMA) is employed to serve multiple single-antenna mobile users on each sub-channel. Let $\mathcal{B}_m^{(n)}$ be the set of users scheduled by BS $m \in \mathcal{M}$ on sub-channel $n \in \mathcal{N}$. For simplicity and without loss of generality, we assume that $|\mathcal{B}_m^{(n)}| = Q, \forall m, \forall n$. We further assume that each user is served by only one BS.

For data transmission, BS m on sub-channel n transmits complex symbols $b_{m,k}^{(n)} \in \mathbb{C}$ through T transmit antennas using a beam-vector $\mathbf{w}_{m,k}^{(n)} \in \mathbb{C}^T$ to user $k \in \mathcal{B}_m^{(n)}$. We assume that $\mathbb{E}\{|b_{m,k}^{(n)}|^2\} = 1$, and $\mathbb{E}\{b_{m_1,k_1}^{(n_1)} b_{m_2,k_2}^{(n_2)}\} = 0$, for $(n_1, m_1, k_1) \neq (n_2, m_2, k_2)$, where $\mathbb{E}\{\cdot\}$ is the expectation operator. Then after normalized by the noise standard deviation, the received signal by user $k \in \mathcal{B}_m^{(n)}$ on sub-channel n can be written as

$$\begin{aligned} y_{m,k}^{(n)} = \underbrace{\vec{\mathbf{h}}_{m,k}^{(n)} \mathbf{w}_{m,k}^{(n)} b_{m,k}^{(n)}}_{\text{useful signal}} + \underbrace{\sum_{k' \in \mathcal{B}_m^{(n)} \setminus k} \vec{\mathbf{h}}_{m,k}^{(n)} \mathbf{w}_{m,k'}^{(n)} b_{m,k'}^{(n)}}_{\text{in-cell co-channel interference}} \\ + \underbrace{\sum_{j \in \mathcal{M} \setminus m} \sum_{u \in \mathcal{B}_j^{(n)}} \vec{\mathbf{h}}_{j,k}^{(n)} \mathbf{w}_{j,u}^{(n)} b_{j,u}^{(n)}}_{\text{out-cell co-channel interference}} + \underbrace{z_{m,k}^{(n)}}_{\text{noise}}, \end{aligned} \tag{1}$$

where $\mathbf{h}_{m,k}^{(n)} \in \mathbb{C}^T$ is the complex channel vector between BS m and user $k \in \mathcal{B}_m^{(n)}$ on sub-channel n, $z_{m,k}^{(n)} \sim \mathcal{N}_{\mathbb{C}}(0, 1)$ denotes the circularly symmetric complex Gaussian noise sample, and $\vec{\cdot}$ is the Hermitian transpose operator.

The SINR for user $k \in \mathcal{B}_m^{(n)}$ on sub-channel n can then be expressed as[2]

$$\Gamma_{m,k}^{(n)} = \frac{\left|\vec{\mathbf{h}}_{m,k}^{(n)} \mathbf{w}_{m,k}^{(n)}\right|^2}{1 + \mathcal{I}_{m,k}^{(n)}}, \tag{2}$$

where $\mathbf{W}^{(n)} = \{\mathbf{w}_{m,k}^{(n)}, k \in \mathcal{B}_m^{(n)}, m \in \mathcal{M}\}, n \in \mathcal{N}$, and

$$\mathcal{I}_{m,k}^{(n)} = \underbrace{\sum_{k' \in \mathcal{B}_m^{(n)} \setminus k} \left|\vec{\mathbf{h}}_{m,k}^{(n)} \mathbf{w}_{m,k'}^{(n)}\right|^2}_{\mathcal{I}_{m,k}^{(n),\text{in}}} + \underbrace{\sum_{j \in \mathcal{M} \setminus m} \underbrace{\sum_{u \in \mathcal{B}_j^{(n)}} \left|\vec{\mathbf{h}}_{j,k}^{(n)} \mathbf{w}_{j,u}^{(n)}\right|^2}_{\mathcal{I}_{m,k}^{(n),\text{out},j}}}_{\mathcal{I}_{m,k}^{(n),\text{out}}}, \tag{3}$$

where the terms $\mathcal{I}_{m,k}^{(n),\text{in}}$ and $\mathcal{I}_{m,k}^{(n),\text{out}}$ account for the in-cell and out-cell interference, respectively.

Now, we consider the following general linear beamforming optimization problem where we wish to maximize a network-wide utility function across all users of all coordinated BSs and all sub-channels, by choosing the set of beam-vectors $\mathbf{W} = \{\mathbf{W}^{(n)}, n \in \mathcal{N}\}$, subject to the per-base-station power constraints:

$$\begin{aligned} \max_{\mathbf{W}} \quad & U_{\text{network}}(\mathbf{W}) = \sum_{m \in \mathcal{M}} \sum_{n \in \mathcal{N}} \sum_{k \in \mathcal{B}_m^{(n)}} U_{m,k}^{(n)}\Big(\Gamma_{m,k}^{(n)}\Big), \\ \text{s.t.} \quad & \sum_{n \in \mathcal{N}} \sum_{k \in \mathcal{B}_m^{(n)}} \vec{\mathbf{w}}_{m,k}^{(n)} \mathbf{w}_{m,k}^{(n)} \leq P_m, \ \forall m \in \mathcal{M}, \end{aligned} \tag{4}$$

where P_m is the total transmit power at BS m. We assume that the above optimization problem has a set of feasible solutions, which can be facilitated through some form of admission control or/and scheduling strategies.

In the above formulation, each user $k \in \mathcal{B}_m^{(n)}$ is assigned a utility function $U_{m,k}^{(n)}(\Gamma_{m,k}^{(n)})$, which is assumed to be a monotonically nondecreasing, concave and twice differentiable function of the received SINR $\Gamma_{m,k}^{(n)}$. Typical utility functions include the following:

[1] The sub-channel refers to a logical collection of physical sub-carriers, which is regarded as the minimum granularity of the radio resource allocation unit in this paper.

[2] We drop the explicit dependency of $\Gamma_{m,k}^{(n)}$ and $\mathcal{I}_{m,k}^{(n)}$ on $\mathbf{W}^{(\mathbf{n})}$.

- Proportional fairness utility [20]: $U(\Gamma) = \log(\Gamma)$;
- Rate utility: $U(\Gamma) = \log(1+\Gamma)$;
- α-fairness utility [20]: $U(\Gamma) = (1-\alpha)^{-1}(\Gamma)^{1-\alpha},\ \alpha \neq 1$.

Note that the constraint set in (4) is convex. However, due to the SINR expression (2), even though the utility function $U^{(n)}_{m,k}(\Gamma^{(n)}_{m,k})$ is concave in terms of the SINR $\Gamma^{(n)}_{m,k}$, it is in general nonconcave in terms of the set of beam-vectors $\mathbf{W}^{(n)}$. Numerically finding the global optimal solution to the optimization problem (4) is known to be a difficult problem.Our objective is to develop a distributed solution to (4) where each BS updates its beam-vectors locally; and with the aid of limited information exchange among the BSs, some form of optimality can be achieved. To that end, we resort to the game theoretical tool of pricing mechanism.

III. Pricing Mechanism and Distributed Algorithm for Non-Cooperative Beamforming Game

An extreme example of distributed beamforming scheme is for each BS to independently update its own beam-vectors without considering the actions of other BSs. However, such a pure non-cooperative approach may result in non-convergence or some undesirable Nash equilibrium (NE) with low individual as well as system-wise performance [22]. For instance, it is shown in [23] that for a two-user MISO system, the NE point achieved through the pure non-cooperative game over all possible choices of beams is far away from the Pareto boundary of the achievable rate region.

The pricing mechanism [24]–[26] has been employed as an effective means to stimulate cooperation among players, and to guide the players' behaviors toward a more efficient NE that improves the system performance, by introducing a certain degree of coordination in a non-cooperative game. In this section, we propose a pricing mechanism for the non-cooperative multicell beamforming game and the corresponding distributed beamforming algorithm. We then prove the convergence of this algorithm. Finally we analyze the information exchange overhead among the BSs.

A. Pricing Mechanism

We model the pricing-based non-cooperative multicell beamforming game as

$$\mathcal{G} = \{\mathcal{M}, \{\mathcal{W}_m\}_{m\in\mathcal{M}}, \{\bar{U}_m\}_{m\in\mathcal{M}}\},$$

where the elements are

- Player set: $\mathcal{M} = \{1, 2, \ldots, M\}$, i.e., the set of BSs.
- Strategy set: $\{\mathcal{W}_1, \ldots, \mathcal{W}_M\}$, where the strategy set of player (BS) m is the following

$$\mathcal{W}_m = \Big\{\mathbf{w}^{(n)}_{m,k} \in \mathbb{C}^T, k \in \mathcal{B}^{(n)}_m, n \in \mathcal{N} : \sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{B}^{(n)}_m} \vec{\mathbf{w}}^{(n)}_{m,k}\mathbf{w}^{(n)}_{m,k} \leq P_m\Big\}. \tag{5}$$

- Payoff functions set: $\{\bar{U}_1, \ldots, \bar{U}_M\}$, with

$$\bar{U}_m(\mathbf{W}_m, \mathbf{W}_{-m}) = \sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{B}^{(n)}_m} U^{(n)}_{m,k}(\Gamma^{(n)}_{m,k}) - C(\mathbf{W}_m, \mathbf{W}_{-m}), \tag{6}$$

where $\mathbf{W}_m = \{\mathbf{w}^{(n)}_{m,k}, k \in \mathcal{B}^{(n)}_m, n \in \mathcal{N}\}$ and $\mathbf{W}_{-m} = \{\mathbf{W}_1, ..., \mathbf{W}_{m-1}, \mathbf{W}_{m+1}, ...\mathbf{W}_M\}$ denote the set of beam-vectors of BS m, and that of all other BSs, respectively. $C(\mathbf{W}_m, \mathbf{W}_{-m})$ is a cost function associated with a pricing mechanism.

An efficient pricing mechanism should take into account the nature of the service requirement of each player and reflect accurately the cost of resource consumption for fullfilling each player's requirement. Inspired by [24]–[26], we will apply the usage-based pricing mechanism to solve our problem, where the price a player pays for using the resource is proportional to the amount of resource consumed by the player.

First, we introduce a quantity called the interference pricing rate of user $k \in \mathcal{B}^{(n)}_m$, which measures the marginal decrease in utility due to a marginal increase in interference, given by

$$\pi^{(n)}_{m,k} \triangleq -\frac{\partial U^{(n)}_{m,k}}{\partial \mathcal{I}^{(n)}_{m,k}} = (U^{(n)}_{m,k})' \frac{|\vec{\mathbf{h}}^{(n)}_{m,k}\mathbf{w}^{(n)}_{m,k}|^2}{(1+\mathcal{I}^{(n)}_{m,k})^2}, \tag{7}$$

where $(U^{(n)}_{m,k})'$ denote the derivative of the utility function with respect to the SINR $\Gamma^{(n)}_{m,k}$. When BS m transmits signal to user $k \in \mathcal{B}^{(n)}_m$ on sub-channel n using the beam-vector $\mathbf{w}^{(n)}_{m,k}$, it induces the interference $\left|\vec{\mathbf{h}}^{(n)}_{m,u}\mathbf{w}^{(n)}_{m,k}\right|^2$ to all other users $u \in \mathcal{B}^{(n)}_j, (j,u) \neq (m,k), j \in \mathcal{M}$. Thus, under the pricing mechanism, when serving user $k \in \mathcal{B}^{(n)}_m$, BS m needs to pay a total cost:

$$\sum_{j\in\mathcal{M}}\sum_{u\in\mathcal{B}^{(n)}_j} \pi^{(n)}_{j,u}\left|\vec{\mathbf{h}}^{(n)}_{m,u}\mathbf{w}^{(n)}_{m,k}\right|^2 = \vec{\mathbf{w}}^{(n)}_{m,k}\mathbf{L}^{(n)}_{m,k}\mathbf{w}^{(n)}_{m,k}, \tag{8}$$

where $\mathbf{L}^{(n)}_{m,k}$ is defined in (9). We called $\mathbf{L}^{(n)}_{m,k}$ as the leakage matrix of user $k \in \mathcal{B}^{(n)}_m$ on sub-channel n, which accounts for the amount of interference caused by BS m to other co-channel users on sub-channel n when serving user $k \in \mathcal{B}^{(n)}_m$. Note that $\mathbf{L}^{(n)}_{m,k}$ is Hermitian symmetric, i.e., $\mathbf{L}^{(n)}_{m,k} = \vec{\mathbf{L}}^{(n)}_{m,k}$, since $\mathbf{h}^{(n)}_{m,k}\vec{\mathbf{h}}^{(n)}_{m,k}$ is Hermitian symmetric. The terms $\mathbf{L}^{(n),\mathrm{in}}_{m,k}$ and $\mathbf{L}^{(n),\mathrm{out}}_{m}$ in (9) account for the in-cell and out-cell leakages, respectively.

Hence summing across all users served by BS m and across all sub-channels, BS m needs to pay a total cost of

$$C(\mathbf{W}_m, \mathbf{W}_{-m}) = \sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{B}^{(n)}_m} \vec{\mathbf{w}}^{(n)}_{m,k}\mathbf{L}^{(n)}_{m,k}\mathbf{w}^{(n)}_{m,k}. \tag{10}$$

Summarizing the discussion above, in the pricing-based non-cooperative multicell beamforming game, each BS m solves the following optimization problem

$$\begin{aligned} \max_{\mathbf{W}_m}\ & \sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{B}^{(n)}_m} \Big(U^{(n)}_{m,k}\Big(\Gamma^{(n)}_{m,k}\Big) - \vec{\mathbf{w}}^{(n)}_{m,k}\mathbf{L}^{(n)}_{m,k}\mathbf{w}^{(n)}_{m,k}\Big), \\ \text{s.t.}\ & \sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{B}^{(n)}_m} \vec{\mathbf{w}}^{(n)}_{m,k}\mathbf{w}^{(n)}_{m,k} \leq P_m. \end{aligned} \tag{11}$$

Notice that the objective function is still nonconcave with respect to the beam-vectors $\mathbf{W}_m$ associated with BS m; thus the globally optimal solution to (11) cannot be found. In Section IV we drive a dual decomposition algorithm for obtaining the solution to the KKT conditions of (11).

$$\mathbf{L}_{m,k}^{(n)} \triangleq \underbrace{\sum_{k'\in\mathcal{B}_m^{(n)}\setminus k} \pi_{m,k'}^{(n)} \mathbf{h}_{m,k'}^{(n)} \vec{\mathbf{h}}_{m,k'}^{(n)}}_{\mathbf{L}_{m,k}^{(n),\mathrm{in}}} + \underbrace{\sum_{j\in\mathcal{M}\setminus m} \underbrace{\sum_{u\in\mathcal{B}_j^{(n)}} \pi_{j,u}^{(n)} \mathbf{h}_{m,u}^{(n)} \vec{\mathbf{h}}_{m,u}^{(n)}}_{\mathbf{L}_m^{(n),\mathrm{out},j}}}_{\mathbf{L}_m^{(n),\mathrm{out}}}. \tag{9}$$

B. Distributed Multicell Beamforming Algorithm

We propose the following distributed algorithm for implementing the pricing-based non-cooperative multicell beamforming game.

Algorithm 1: Distributed multicell beamforming algorithm.

```
Initialization:
  Each BS m initializes W_m satisfying the power constraint.
Repeat
  For  m = 1 : M
      BS m obtains a solution W̄_m to (11) for given
      W_{-m}, using Algorithm 2.
      If Ū_m(W̄_m, W_{-m}) ≥ Ū_m(W_m, W_{-m})
      Then
        {
          BS m updates its beam-vectors as W̄_m.
          Based on the new beam-vectors W̄_m, BS m
          updates
          {I_{j,u}^{(n),out,m}, u ∈ B_j^{(n)}, j ∈ M\m, n ∈ N},
          and {π_{m,k}^{(n)}, k ∈ B_m^{(n)}, n ∈ N}
          according to (3) and (7) respectively, and
          passes them to BSs j ∈ M\m.
        }
      EndIf
  EndFor
Until convergence
```

We have the following observations on Algorithm 1.

1) Only one BS updates its beam-vectors at a time, based on the latest out-cell interference powers and interference price rates (and thus the latest out-cell leakage matrices) from every other BS in the multicell network. Moreover, after a BS updates its beam-vectors, the new out-cell interference powers and new interference price rates are announced timely to every other BS.
2) Only if $\bar{U}_m(\bar{\mathbf{W}}_m, \mathbf{W}_{-m}) \geq \bar{U}_m(\mathbf{W}_m, \mathbf{W}_{-m})$ holds, BS m updates its beam-vectors as $\bar{\mathbf{W}}_m$. Otherwise, BS m keeps its old beam-vectors. This method is based on the better response strategy in game theory, which refers to an update procedure where the players choose actions that increase their utilities as opposed to maximizing their utilities in the best response strategy. Notice that the best response strategy cannot applied due to the nonconvexity of (11).

These features ensure the convergence of the algorithm, as discussed next.

C. Existence and Convergence of NE

The Nash equilibrium (NE) is a well-known concept for analyzing a game. A set of beam-vectors $\mathbf{W}^* = (\mathbf{W}_1^*, \ldots, \mathbf{W}_M^*)$ is an NE if, for every BS $m \in \mathcal{M}$, $\bar{U}_m(\mathbf{W}_m^*, \mathbf{W}_{-m}^*) \geq \bar{U}_m(\mathbf{W}_m, \mathbf{W}_{-m}^*)$, $\forall \mathbf{W}_m \in \mathcal{W}_m$. That is, given the other BSs' beam-vectors, no BS can increase its utility unilaterally by changing its own beam-vectors. For the multicell beamforming game under consideration, the existence and convergence of NE is heavily dependent on the concavity of the utility function $U_{m,k}^{(n)}$. We first introduce a quantity that measures the relative concavity of a utility function. Specifically, the coefficient of relative risk aversion associated with the utility function $U(\Gamma)$ is defined as

$$\kappa(\Gamma) = -\frac{\Gamma \cdot U(\Gamma)''}{U(\Gamma)'}, \tag{12}$$

where $U(\Gamma)'$ and $U(\Gamma)''$ denote the first- and second-order derivatives, respectively.

We have the following result on a sufficient condition for the convergence of Algorithm 1.

Proposition 1: Suppose that the utility function $U_{m,k}^{(n)}$ satisfies

$$0 \leq \kappa_{m,k}^{(n)} \leq 2, \quad \forall k \in \mathcal{B}_m^{(n)},\ m \in \mathcal{M},\ n \in \mathcal{N},$$

then Algorithm 1 converges to an NE point.

Proof: See Appendix A. □

Remark: The condition $0 \leq \kappa_{m,k}^{(n)} \leq 2$ can be interpreted as requiring that the utility function to be sufficiently concave, but not too concave. If the utility function is too concave (i.e., $\kappa_{m,k}^{(n)} > 2$), the updates may be too aggressive to guarantee convergence. Fortunately, this condition is satisfied by most utility functions of interest, as discussed below.

1) For the proportional fairness utility function $U(\Gamma) = \log_2(\Gamma)$, its coefficient of relative risk aversion is $\kappa = 1$.
2) For the α-fairness utility function $U(\Gamma) = \frac{(\Gamma)^{1-\alpha}}{1-\alpha}$ with $\alpha \neq 1$, we have $\kappa_{m,k}^{(n)} = \alpha$. Hence, for $0 \leq \alpha \leq 2$ ($\alpha \neq 1$), we have $0 \leq \kappa \leq 2$ ($\kappa \neq 1$).
3) For the weighted sum-rate utility $U(\Gamma) = \omega \log_2(1 + \theta\Gamma)$, with $0 < \theta \leq 1$, we have the following.
   - $\theta = 1$ corresponds to the Shannon rate with $0 < \kappa = \frac{\Gamma}{1+\Gamma} < 1$.
   - $0 < \theta < 1$ corresponds to the achievable rate for some practical modulations, where $\theta = -\frac{\phi_1}{\log(\phi_2 \mathrm{BER})}$, and ϕ_1, ϕ_2 are constants depending on the modulation and BER is the required bit-error rate. We have $0 < \kappa = \frac{\theta\Gamma}{1+\theta\Gamma} < 1$.

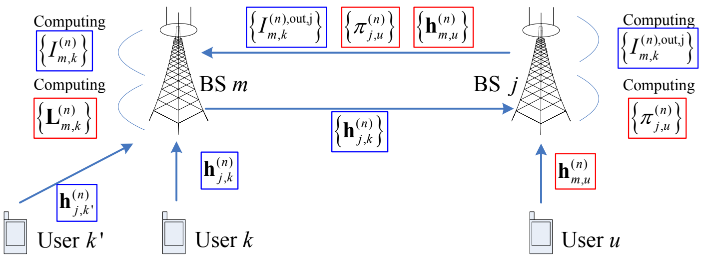


Fig. 1. Local computation at each BS and the information exchange among the BSs.

D. Information Exchange Overhead

In order to compute the SINR $\Gamma_{m,k}^{(n)}$ and the leakage matrix $\mathbf{L}_{m,k}^{(n)}$ in (11), BS m needs to get certain information from the neighboring BSs. In Fig. 1, we give a graphic illustration of the local computation at each BS and the information exchange among the BSs, which is further elaborated as follows.

BS m can compute $\mathcal{I}_{m,k}^{(n)}$ in (2)-(3) for each user $k \in \mathcal{B}_m^{(n)}$ on each sub-channel $n \in \mathcal{N}$ through the following.

1) Computing the first term $\mathcal{I}_{m,k}^{(n),\text{in}} = \sum_{k'\in\mathcal{B}_m^{(n)}\setminus k} \left|\vec{\mathbf{h}}_{m,k}^{(n)}\mathbf{w}_{m,k'}^{(n)}\right|^2$ in (3). BS m only needs the knowledge of the direct downlink channels $\mathbf{h}_{m,k}^{(n)}$ and of the local beam-vectors $\{\mathbf{w}_{m,k'}^{(n)}, k' \in \mathcal{B}_m^{(n)}\setminus k\}$ computed at the previous iteration. Thus, no information exchange is needed.
2) Computing the second term $\mathcal{I}_{m,k}^{(n),\text{out},j} = \sum_{u\in\mathcal{B}_j^{(n)}} \left|\vec{\mathbf{h}}_{j,k}^{(n)}\mathbf{w}_{j,u}^{(n)}\right|^2, j \in \mathcal{M}\setminus m$ in (3). BS $j \in \mathcal{M}\setminus m$ needs the local beam-vectors $\{\mathbf{w}_{j,u}^{(n)}, u \in \mathcal{B}_j^{(n)}\}$ computed at the previous iteration, and the interference downlink channel $\mathbf{h}_{j,k}^{(n)}$, which is sent from user $k \in \mathcal{B}_m^{(n)}$ to the serving BS m, and then from BS m to BS j. BS j calculates $\mathcal{I}_{m,k}^{(n),\text{out},j}$ and then sends it to BS m.

BS m can compute $\mathbf{L}_{m,k}^{(n)}$ in (9) for each user $k \in \mathcal{B}_m^{(n)}$ and each sub-channel $n \in \mathcal{N}$ through the following.

1) Computing the first term $\mathbf{L}_{m,k}^{(n),\text{in}} = \sum_{k'\in\mathcal{B}_m^{(n)}\setminus k} \pi_{m,k'}^{(n)}\mathbf{h}_{m,k'}^{(n)}\vec{\mathbf{h}}_{m,k'}^{(n)}$ in (9). BS m only needs the direct downlink channels $\{\mathbf{h}_{m,k'}^{(n)}, k' \in \mathcal{B}_m^{(n)}\setminus k\}$ and the local interference pricing rates $\{\pi_{m,k'}^{(n)}, k' \in \mathcal{B}_m^{(n)}\setminus k\}$ computed at the previous iteration. Thus no information exchange is needed.
2) Computing the second term $\mathbf{L}_m^{(n),\text{out},j} = \sum_{u\in\mathcal{B}_j^{(n)}} \pi_{j,u}^{(n)}\mathbf{h}_{m,u}^{(n)}\vec{\mathbf{h}}_{m,u}^{(n)}, j \in \mathcal{M}\setminus m$ in (9). BS j needs the interference downlink channels $\{\mathbf{h}_{m,u}^{(n)}, u \in \mathcal{B}_j^{(n)}\}$, each of which is sent from user u to the serving BS j. Notice that these channels have already been sent when computing $\mathcal{I}_{j,u}^{(n),\text{out},m}$. Thus this incurs no additional information exchange. On the other hand, the interference pricing rates $\{\pi_{j,u}^{(n)}, u \in \mathcal{B}_j^{(n)}\}$ can be computed locally according to (7) for which the quantity $\mathcal{I}_{j,u}^{(n)}$ is needed, which in turn has been computed at the previous iteration. After computing $\{\pi_{j,u}^{(n)}, u \in \mathcal{B}_j^{(n)}\}$, BS j sends them to BS m.

Remark: Note that compared with [13], in our scheme, BS j sends $\{\pi_{j,u}^{(n)}, u \in \mathcal{B}_j^{(n)}\}$ instead of $\mathbf{L}_m^{(n),\text{out},j}$ itself to BS m. Since $\mathbf{L}_m^{(n),\text{out},j}$ is a $T \times T$ complex-valued matrix, while $\{\pi_{j,u}^{(n)}, u \in \mathcal{B}_j^{(n)}\}$ is a $Q \times 1$ real-valued vector, and typically $Q \leq T$, our scheme incurs a much lower information exchange overhead.

IV. Per-BS Beam-vector Update via Dual Decomposition

In this section, we derive a dual decomposition algorithm for obtaining the solution to the KKT conditions of Problem (11) at each BS, which is the key step in Algorithm 1 (line 5 and 6).

A. Dual Decomposition

The dual decomposition technique is an effective method for decoupling the coupled constraints and performing distributed optimization.

First, by introducing a set of scalar variables $\mathbf{p}_m = \left\{p_{m,k}^{(n)}, k \in \mathcal{B}_m^{(n)}, n \in \mathcal{N}\right\}$, we can rewrite the optimization problem (11) as follows:

$$\begin{aligned} \max_{\mathbf{W}_m} \quad & \sum_{n\in\mathcal{N}} \sum_{k\in\mathcal{B}_m^{(n)}} \left(U_{m,k}^{(n)}(\Gamma_{m,k}^{(n)}) - \vec{\mathbf{w}}_{m,k}^{(n)}\mathbf{L}_{m,k}^{(n)}\mathbf{w}_{m,k}^{(n)}\right), \\ \text{s.t.} \quad & \sum_{n\in\mathcal{N}} \sum_{k\in\mathcal{B}_m^{(n)}} p_{m,k}^{(n)} \leq P_m, \\ & \vec{\mathbf{w}}_{m,k}^{(n)}\mathbf{w}_{m,k}^{(n)} \leq p_{m,k}^{(n)}, \quad k \in \mathcal{B}_m^{(n)},\ n \in \mathcal{N}. \end{aligned} \tag{13}$$

Notice that the optimization problem (13) has only one single coupled constraint $\sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{B}_m^{(n)}} p_{m,k}^{(n)} \leq P_m$. Then, we form the Lagrangian (14) of the optimization problem (13) with respect to the coupled constraint. In (14), λ_m denotes the Lagrangian dual variable.

Define the dual problem as

$$\min_{\lambda_m} D_m(\lambda_m), \tag{15}$$

where the objective function $D_m(\lambda_m)$ is

$$\begin{aligned} \max_{\mathbf{W}_m, \mathbf{p}_m} \quad & \tilde{L}_m(\mathbf{W}_m, \mathbf{p}_m, \lambda_m) \\ \text{s.t.} \quad & \vec{\mathbf{w}}_{m,k}^{(n)}\mathbf{w}_{m,k}^{(n)} \leq p_{m,k}^{(n)}, \quad k \in \mathcal{B}_m^{(n)},\ n \in \mathcal{N}. \end{aligned} \tag{16}$$

Notice that the dual function $D_m(\lambda_m)$ is the pointwise maximum of a family of affine functions of λ_m, hence it is a convex function of λ_m.

B. Decoupled Subproblems

First, we need to compute $D_m(\lambda_m)$ for a fixed λ_m. Due to its separable structure, the dual function D_m can be decomposed into NQ subproblems $D_{m,k}^{(n)}, n \in \mathcal{N}, k \in \mathcal{B}_m^{(n)}$ as follows:

$$\begin{aligned} \max_{\left\{\mathbf{w}_{m,k}^{(n)}, p_{m,k}^{(n)}\right\}} \quad & U_{m,k}^{(n)}(\Gamma_{m,k}^{(n)}) - \vec{\mathbf{w}}_{m,k}^{(n)}\mathbf{L}_{m,k}^{(n)}\mathbf{w}_{m,k}^{(n)} - \lambda_m p_{m,k}^{(n)}, \\ \text{s.t.} \quad & \vec{\mathbf{w}}_{m,k}^{(n)}\mathbf{w}_{m,k}^{(n)} \leq p_{m,k}^{(n)}. \end{aligned} \tag{17}$$

$$\begin{aligned}\tilde{L}_m(\mathbf{W}_m,\mathbf{p}_m,\lambda_m) &= \sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{B}_m^{(n)}}\left(U_{m,k}^{(n)}(\Gamma_{m,k}^{(n)})-\vec{\mathbf{w}}_{m,k}^{(n)}\mathbf{L}_{m,k}^{(n)}\mathbf{w}_{m,k}^{(n)}\right)-\lambda_m\Bigl(\sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{B}_m^{(n)}}p_{m,k}^{(n)}-P_m\Bigr)\\ &= \sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{B}_m^{(n)}}\left(U_{m,k}^{(n)}(\Gamma_{m,k}^{(n)})-\vec{\mathbf{w}}_{m,k}^{(n)}\mathbf{L}_{m,k}^{(n)}\mathbf{w}_{m,k}^{(n)}-\lambda_m p_{m,k}^{(n)}\right)+\lambda_m P_m,\end{aligned} \tag{14}$$

The Lagrangian of the subproblem $D_{m,k}^{(n)}$ is given by

$$\begin{aligned}&\widehat{L}_{m,k}^{(n)}\left(\mathbf{w}_{m,k}^{(n)},p_{m,k}^{(n)},\lambda_m,\nu_{m,k}^{(n)}\right)\\ &=U_{m,k}^{(n)}\left(\Gamma_{m,k}^{(n)}\right)-\vec{\mathbf{w}}_{m,k}^{(n)}\mathbf{L}_{m,k}^{(n)}\mathbf{w}_{m,k}^{(n)}-\lambda_m p_{m,k}^{(n)}\\ &\quad-\nu_{m,k}^{(n)}\left(\vec{\mathbf{w}}_{m,k}^{(n)}\mathbf{w}_{m,k}^{(n)}-p_{m,k}^{(n)}\right),\end{aligned} \tag{18}$$

where $\nu_{m,k}^{(n)}$ is a dual variable associated with the constraint $\vec{\mathbf{w}}_{m,k}^{(n)}\mathbf{w}_{m,k}^{(n)}\le p_{m,k}^{(n)}$.

We next obtain the KKT conditions, given by

$$\begin{aligned}\frac{\partial\widehat{L}_{m,k}^{(n)}}{\partial\mathbf{w}_{m,k}^{(n)}} &= U_{m,k}^{(n)}\left(\Gamma_{m,k}^{(n)}\right)'\cdot\frac{2\vec{\mathbf{h}}_{m,k}^{(n)}\mathbf{h}_{m,k}^{(n)}\mathbf{w}_{m,k}^{(n)}}{1+\mathcal{I}_{m,k}^{(n)}}\\ &\quad-2\mathbf{L}_{m,k}^{(n)}\mathbf{w}_{m,k}^{(n)}-2\nu_{m,k}^{(n)}\mathbf{I}_T\mathbf{w}_{m,k}^{(n)}=0\end{aligned} \tag{19}$$

$$\frac{\partial\widehat{L}_{m,k}^{(n)}}{\partial p_{m,k}^{(n)}} = -\lambda_m+\nu_{m,k}^{(n)}=0 \tag{20}$$

where $\mathbf{I}_T$ indicates the $T\times T$ identity matrix.

By combining the above two equations, we obtain

$$U_{m,k}^{(n)}\Bigl(\frac{\bigl|\vec{\mathbf{h}}_{m,k}^{(n)}\mathbf{w}_{m,k}^{(n)}\bigr|^2}{1+\mathcal{I}_{m,k}^{(n)}}\Bigr)'\frac{\mathbf{h}_{m,k}^{(n)}\vec{\mathbf{h}}_{m,k}^{(n)}\mathbf{w}_{m,k}^{(n)}}{1+\mathcal{I}_{m,k}^{(n)}}=\mathbf{T}_{m,k}^{(n)}\mathbf{w}_{m,k}^{(n)}, \tag{21}$$

$$\text{with}\quad \mathbf{T}_{m,k}^{(n)}=\mathbf{L}_{m,k}^{(n)}+\lambda_m\mathbf{I}_T. \tag{22}$$

Solving $\mathbf{w}_{m,k}^{(n)}$ from (21), we obtain the expression for the beam-vectors associated with user $k\in\mathcal{B}_m^{(n)}$ for a fixed λ_m, as follows.

Proposition 2: For a fixed $\lambda_m\ge 0$, the solution to the KKT conditions of problem (17) is of the following form[3]:

$$\mathbf{w}_{m,k}^{(n)*}=\mathbf{T}_{m,k}^{(n)\dagger}\mathbf{h}_{m,k}^{(n)}\sqrt{\left(1+\mathcal{I}_{m,k}^{(n)}\right)\Phi_{m,k}^{(n)}\Upsilon_{m,k}^{(n)}}, \tag{23}$$

$$p_{m,k}^{(n)*}=\vec{\mathbf{w}}_{m,k}^{(n)*}\mathbf{w}_{m,k}^{(n)*}=\left(1+\mathcal{I}_{m,k}^{(n)}\right)\Phi_{m,k}^{(n)}\Psi_{m,k}^{(n)}, \tag{24}$$

$$\text{with}\quad \Phi_{m,k}^{(n)}=\mathrm{Inv}\Bigl\{U_{m,k}^{(n)}\Bigl(\frac{1+\mathcal{I}_{m,k}^{(n)}}{\vec{\mathbf{h}}_{m,k}^{(n)}\mathbf{T}_{m,k}^{(n)\dagger}\mathbf{h}_{m,k}^{(n)}}\Bigr)'\Bigr\}, \tag{25}$$

$$\Upsilon_{m,k}^{(n)}=1/\Bigl(\vec{\mathbf{h}}_{m,k}^{(n)}\mathbf{T}_{m,k}^{(n)\dagger}\mathbf{h}_{m,k}^{(n)}\Bigr)^2, \tag{26}$$

$$\Psi_{m,k}^{(n)}=\Bigl\|\mathbf{T}_{m,k}^{(n)\dagger}\mathbf{h}_{m,k}^{(n)}\Bigr\|^2\Upsilon_{m,k}^{(n)}. \tag{27}$$

Furthermore, if $\mathbf{w}_{m,k}^{(n)*}\neq\mathbf{0}$, $\lambda_m=0$ is feasible only if $\mathbf{h}_{m,k}^{(n)}\in\varrho(\mathbf{L}_{m,k}^{(n)})$, where $\varrho(\mathbf{L}_{m,k}^{(n)})$ denotes the column span of the matrix $\mathbf{L}_{m,k}^{(n)}$.

Proof: See Appendix B. □

[3] $(\cdot)^\dagger$ denotes the pseudo-inverse; $\mathrm{Inv}\{U'\}$ is the inverse function of U'.

C. Master Problem

We need to solve the master problem (15) on top of the NQ subproblems. Since the master problem is convex in λ_m, we will apply the subgradient method. Notice that where $\mathbf{W}_m^*=\{\mathbf{w}_{m,k}^{(n)*},n\in\mathcal{N},k\in\mathcal{B}_m^{(n)}\}$ is the optimizer for (15) in the definition of $D_m\left(\lambda_m\right)$.

Thus, we can set the subgradient of $D_m\left(\lambda_m\right)$ as $P_m-\sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{B}_m^{(n)}}p_{m,k}^{(n)*}$. The subgradient search suggests that we should increase λ_m if $P_m<\sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{B}_m^{(n)}}p_{m,k}^{(n)*}$; and decrease λ_m otherwise. Notice that the adjustment occurs in a one-dimensional space, thus a simple bisection method can be employed.

D. The Beamforming Algorithm at Each BS

Finally we summarize in Algorithm 2 the dual-decomposition-based beamforming algorithm at each BS.

Algorithm 2: An algorithm for updating the beam-vectors at each BS m.

Initialize $\lambda_m^{\min}$ and $\lambda_m^{\max}$;
Repeat
  Set $\lambda_m\leftarrow\left(\lambda_m^{\min}+\lambda_m^{\max}\right)/2$;
  Repeat through $k=1,\dots,Q;1,\dots,Q;\dots$
    Compute $\{\mathcal{I}_{m,k}^{(n),\mathrm{in}}\}_{n=1}^N$ and thus $\{\mathcal{I}_{m,k}^{(n)}\}_{n=1}^N$
    according to (3);
    Compute $\{\mathbf{L}_{m,k}^{(n),\mathrm{in}}\}_{n=1}^N$ and thus $\{\mathbf{L}_{m,k}^{(n)}\}_{n=1}^N$
    according to (9);
    Compute $\{\mathbf{w}_{m,k}^{(n),*},p_{m,k}^{(n),*}\}_{n=1}^N$ for (17) according to
    (23) and (24).
  Until convergence
  If $\sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{B}_m^{(n)}}p_{m,k}^{(n)*}>P_m$
    Then set $\lambda_m^{\min}\leftarrow\lambda_m$
    Else set $\lambda_m^{\max}\leftarrow\lambda_m$
  End If
Until $|\lambda_m^{\max}-\lambda_m^{\min}|\to 0$

Note that Algorithm 2 can also be viewed as an iterative procedure for solving the KKT conditions of problem (11), which consists of the following.

1) The stationarity condition (21) for $k\in\mathcal{B}_m^{(n)}$, $n\in\mathcal{N}$.
2) The sum-power constraint

$$\sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{B}_m^{(n)}}\vec{\mathbf{w}}_{m,k}^{(n)}\mathbf{w}_{m,k}^{(n)}\le P_m. \tag{29}$$

$$
\begin{aligned}
D_m\left(\tilde{\lambda}_m\right) &= \max_{\mathbf{W}_m,\mathbf{p}_m} \sum_{n\in\mathcal{N}} \sum_{k\in\mathcal{B}_m^{(n)}} \left\{U_{m,k}^{(n)}(\Gamma_{m,k}) - \vec{\mathbf{w}}_{m,k}^{(n)}\mathbf{L}_{m,k}^{(n)}\mathbf{w}_{m,k}^{(n)}\right\} - \tilde{\lambda}_m\Big(\sum_{n\in\mathcal{N}} \sum_{k\in\mathcal{B}_m^{(n)}} p_{m,k}^{(n)} - P_m\Big) \\
&\geq \sum_{n\in\mathcal{N}} \sum_{k\in\mathcal{B}_m^{(n)}} \left\{U_{m,k}^{(n)} - \vec{\mathbf{w}}_{m,k}^{(n)*}\mathbf{L}_{m,k}^{(n)}\mathbf{w}_{m,k}^{(n)*}\right\} - \tilde{\lambda}_m\Big(\sum_{n\in\mathcal{N}} \sum_{k\in\mathcal{B}_m^{(n)}} p_{m,k}^{(n)*} - P_m\Big) \\
&= D_m\left(\lambda_m\right) - \Big(\sum_{n\in\mathcal{N}} \sum_{k\in\mathcal{B}_m^{(n)}} p_{m,k}^{(n)*} - P_m\Big)\left(\tilde{\lambda}_m - \lambda_m\right), \qquad (28)
\end{aligned}
$$

3) The complementary slackness conditions:

$$\lambda_m(\sum_{n\in\mathcal{N}} \sum_{k\in\mathcal{B}_m^{(n)}} \vec{\mathbf{w}}_{m,k}^{(n)}\mathbf{w}_{m,k}^{(n)} - P_m) = 0, \qquad (30)$$

with $\lambda_m \geq 0$.

Starts with a given λ_m, Algorithm 2 solves (21) for $k \in \mathcal{B}_m^{(n)}$, $n \in \mathcal{N}$, and then adjusts λ_m according to the search direction suggested by the power constraint (29), such as to satisfy the complementary slackness conditions (30).

Notice that since $\mathcal{N} = \{1, 2, ..., N\}$ is a set of orthogonal sub-channels, for $n_1, n_2 \in \mathcal{N}, n_1 \neq n_2$, and $k_1, k_2 \in \mathcal{B}_m^{(n)}, k_1 \neq k_2$, we can update $\{\mathbf{w}_{m,k_1}^{(n_1)}, p_{m,k_1}^{(n_1)}\}$ for the subproblem $D_{m,k_1}^{(n_1)}$, and $\{\mathbf{w}_{m,k_2}^{(n_2)}, p_{m,k_2}^{(n_2)}\}$ for the subproblem $D_{m,k_2}^{(n_2)}$ in parallel. Such a simultaneous update can improve the convergence speed of the iterative procedure, especially when N is large.

Moreover, it is easy to derive the following KKT conditions of the original problem (4):

$$
\begin{aligned}
& U_{m,k}^{(n)}\Big(\frac{\left|\vec{\mathbf{h}}_{m,k}^{(n)}\mathbf{w}_{m,k}^{(n)}\right|^2}{1+\mathcal{I}_{m,k}^{(n)}}\Big)' \frac{\mathbf{h}_{m,k}^{(n)}\vec{\mathbf{h}}_{m,k}^{(n)}\mathbf{w}_{m,k}^{(n)}}{1+\mathcal{I}_{m,k}^{(n)}} = \mathbf{T}_{m,k}^{(n)}\mathbf{w}_{m,k}^{(n)}, \\
& k \in \mathcal{B}_m^{(n)}, n \in \mathcal{N}, m \in \mathcal{M}, \qquad (31) \\
& \textstyle\sum_{n\in\mathcal{N}} \sum_{k\in\mathcal{B}_m^{(n)}} \vec{\mathbf{w}}_{m,k}^{(n)}\mathbf{w}_{m,k}^{(n)} \leq P_m, \quad m \in \mathcal{M}, \qquad (32) \\
& \lambda_m\Big(\textstyle\sum_{n\in\mathcal{N}} \sum_{k\in\mathcal{B}_m^{(n)}} \vec{\mathbf{w}}_{m,k}^{(n)}\mathbf{w}_{m,k}^{(n)} - P_m\Big) = 0, m \in \mathcal{M}. \quad (33)
\end{aligned}
$$

Let $\mathbf{W}_m^{KKT}$ the beam-vector set satisfying the KKT conditions of (11) for BS $m \in \mathcal{M}$. By comparing (21), (29), (30) with (31), (32), (33), respectively, it is obvious that $\mathbf{W}^{KKT} = \{\mathbf{W}_1^{KKT}, \ldots, \mathbf{W}_M^{KKT}\}$ is the beam-vector set satisfying the KKT conditions of the original problem (4).

In Algorithm 2, each BS $m \in \mathcal{M}$ can achieve the KKT solution to (11). From Proposition 1, we know that Algorithm 1 converges to an NE point. Furthermore, line 6 in Algorithm 1 guarantees that the beam-vector update at each BS cannot decrease the total utility. Thus, the total utility at the NE point is *not* smaller than that at the point satisfying the KKT conditions of the original problem (4).

V. Simulation Results

A. Simulation Setup

We consider a network with hexagonal cells $\mathcal{M}_t = \{1, ..., 27\}$ shown in Fig. 2. The distance between adjacent BSs is D_{BS} =2000m. Let $\mathcal{M}_{co} = \{1, ..., M\}$ be the set of coordinated BSs, and $\mathcal{M}_{un} = \mathcal{M}_t \backslash \mathcal{M}_{co}$ be the set of uncoordinated BSs. On each sub-channel, Q users are uniformly displaced around the serving BS within a circular annulus of external and internal radii of D and $0.9D$, respectively. Since the proposed method is expected to benefit most the cell-edge users, by setting D =1000m as the default value all users are around the cell edges.

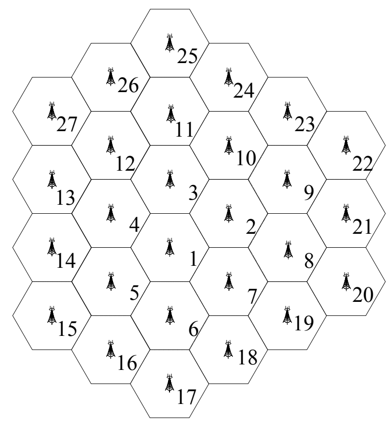


Fig. 2. The simulated network.

The base-band fading channel from the m-th BS to the k-th user on sub-channel n is modeled as

$$\mathbf{h}_{m,k}^{(n)} = \left(\frac{200}{d_{m,k}^{(n)}}\right)^{3.5} l_{m,k}^{(n)}\bar{\mathbf{h}}_{m,k}^{(n)}, \qquad (34)$$

where $d_{m,k}^{(n)}$ is the distance from the m-th BS to the k-th user on sub-channel n; $10\log_{10} l_{m,k}^{(n)}$ is a real Gaussian random variable with zero mean and a standard deviation of 8 accounting for the large scale log-normal shadowing; finally, $\bar{\mathbf{h}}_{m,k}^{(n)} \sim \mathcal{N}_c(\mathbf{0}_T, \mathbf{I}_T)$ is a circularly symmetric complex Gaussian random vector accounting for Rayleigh fast fading.

The total noise power $\eta_{m,k}^{(n)}$ at each user is modeled as

$$\eta_{m,k}^{(n)} = \sigma^2 + \sum_{m\in\mathcal{M}_{un}} \left(\frac{200}{d_{m,k}^{(n)}}\right)^{3.5} l_{m,k}^{(n)}\frac{P}{N}, \qquad (35)$$

where σ^2 is the thermal noise power, and the second term accounts for the uncoordinated inter-cell interference. Note that $\eta_{m,k}^{(n)}$ is used to obtain the normalized signal model (1). We assume equal power (i.e., $P_m = P$) for each BS m in the following simulation, and consider the system the performance under different signal-to-noise ratio (SNR), which is defined as $\gamma \triangleq P/\sigma^2$.

B. Convergence Behavior

We first illustrate the convergence behavior of the proposed distributed beamforming scheme, under three different utility functions, namely,

1) Proportional fairness utility: $U_{m,k}^{(n)}(\Gamma_{m,k}^{(n)}) = \frac{1}{NM}\log_2(\Gamma_{m,k}^{(n)})$;
2) α-fairness utility: $U_{m,k}^{(n)}(\Gamma_{m,k}^{(n)}) = \frac{1}{NM}\frac{(\Gamma_{m,k}^{(n)})^{(1-\alpha)}}{1-\alpha}$ with $(\alpha = 2)$;
3) Sum-rate utility: $U_{m,k}^{(n)}(\Gamma_{m,k}^{(n)}) = \frac{1}{NM}\log_2(1+\Gamma_{m,k}^{(n)})$.

The number of coordinated cells is $M = 7$; the number of sub-channels is $N = 3$; the number of transmit antennas at each BS is $T = 6$; the number of SDMA users is $Q = 3$; and the location parameter of the users is $D = 1000$m. Algorithm 1 is initialized by the channel-matched (CM) beamformers. Note that initializing with the more sophisticated beamformers, such as the in-cell zero-forcing (ICZF) beamformer, may only slightly increase the convergence speed.

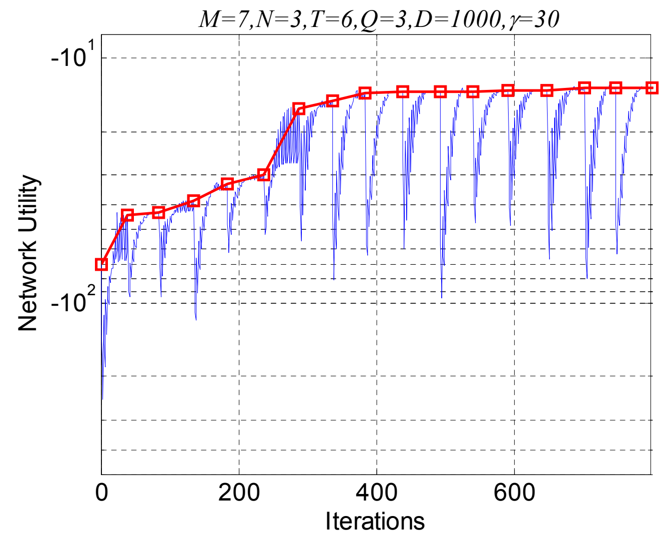


Fig. 4. Convergence of Algorithms 1 & 2 (α-fairness utility).

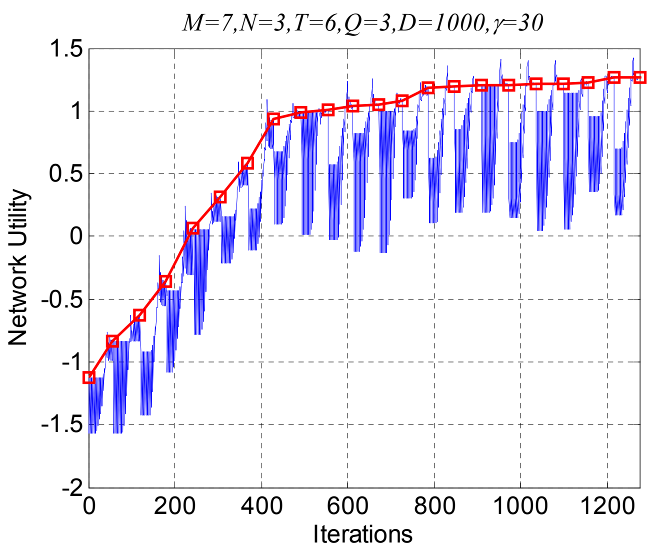


Fig. 3. Convergence of Algorithms 1 & 2 (proportion fairness utility).

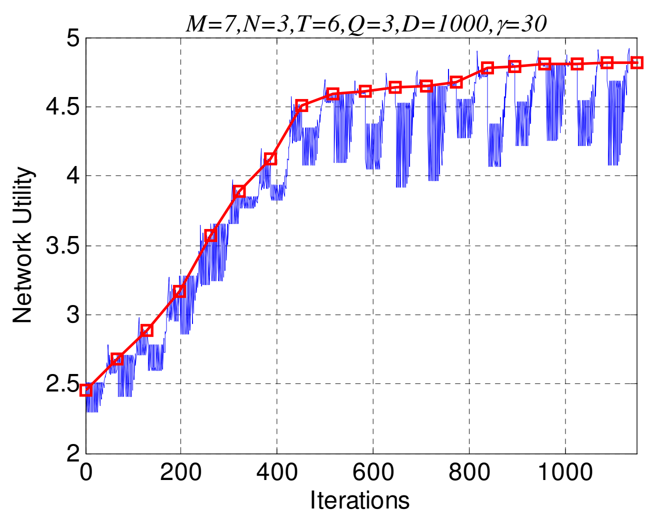


Fig. 5. Convergence of Algorithms 1 & 2 (weighted sum-rate utility).

The convergence behaviors of Algorithm 1 and Algorithm 2 at SNR $\gamma = 30$dB are shown in Figs. 3-5 for the above three utility functions, respectively. In these figures, the squares correspond to the outer iteration, i.e., the iterations of Algorithm 1. In each outer iteration, one BS updates its beam-vectors for all its users. The solid lines correspond to the inner iterations, i.e., the iterations of Algorithm 2. In each inner iteration, the BS updates the beam-vectors of one of its users. It is seen that the proposed distributed beamforming technique can significantly improve upon the initial network utility through optimizing the power allocation across the beams and the beam directions according to the conditions of the in-cell and out-cell interference. Moreover, the value of the network utility monotonically increases at each outer iteration, which confirms the theoretic result of Proposition 1.

Our extensive simulations reveal that the convergence speed of Algorithm 1 is affected by the system parameters as follows.

- The number of coordinated BSs M: A larger M corresponds to a slower convergence.

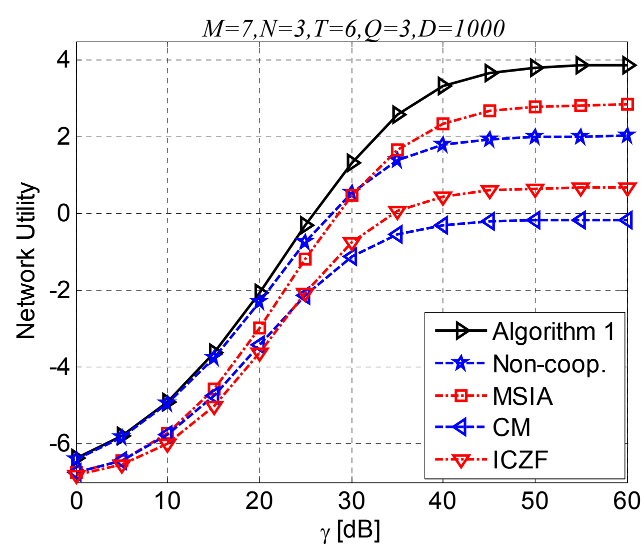


Fig. 6. Total proportional fairness utility versus SNR.

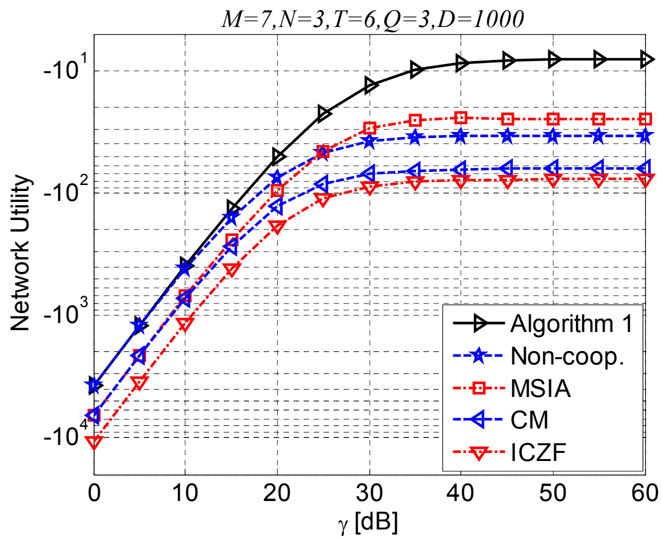


Fig. 7. Total α-fairness utility versus SNR.

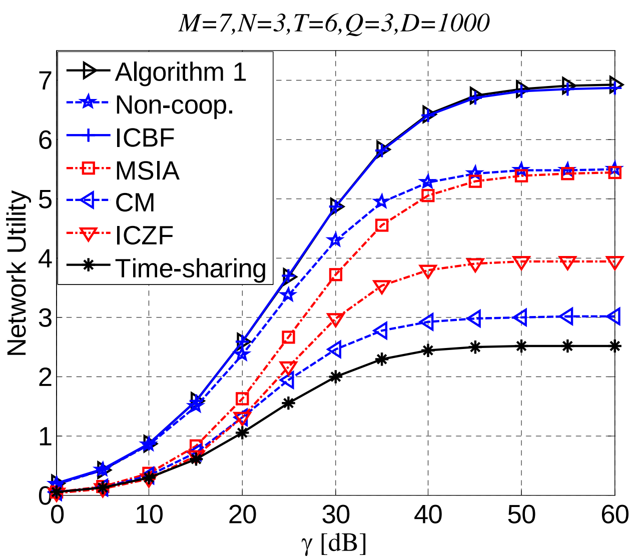


Fig. 8. Total weighted sum-rate utility versus SNR.

- The SNR γ: A larger γ corresponds to a slower convergence.
- The number of antennas T and the number of co-channel users Q: They only slightly affect the convergence of the outer iteration, i.e., Algorithm 1; but significantly affect the convergence of the inner iteration, i.e., Algorithm 2. Specifically, larger T and Q correspond to slower convergence of Algorithm 2.

In general it is seen that a relatively small number of outer iterations are sufficient for Algorithm 1 to converge. Note that there is no step-size parameter in Algorithm 1 and the overall utility could change dramatically in a single update, leading to rapid convergence. This is in contrast to some conventional distributed optimization algorithms, in which some step-size parameter controls the speed of convergence.

C. Performance Comparisons

We next compare the performance of the proposed distributed beamforming method with some existing techniques, including the simple channel matching (CM) method, the in-cell zero-forcing (ICZF) method, the iterative coordinated beam-forming (ICBF) method proposed in [13], and the more recent maximum SINR interference alignment (MSIA) method [21], as well as the approach based on the pure non-cooperative game (i.e., without the pricing mechanism) and the full-cooperation based method. Furthermore, the sum-rate performance of the time-sharing scheme is also considered, where the Q users in each cell access each sub-channel via TDMA.

In Figs. 6-8, the total utility values versus the SNR for the above-mentioned methods are plotted for the three utility functions, respectively. Note that for CM, ICZF and MSIA, the beamformer solutions are independent of the utility function. It is seen that the proposed distributed beamforming method outperform all other techniques in the sense of offering higher total network utilities. Moreover, we note that the network utility gain provided by Algorithm 1 is affected by the system parameters as follows.

- The distance from the BS to the user D: The network utility gain provided by Algorithm 1 is larger when the users are closer to the cell edge. Intuitively, the cell-edge users experience higher path losses and suffer from higher out-cell interference. Through Algorithm 1, both the available power across beams and the beam directions are optimized to mitigate these effects. On the other hand, when the users are close to the serving BS and away from the cell-edge, the per-cell optimized solution based on the non-cooperative game can achieve high performance without any information exchange among the BSs.
- The number of antennas T and the number of cochannel users Q: The network utility gain provided by Algorithm 1 is larger for larger T and Q. Intuitively, when the number of co-channel users Q is large, the users suffer from the high out-cell and in-cell interference. Due to the large number of antennas, Algorithm 1 has enough degrees of freedom to ensure the good quality of service of each user while causing minimum amount of in-cell and out-cell interference.
- The number of coordinated BSs M: The performance gain provided by Algorithm 1 becomes larger for larger M. This is because a larger M corresponds to a larger number of degrees of freedom to mitigate the interference.
- The signal-to-noise ratio γ: Algorithm 1 only provides marginal gains at low SNR, while the gain is more prominent at high SNR. This is because at high SNR, the interference becomes dominant factor for limiting the system performance, which can be effectively mitigated by Algorithm 1.

Comparison with non-cooperative game: As expected, the approach based on the pure non-cooperative game, denoted by Non-coop in Figs. 6-8, yields inferior performance in terms of the network utility. The reason is that when each BS optimizes only its own utility function, it does not account for the disutility it causes to the users served by other BSs due to the interference it generates. In economic terms, a disutility of one agent due to the action of another is referred to as a negative externality, which is the root of the inefficiencies of the non-cooperative game. For large M, the system becomes interference limited, and the pricing mechanism in Algorithm 1 can significantly increase the achievable network utility by implicitly inducing cooperation and yet maintaining the non-cooperative nature of the beamforming solution.

Comparison with full-cooperation based algorithm: The full-cooperation based algorithm has also been simulated. The results show that its utility performance is always similar to that of our proposed Algorithm 1. Thus, for clarity we did not plot its utility performance in Figs. 6-8. Notice that this performance similarity is reasonable: because the optimization problem of downlink beamforming in multi-cell OFDMA networks is non-convex, even full-cooperation based algorithm only achieves the KKT solution. For our proposed Algorithm 1, the total utility at the NE point is no less than that at the point satisfying the KKT conditions of the original problem.

Comparison with distributed interference alignment: The maximum SINR interference alignment (MSIA) method is an extension of the interference alignment algorithm proposed in [21], where the receive filters are chosen to maximize the SINR, and in the meantime to minimize the leakage interference at the receivers. For the MISO scenario discussed in this paper, interference alignment can be accomplished through symbol extension. In the simulations, the number of symbol extension S is set from 2 to 10, and the degree of freedom d for a user's message is set from 1 to its upper bound S. In Fig. (6)-(8), we plot the performance of MSIA for the case of the best choices of symbol extension S and the degree of freedom d.

It is seen that the performance of MSIA is inferior to that of Algorithm 1, especially at the strong interference scenario. In fact, the MSIA is even inferior to the per-cell optimized non-cooperative game solution in some scenarios. The reasons are as follows.

- The MSIA only optimizes the beam directions. In contrast, Algorithm 1 optimizes both the beam directions and the power distribution across the beams.
- The MSIA is designed to maximize the SINR and in the meantime to minimize the leakage interference. In contrast, Algorithm 1 is designed to optimize a general utility function.
- In MSIA, the iterative algorithm alternates between the original and reciprocal networks. Within each network, the receivers update their interference suppression filters. In contrast, Algorithm 1 is iteratively implemented only at the transmitters (BSs) and it can converge only in a small number of iterations. Thus, Algorithm 1 incurs a much lower information exchange overhead, in comparison with MSIA.

VI. Conclusions

We have considered the downlink beamforming problem for co-channel interference mitigation in multi-cell OFDMA networks. The problem is formulated as a general utility maximization problem subject to the per-cell power constraints, which is non-convex. We have proposed a distributed solution based on the non-cooperative game with pricing mechanism. We have shown that for some popular utility functions, such as the weighted sum rate utility, the proportional fairness utility, and the α-fairness utility, the proposed algorithm converges to a Nash equilibrium point. Moreover, we have developed an efficient algorithm to solve the KKT condition at each base station based on the dual decomposition technique. We have provided extensive simulation results to illustrate that the proposed method can converge to a Nash equilibrium in a small number of iterations, and it outperforms several state-of-the-art approaches to multicell interference mitigation, including the recently developed distributed interference alignment method.

In this paper, we have assumed the perfect instantaneous channel state information. From the practical point of view, the following issues remain to be investigated in future: (1) the impact of the reduced information exchange; (2) the robustness of the proposed method in the presence of transmission delay, packet loss and estimation error.

Appendix A: Proof of Proposition 1

After some simple manipulations, we have [4]

$$\frac{\partial^2 U_{m,k}^{(n)}(\mathbf{W}^{(n)})}{\partial\left(\mathcal{I}_{m,k}^{(n)}\right)^2} = \left(U_{m,k}^{(n)}\right)' \frac{\left|\vec{\mathbf{h}}_{m,k}^{(n)}\mathbf{w}_{m,k}^{(n)}\right|^2}{\left(1+\mathcal{I}_{m,k}^{(n)}\right)^3}\left(2-\kappa_{m,k}^{(n)}\right). \tag{36}$$

By the assumptions on the utility function $U_{m,k}^{(n)}$, we have $\left(U_{m,k}^{(n)}\right)' > 0$, and $0 \le \kappa_{m,k}^{(n)} \le 2$. Thus

$$\frac{\partial^2 U_{m,k}^{(n)}(\mathbf{W}^{(n)})}{\partial\left(\mathcal{I}_{m,k}^{(n)}\right)^2} \ge 0. \tag{37}$$

Hence $U_{m,k}^{(n)}$ is a convex function of $\mathcal{I}_{m,k}^{(n)}$. Then we have

$$\begin{aligned} U_{m,k}^{(n)}(\widehat{\mathbf{W}}^{(n)}) &\ge U_{m,k}^{(n)}(\mathbf{W}^{(n)}) + \frac{\partial U_{m,k}^{(n)}}{\partial \mathcal{I}_{m,k}^{(n)}}\Big|_{\mathcal{I}_{m,k}^{(n)}}\left(\widehat{\mathcal{I}}_{m,k}^{(n)} - \mathcal{I}_{m,k}^{(n)}\right) \\ &= U_{m,k}^{(n)}(\mathbf{W}^{(n)}) - \pi_{m,k}^{(n)}(\mathbf{W}^{(n)})\left(\widehat{\mathcal{I}}_{m,k}^{(n)} - \mathcal{I}_{m,k}^{(n)}\right), \end{aligned} \tag{38}$$

where $\mathcal{I}_{m,k}^{(n)}$ and $\widehat{\mathcal{I}}_{m,k}^{(n)}$ denote the interferences at the current operating point $\mathbf{W} = \{\mathbf{W}^{(n)}, n \in \mathcal{N}\}$ and at any new operating point $\widehat{\mathbf{W}} = \{\widehat{\mathbf{W}}^{(n)}, n \in \mathcal{N}\}$, respectively.

Summing up (38) over all users served by BS m and over all sub-channels, we have (39).

Hereafter we will drop the explicit dependency of $\pi_{m,k}^{(n)}$ on $\mathbf{W}^{(n)}$.

[4]Here, we drop the explicit dependency of $U_{m,k}^{(n)}$ on $\Gamma_{m,k}^{(n)}$, and denote $U_{m,k}^{(n)}(\Gamma_{m,k}^{(n)}(\mathbf{W}^{(n)}))$ as $U_{m,k}^{(n)}(\mathbf{W}^{(n)})$.

$$\sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{B}_m^{(n)}} U_{m,k}^{(n)}(\widehat{\mathbf{W}}^{(n)}) \geq \sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{B}_m^{(n)}} \left[U_{m,k}^{(n)}(\mathbf{W}^{(n)}) - \pi_{m,k}^{(n)}(\mathbf{W}^{(n)})\left(\widehat{\mathcal{I}}_{m,k}^{(n)} - \mathcal{I}_{m,k}^{(n)}\right)\right]. \tag{39}$$

Assume that BS m applies Algorithm 1 to update its beam-vectors, given the current beam-vectors $\mathbf{W} = \{\mathbf{W}^{(n)}, n \in \mathcal{N}\}$. According to the condition of STEP 6 in Algorithm 1, we have

$$\bar{U}_m(\bar{\mathbf{W}}_m, \mathbf{W}_{-m}) \geq \bar{U}_m(\mathbf{W}_m, \mathbf{W}_{-m}), \tag{40}$$

where $\bar{\mathbf{W}} = \{\mathbf{W}_1, \cdots, \mathbf{W}_{m-1}, \bar{\mathbf{W}}_m, \mathbf{W}_{m+1}, \cdots, \mathbf{W}_M\}$ denotes the operating point after BS m updates its beam-vectors $\bar{\mathbf{W}}_m = \{\bar{\mathbf{w}}_{m,k}^{(n)}, k \in \mathcal{B}_m^{(n)}, n \in \mathcal{N}\}$.

Plugging (6)-(10) into (40), we have (41).

Because of BS m's beam-vector update, the received interference by user $u \in \mathcal{B}_j^{(n)}, j \in \mathcal{M}\backslash m$ is changed from $\mathcal{I}_{j,u}^{(n)}$ to $\bar{\mathcal{I}}_{j,u}^{(n)}$, and

$$\bar{\mathcal{I}}_{j,u}^{(n)} - \mathcal{I}_{j,u}^{(n)} = \sum_{k\in\mathcal{B}_m^{(n)}}\left(\left|\vec{\mathbf{h}}_{m,u}^{(n)}\bar{\mathbf{w}}_{m,k}^{(n)}\right|^2 - \left|\vec{\mathbf{h}}_{m,u}^{(n)}\mathbf{w}_{m,k}^{(n)}\right|^2\right). \tag{42}$$

Thus, we have (43).

Adding the constant terms $\sum\limits_{\substack{j\in\mathcal{M}\\ j\neq m}}\left[\sum\limits_{n\in\mathcal{N}}\sum\limits_{u\in\mathcal{B}_j^{(n)}} U_{j,u}^{(n)}(\mathbf{W}^{(n)})\right]$ to both sides of (41), we have (44).

Subtracting the terms $\sum\limits_{n\in\mathcal{N}}\sum\limits_{k\in\mathcal{B}_m^{(n)}}\sum\limits_{j\in\mathcal{M}}\sum\limits_{\substack{u\in\mathcal{B}_j^{(n)}\\ (j,u)\neq(m,k)}} \pi_{j,u}^{(n)}\left|\vec{\mathbf{h}}_{m,u}^{(n)}\mathbf{w}_{m,k}^{(n)}\right|^2$ from both sides of (44), and using (43), we have (45).

Summing up (39) over $j \in \mathcal{M}\backslash m$ at the updated operating point $\bar{\mathbf{W}}$, we have

$$\begin{aligned}&\sum_{\substack{j\in\mathcal{M}\\ j\neq m}}\sum_{n\in\mathcal{N}}\sum_{u\in\mathcal{B}_j^{(n)}} U_{j,u}^{(n)}(\bar{\mathbf{W}}^{(n)}).\\ \geq &\sum_{\substack{j\in\mathcal{M}\\ j\neq m}}\sum_{n\in\mathcal{N}}\sum_{u\in\mathcal{B}_j^{(n)}}\left[U_{j,u}^{(n)}(\mathbf{W}^{(n)}) - \pi_{j,u}^{(n)}\left(\bar{\mathcal{I}}_{j,u}^{(n)} - \mathcal{I}_{j,u}^{(n)}\right)\right]\end{aligned} \tag{46}$$

Adding $\sum\limits_{n\in\mathcal{N}}\sum\limits_{k\in\mathcal{B}_m^{(n)}} U_{m,k}^{(n)}(\bar{\mathbf{W}}^{(n)})$ to both sides of (55), we have (47).

Combining (45) and (47) yields

$$\underbrace{\sum_{j\in\mathcal{M}}\sum_{n\in\mathcal{N}}\sum_{u\in\mathcal{B}_j^{(n)}} U_{j,u}^{(n)}(\bar{\mathbf{W}}^{(n)})}_{U_{\text{network}}(\bar{\mathbf{W}})} \geq \underbrace{\sum_{j\in\mathcal{M}}\sum_{n\in\mathcal{N}}\sum_{u\in\mathcal{B}_j^{(n)}} U_{j,u}^{(n)}(\mathbf{W}^{(n)})}_{U_{\text{network}}(\mathbf{W})}. \tag{48}$$

Hence, when BS m adjusts its beam-vectors, the total utility cannot decrease. Since at most one BS updates its beam-vectors at anytime, the total utility is non-decreasing in each iteration. As both the number of players and the size of the strategy sets are finite, the total utility is bounded. Thus, the total utility will convergence,

Now, we assume that Algorithm 1 converges to a fixed point $\mathbf{W}^* = (\mathbf{W}_1^*, \ldots, \mathbf{W}_M^*)$. If $\mathbf{W}^*$ is not an NE point, then there exists $\widetilde{\mathbf{W}} = (\mathbf{W}_1^*, \ldots, \widetilde{\mathbf{W}}_m, \ldots, \mathbf{W}_M^*)$, such that

$$\bar{U}_m(\widetilde{\mathbf{W}}_m, \mathbf{W}_{-m}^*) \geq \bar{U}_m(\mathbf{W}_m^*, \mathbf{W}_{-m}^*). \tag{49}$$

Applying the similar deduction in (40)-(48), we have

$$U_{\text{network}}(\widetilde{\mathbf{W}}) \geq U_{\text{network}}(\mathbf{W}^*), \tag{50}$$

which contradicts the assumption that $\mathbf{W}^*$ is a fixed point. Thus $\mathbf{W}^*$ is an NE point. $\square$

Appendix B: Proof of Proposition 2

The proof is along the similar line of that for Proposition 1 in [13]. There are two cases for the solution to the KKT conditions of (17).

Case 1: $\lambda_m > 0$ and $\mathbf{w}_{m,k}^{(n)} \neq \mathbf{0}$.

Notice that $\mathbf{T}_{m,k}^{(n)}$ is a positive-definite matrix. Thus, it is easy to have

$$\mathbf{T}_{m,k}^{(n)}\mathbf{w}_{m,k}^{(n)} \neq \mathbf{0}. \tag{51}$$

Obviously, (51) and (21) imply that $\vec{\mathbf{h}}_{m,k}^{(n)}\mathbf{w}_{m,k}^{(n)} \neq \mathbf{0}$ and $\mathbf{h}_{m,k}^{(n)} \propto \mathbf{T}_{m,k}^{(n)}\mathbf{w}_{m,k}^{(n)}$. Hence, a non-zero solution $\mathbf{w}_{m,k}^{(n)*}$ to the KKT conditions of (17) must be of the form

$$\mathbf{w}_{m,k}^{(n)*} \propto \mathbf{T}_{m,k}^{(n)\dagger}\mathbf{h}_{m,k}^{(n)}. \tag{52}$$

Case 2: $\lambda_m = 0$ and $\mathbf{w}_{m,k}^{(n)} \neq \mathbf{0}$.

It is easy to see that (21) is satisfied only if one of the following two conditions holds:

(a) $\vec{\mathbf{h}}_{m,k}^{(n)}\mathbf{w}_{m,k}^{(n)} = 0$ and $\mathbf{L}_{m,k}^{(n)}\mathbf{w}_{m,k}^{(n)} = 0$;

(b) $\mathbf{h}_{m,k}^{(n)} \in \varrho(\mathbf{L}_{m,k}^{(n)})$.

If (a) holds, $\vec{\mathbf{h}}_{m,k}^{(n)}\mathbf{w}_{m,k}^{(n)} = 0$ implies that the non-zero beam-vector $\mathbf{w}_{m,k}^{(n)}$ is orthogonal to the channel vector $\vec{\mathbf{h}}_{m,k}^{(n)}$. In this case, user $k \in \mathcal{B}_m^{(n)}$ cannot receive any information from the serving base station. Hence we discard the solutions $\vec{\mathbf{h}}_{m,k}^{(n)}\mathbf{w}_{m,k}^{(n)} = 0$ and $\mathbf{L}_{m,k}^{(n)}\mathbf{w}_{m,k}^{(n)} = \mathbf{0}$ for the case $\lambda_m \neq 0$ and $\mathbf{w}_{m,k}^{(n)} \neq \mathbf{0}$.

If (b) holds and $\lambda_m = 0$, a non-zero beam-vector which satisfies (21) must be of the form

$$\mathbf{w}_{m,k}^{(n)*} \propto \mathbf{L}_{m,k}^{(n)\dagger}\mathbf{h}_{m,k}^{(n)}. \tag{53}$$

Notice that (53) is equivalent to (52) with $\lambda_m = 0$. Meanwhile, notice that multiplying $\mathbf{w}_{m,k}^{(n)*}$ by any unit-norm complex number does not affect either the objective function or the power constraint in (17). Hence, for the two cases discussed above, we can set the unique solution to the KKT conditions of (17) as

$$\mathbf{w}_{m,k}^{(n)*} = \beta_{m,k}^{(n)}\mathbf{T}_{m,k}^{(n)\dagger}\mathbf{h}_{m,k}^{(n)}, \tag{54}$$

where $\beta_{m,k}^{(n)}$ is some scalar constant.

To determine $\beta_{m,k}^{(n)}$, we plug (54) into (21) to obtain

$$\begin{aligned}&U_{m,k}^{(n)}\left(\frac{\left|\vec{\mathbf{h}}_{m,k}^{(n)}\beta_{m,k}^{(n)}\mathbf{T}_{m,k}^{(n)\dagger}\mathbf{h}_{m,k}^{(n)}\right|^2}{1+\mathcal{I}_{m,k}^{(n)}}\right)' \frac{\mathbf{h}_{m,k}^{(n)}\vec{\mathbf{h}}_{m,k}^{(n)}\beta_{m,k}^{(n)}\mathbf{T}_{m,k}^{(n)\dagger}\mathbf{h}_{m,k}^{(n)}}{1+\mathcal{I}_{m,k}^{(n)}}\\ &= \mathbf{T}_{m,k}^{(n)}\beta_{m,k}^{(n)}\mathbf{T}_{m,k}^{(n)\dagger}\mathbf{h}_{m,k}^{(n)}\end{aligned} \tag{55}$$

$$\sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{B}_m^{(n)}}\left(U_{m,k}^{(n)}\left(\bar{\mathbf{W}}^{(n)}\right)-\sum_{j\in\mathcal{M}}\sum_{\substack{u\in\mathcal{B}_j^{(n)}\\(j,u)\neq(m,k)}}\pi_{j,u}^{(n)}\left|\vec{\mathbf{h}}_{m,u}^{(n)}\bar{\mathbf{w}}_{m,k}^{(n)}\right|^2\right)$$
$$\geq\sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{B}_m^{(n)}}\left(U_{m,k}^{(n)}\left(\mathbf{W}^{(n)}\right)-\sum_{j\in\mathcal{M}}\sum_{\substack{u\in\mathcal{B}_j^{(n)}\\(j,u)\neq(m,k)}}\pi_{j,u}^{(n)}\left|\vec{\mathbf{h}}_{m,u}^{(n)}\mathbf{w}_{m,k}^{(n)}\right|^2\right). \tag{41}$$

$$\sum_{\substack{j\in\mathcal{M}\\j\neq m}}\sum_{n\in\mathcal{N}}\sum_{u\in\mathcal{B}_j^{(n)}}\pi_{j,u}^{(n)}\left(\bar{\mathcal{I}}_{j,u}^{(n)}-\mathcal{I}_{j,u}^{(n)}\right)=\sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{B}_m^{(n)}}\sum_{j\in\mathcal{M}}\sum_{\substack{u\in\mathcal{B}_j^{(n)}\\(j,u)\neq(m,k)}}\pi_{j,u}^{(n)}\left(\left|\vec{\mathbf{h}}_{m,u}^{(n)}\bar{\mathbf{w}}_{m,k}^{(n)}\right|^2-\left|\vec{\mathbf{h}}_{m,u}^{(n)}\mathbf{w}_{m,k}^{(n)}\right|^2\right). \tag{43}$$

$$\sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{B}_m^{(n)}}\left[U_{m,k}^{(n)}\left(\bar{\mathbf{W}}^{(n)}\right)-\sum_{j\in\mathcal{M}}\sum_{\substack{u\in\mathcal{B}_j^{(n)}\\(j,u)\neq(m,k)}}\pi_{j,u}^{(n)}\left|\vec{\mathbf{h}}_{m,u}^{(n)}\bar{\mathbf{w}}_{m,k}^{(n)}\right|^2\right]+\sum_{\substack{j\in\mathcal{M}\\j\neq m}}\sum_{n\in\mathcal{N}}\sum_{u\in\mathcal{B}_j^{(n)}}U_{j,u}^{(n)}(\mathbf{W}^{(n)})$$
$$=\sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{B}_m^{(n)}}U_{m,k}^{(n)}\left(\bar{\mathbf{W}}^{(n)}\right)+\sum_{\substack{j\in\mathcal{M}\\j\neq m}}\sum_{n\in\mathcal{N}}\sum_{u\in\mathcal{B}_j^{(n)}}U_{j,u}^{(n)}(\mathbf{W}^{(n)})-\sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{B}_m^{(n)}}\sum_{j\in\mathcal{M}}\sum_{\substack{u\in\mathcal{B}_j^{(n)}\\(j,u)\neq(m,k)}}\pi_{j,u}^{(n)}\left|\vec{\mathbf{h}}_{m,u}^{(n)}\bar{\mathbf{w}}_{m,k}^{(n)}\right|^2 \tag{44}$$
$$\geq\sum_{j\in\mathcal{M}}\sum_{n\in\mathcal{N}}\sum_{u\in\mathcal{B}_j^{(n)}}U_{j,u}^{(n)}(\mathbf{W}^{(n)})-\sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{B}_m^{(n)}}\sum_{j\in\mathcal{M}}\sum_{\substack{u\in\mathcal{B}_j^{(n)}\\(j,u)\neq(m,k)}}\pi_{j,u}^{(n)}\left|\vec{\mathbf{h}}_{m,u}^{(n)}\mathbf{w}_{m,k}^{(n)}\right|^2.$$

$$\sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{B}_m^{(n)}}U_{m,k}^{(n)}\left(\bar{\mathbf{W}}^{(n)}\right)+\sum_{\substack{j\in\mathcal{M}\\j\neq m}}\sum_{n\in\mathcal{N}}\sum_{u\in\mathcal{B}_j^{(n)}}U_{j,u}^{(n)}(\mathbf{W}^{(n)})-\sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{B}_m^{(n)}}\sum_{j\in\mathcal{M}}\sum_{\substack{u\in\mathcal{B}_j^{(n)}\\(j,u)\neq(m,k)}}\pi_{j,u}^{(n)}\left(\left|\vec{\mathbf{h}}_{m,u}^{(n)}\bar{\mathbf{w}}_{m,k}^{(n)}\right|^2-\left|\vec{\mathbf{h}}_{m,u}^{(n)}\mathbf{w}_{m,k}^{(n)}\right|^2\right)$$
$$=\sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{B}_m^{(n)}}U_{m,k}^{(n)}\left(\bar{\mathbf{W}}^{(n)}\right)+\sum_{\substack{j\in\mathcal{M}\\j\neq m}}\sum_{n\in\mathcal{N}}\sum_{u\in\mathcal{B}_j^{(n)}}U_{j,u}^{(n)}(\mathbf{W}^{(n)})-\sum_{\substack{j\in\mathcal{M}\\j\neq m}}\sum_{n\in\mathcal{N}}\sum_{u\in\mathcal{B}_j^{(n)}}\pi_{j,u}^{(n)}\left(\bar{\mathcal{I}}_{j,u}^{(n)}-\mathcal{I}_{j,u}^{(n)}\right)$$
$$=\sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{B}_m^{(n)}}U_{m,k}^{(n)}\left(\bar{\mathbf{W}}^{(n)}\right)+\sum_{\substack{j\in\mathcal{M}\\j\neq m}}\sum_{n\in\mathcal{N}}\sum_{u\in\mathcal{B}_j^{(n)}}\left[U_{j,u}^{(n)}(\mathbf{W}^{(n)})-\pi_{j,u}^{(n)}\left(\bar{\mathcal{I}}_{j,u}^{(n)}-\mathcal{I}_{j,u}^{(n)}\right)\right]\geq\sum_{j\in\mathcal{M}}\sum_{n\in\mathcal{N}}\sum_{u\in\mathcal{B}_j^{(n)}}U_{j,u}^{(n)}(\mathbf{W}^{(n)}). \tag{45}$$

$$\sum_{j\in\mathcal{M}}\sum_{n\in\mathcal{N}}\sum_{u\in\mathcal{B}_j^{(n)}}U_{j,u}^{(n)}(\bar{\mathbf{W}}^{(n)})\geq\sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{B}_m^{(n)}}U_{m,k}^{(n)}(\bar{\mathbf{W}}^{(n)})+\sum_{\substack{j\in\mathcal{M}\\j\neq m}}\sum_{n\in\mathcal{N}}\sum_{u\in\mathcal{B}_j^{(n)}}\left[U_{j,u}^{(n)}(\mathbf{W}^{(n)})-\pi_{j,u}^{(n)}\left(\bar{\mathcal{I}}_{j,u}^{(n)}-\mathcal{I}_{j,u}^{(n)}\right)\right]. \tag{47}$$

Considering $\mathbf{T}_{m,k}^{(n)}\beta_{m,k}^{(n)}\mathbf{T}_{m,k}^{(n)\dagger}=\beta_{m,k}^{(n)}\mathbf{I}$, and $\mathbf{h}_{m,k}^{(n)}\neq\mathbf{0}$, we have

$$U_{m,k}^{(n)}\Big(\frac{|\vec{\mathbf{h}}_{m,k}^{(n)}\beta_{m,k}^{(n)}\mathbf{T}_{m,k}^{(n)\dagger}\mathbf{h}_{m,k}^{(n)}|^2}{1+\mathcal{I}_{m,k}^{(n)}}\Big)'=\frac{1+\mathcal{I}_{m,k}^{(n)}}{\vec{\mathbf{h}}_{m,k}^{(n)}\mathbf{T}_{m,k}^{(n)\dagger}\mathbf{h}_{m,k}^{(n)}} \tag{56}$$

which is equivalent to

$$\frac{\left|\vec{\mathbf{h}}_{m,k}^{(n)}\beta_{m,k}^{(n)}\mathbf{T}_{m,k}^{(n)\dagger}\mathbf{h}_{m,k}^{(n)}\right|^2}{1+\mathcal{I}_{m,k}^{(n)}}=\mathrm{Inv}\Big\{U_{m,k}^{(n)}\Big(\frac{1+\mathcal{I}_{m,k}^{(n)}}{\vec{\mathbf{h}}_{m,k}^{(n)}\mathbf{T}_{m,k}^{(n)\dagger}\mathbf{h}_{m,k}^{(n)}}\Big)'\Big\}. \tag{57}$$

Thus, we obtain

$$\beta_{m,k}^{(n)}=\sqrt{\left(1+\mathcal{I}_{m,k}^{(n)}\right)\Phi_{m,k}^{(n)}\Upsilon_{m,k}^{(n)}}. \tag{58}$$

□

References

[1] D. Gesbert, S. Hanly, H. Huang, S. Shamai, O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE J. Select. Areas Commun.*, vol. 28, no. 9, pp. 1–29, 2010.

[2] H. Zhang and H. Dai, "Cochannel interference mitigation and cooperative processing in downlink multicell multiuser mimo networks," *EURASIP J. Wireless Commun. Netw.*, vol. 2004, no. 2, pp. 222–235, 2004.

[3] B. Ng, J. Evans, S. Hanly, and D. Aktas, "Distributed downlink beamforming with cooperative base stations," *IEEE Tran. Inform. Theory*, vol. 54, no. 12, pp. 5491–5499, 2008.

[4] B. Somekh, O.and Zaidel and S. Shamai, "Sum rate characterization of joint multiple cell-site processing," *IEEE Trans. Inform. Theory*, vol. 53, no. 12, pp. 4473–4497, 2007.

[5] S. Shamai, O. S. A. Z. B. M. Somekh, O.and Simeone, and H. V. Poor, "Cooperative multi-cell networks: Impact of limited-capacity backhaul and inter-users links," in *Proc. Joint Workshop on Coding and Communications*, Durnstein, Austria, pp. 14–16, 2007.

[6] O. Simeone, O. Somekh, H. V. Poor, and S. Shamai, "Local base station cooperation via finite-capacity links for the uplink of linear cellular networks," *IEEE Trans. Inform. Theory*, vol. 55, no. 1, pp. 190–204, 2009.

[7] T. Tamaki, K. Seong, and J. Cioffi, "Downlink MIMO systems using cooperation among base stations in a slow fading channel," *Proc. 2007 IEEE Int. Conf. Commun. (ICC'07)*, pp. 4728–4733, 2007.

[8] P. Marsch and G. Fettweis, "On downlink network mimo under a constrained backhaul and imperfect channel knowledge," *Proc. 2009 IEEE Global Telecommun. Conf. (Globecom'09)*, 2009.

[9] M. Awad, V. Mahinthan, M. Mehrjoo, X. Shen, and J.W. Mark, "A Dual Decomposition-based Resource Allocation for OFDMA Networks with Imperfect CSI," *IEEE Trans. Veh. Commun.*, vol. 59, no. 5, pp. 2394–2403, 2010.

[10] A. Papadogiannis, E. Hardouin, and D. Gesbert, "Decentralising multi-cell cooperative processing on the downlink: a novel robust framework," *EURASIP J. Wireless Commun. Netw.*, August 2009.

[11] H. Dahrouj and W. Yu, "Coordinated beamforming for the multicell multi-antenna wireless systems," *IEEE Trans. Wireless Commun.*, vol. 9, pp. 1748–1759, 2010.

[12] R. Zakhour and D. Gesbert, "Distributed multicell MISO precoding using the layered virtual SINR framework," *IEEE Trans. Wireless Commun.*, vol. 9, pp. 2444–2448, 2010.

[13] L. Venturino, N. Prasad, and X. Wang, "Coordinated linear beamforming in downlink multi-cell wireless networks," *IEEE Trans. Wireless Commun.*, vol. 9, pp. 1451–1461, 2010.

[14] R. Zhang and S. Cui, "Cooperative interference management with MISO beamforming," *IEEE Trans. Sig. Proc.*, vol. 58, no. 10, pp. 5450–5458, 2010.

[15] E. Jorswieck, E. Larsson, M. Luise, and H. Poor, "Game theory in signal processing and communications," *IEEE Sig. Proc. Mag.*, vol. 26, no. 5, pp. 17–132, 2009.

[16] J. Chen, Q. Yu, P. Cheng, Y. Sun, Y. Fan, and X. Shen, "Game Theoretical Approach for Channel Allocation in Wireless Sensor and Actuator Networks," DOI: 10.1109/TAC.2011.2164014, *to appear in IEEE Trans. Autom. Control*, 2011.

[17] Q. Yu, J. Chen, Y. Fan, X. Shen and Y. Sun, "Multi-Channel Assignment in Wireless Sensor Networks: A Game Theoretic Approach," *Proceedings of IEEE INFOCOM*, San Diego, Ca, USA, March 15-19, 2010.

[18] X. Xiao, Q. Zhang, Y. Shi, and Y. Gao, "How Much to Share: A Repeated Game Model for Peer-to-Peer Streaming under Service Differentiation Incentives," *to appear in IEEE Trans. Parallel Distrib. Syst.*, 2011.

[19] Z. Han, D. Niyato, W. Saad, T. Basar, and A. Hjorungnes, "Game Theory in Wireless and Communication Networks: Theory, Models and Applications," *Cambridge University Press*, UK, 2011.

[20] J. Mo, J.and Walrand, "Fair end-to-end window-based congestion control," *IEEE/ACM Trans. Netw.*, vol. 8, no. 5, pp. 556–567, 2000.

[21] K. Gomadam, V. R. Cadambe, and S. A. Jafar, "Approaching the capacity of wireless networks through distributed interference alignment," *arXiv: 0803.3816, e-print*.

[22] E. G. Larsson and E. Jorswieck, "Competition versus cooperation on the miso interference channel," *IEEE J. Select. Areas Commun.*, vol. 26, no. 9, pp. 1059–1069, Sep. 2008.

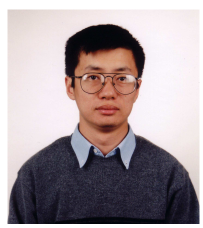
[23] E. A. Jorswieck, E. G. Larsson, and D. Danev, "Complete characterization of the Pareto boundary for the MISO interference channel," *IEEE Trans. Sig. Proc.*, vol. 56, no. 10-2, pp. 5292–5296, 2008.

[24] C. Saraydar, N. Mandayam, and D. Goodman, "Efficient power control via pricing in wireless data networks," *IEEE Tran. Commun.*, vol. 50, no. 2, pp. 291–303, 2002.

[25] J. Huang, R. A. Berry, and M. L. Honig, "Distributed interference compensation for wireless networks," *IEEE J. Select. Areas Commun.*, vol. 24, pp. 1074–1084, May 2006.

[26] D. A. Schmidt, C. Shi, R. A. Berry, M. L. Honig, and W. Utschick, "Distributed resource allocation schemes: Pricing algorithms for power control and beamformer design in interference networks," *IEEE Sig. Proc. Mag.*, vol. 26, no. 5, pp. 53–63, 2009.

Weiqiang Xu (M'09) received his M.Sc. degree in Communications and Information System from Southwest Jiao-Tong University, China, and his Ph.D. degree in Control Science and Engineering from Zhejiang University, China, in 2003 and 2006, respectively. He also was a postdoctor research fellow with the group of Networked Sensing and Control in the State Key laboratory of Industrial Control Technology, Zhejiang University, China. From Oct. 2009 to Oct. 2010, he visited Prof. Xiaodong Wang's research group in Electrical Engineering Department at Columbia University, New York. He is currently a professor with the School of Information Science and Technology, Zhejiang Sci-Tech University, Hangzhou, China. His research interests include multi-cell networks, Ad Hoc networks, wireless sensor networks, wireless optical networks, congestion control, and networked control system, etc. He has served as a TPC member for IEEE Globecom 2012, IWCMC 2009, IWCMC 2010, PMSN 2009, IHMSC 2009, IHMSC 2010, IHMSC 2011, IHMSC 2012. He has also served as a peer reviewer for a variety of IEEE journals and conferences.



Xiaodong Wang (S'97-M'98-SM'04-F'08) received the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ.

He is a Professor in the Department of Electrical Engineering, Columbia University, New York. His research interests fall in the general areas of computing, signal processing, and communications, and he has published extensively in these areas. Among his publications is a book entitled Wireless Communication Systems: Advanced Techniques for Signal Reception (Prentice-Hall, 2003). His current research interests include wireless communications, statistical signal processing, and genomic signal processing. Dr. Wang received the 1999 NSF CAREER Award, and the 2001 IEEE Communications Society and Information Theory Society Joint Paper Award. He has served as an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS, the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and the IEEE TRANSACTIONS ON INFORMATION THEORY.