This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2014.2343255, IEEE Transactions on Vehicular Technology

Energy-Efficient Coordinated Beamforming Under Minimal Data Rate Constraint of Each User

Yang Li, *Student Member, IEEE,* Yafei Tian, *Member, IEEE,* and Chenyang Yang, *Senior Member, IEEE*

***Abstract*—Coordinated beamforming has been optimized to maximize the sum rate under the transmit power constraints, or to minimize the transmit power under the data rate constraints. In this paper, we design precoder that maximizes the energy efficiency (EE) of multi-cell multi-antenna downlink network with coordination and meanwhile ensures the minimal data rate requirement of each user. To find a solution of such a non-convex optimization problem, we construct a convex subset of the original constraint set and a quasi-concave lower bound of the EE. Then, we propose an iterative algorithm to maximize the lower bound of the EE within the convex subset, which is shown converging rapidly to a local optimum of the original problem. We evaluate the EE of the proposed algorithm through simulations under different user locations, data rate requirements, and numbers of antennas. The results demonstrate that the proposed precoder performs closely to an upper bound derived from interference-free assumption. When the circuit power consumption is dominant, the proposed precoder is much more energy-efficient than the transmit power minimization precoder as well as a precoder including an interference alignment beamformer and the optimized transmit power maximizing the EE.**

***Index Terms*—Coordinated beamforming, energy efficiency (EE), precoding design, quality of service (QoS)**

I. Introduction

To reduce the operational cost of the networks and the global greenhouse gas emissions, energy efficiency (EE) is becoming an important design goal for cellular systems [2], [3]. For high throughput downlink cellular networks, EE can be defined as the ratio of the sum rate to the total power consumption of the network, which involves transmit power and circuit power. It has been recognized that the well-explored transmission schemes designed toward high spectral efficiency (SE) does not necessarily provide high EE [4], [5]. To improve the EE, we should increase the SE and meanwhile decrease the total power consumption.

Inter-cell interference (ICI) is the major limiting factor for improving the SE of full frequency reuse cellular networks. When the base stations (BSs) with multiple antennas can share some form of information related to channel, coordinated beamforming can effectively alleviate the impact of ICI, which has been optimized from different perspectives.

The coordinated beamforming[1] that optimizes the SE subject to transmit power constraints has been widely studied, e.g., [6]–[8]. In [6], beamforming and power allocation were jointly designed for maximizing the weighted sum rate. In [7], user scheduling, beamforming and power allocation were jointly optimized to maximize sum rate, and a local optimal solution was found with an iterative algorithm. In [8], the precoders with BS coordination were optimized to maximize weighted sum rate, where the proposed algorithm was proved to converge to a stationary point of the original problem.

The coordinated beamforming that minimizes the total transmit power of coordinative BSs subject to the signal-to-interference-plus-noise ratio (SINR) constraint of each user has also been well-investigated, e.g., [9]–[11], all for users with a single antenna. The obtained precoders in [9]–[11] are optimal. However, the optimization approaches in these works are all based on transforming the SINR constraints into a second-order-cone constraint, which was proposed in [12] assuming fixed linear detectors. As a result, when each user is equipped with multiple antennas, these approaches to find the optimal solutions are no longer applicable.

For cellular networks, the basic principle of energy-efficient design is maximizing the EE of the system without sacrificing the required quality of service (QoS) of each user. To reflect typical QoS constraints imposed by different kinds of traffics such as best effort, real and non-real time services, a typical way is to set a minimal data rate requirement for each user [13]. With such QoS constraints, the design minimizing the transmit power is not always energy-efficient. When the circuit power dominates the total power consumption of a network, which is a typical scenario of prevalent cellular networks, transmitting with higher data rate than the required minimal data rate may achieve higher EE.

Although EE has been extensively studied for various systems and important progress has been made in resource allocation and performance evaluation, e.g., [14]–[19], so far transmission strategy design for interference networks was seldom studied in literature. In [20], a non-cooperative power allocation for energy-efficient design was investigated for a single antenna multi-cell multi-carrier system. In [21], an energy-efficient precoding including beamforming and transmit power was proposed for multi-input-multi-output (MIMO)

Manuscript received December 5, 2013; revised March 8, 2014 and June 6, 2014; accepted July 4, 2014. This work was supported by the National Natural Science Foundation of China under Grants 61371077, and by the National Basic Research Program of China (973 Program) under Grant 2012CB316003. Part of this work was previously presented in IEEE PIMRC, London, UK, September 2013 [1]. The review of this paper was coordinated by Prof. C.-X. Wang.

Y. Li, Y. Tian, and C. Yang are with the School of Electronics and Information Engineering, Beihang University, Beijing, 100191, P. R. China (e-mail: liyang2012@ee.buaa.edu.cn; {ytian, cyyang}@buaa.edu.cn). The correspondence author is Yafei Tian.

[1] In this paper, coordinated beamforming is a general notion for precoding with coordination among multiple BSs. Therefore, optimizing precoding explicitly or implicitly includes optimizing beamforming and transmit power.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2014.2343255, IEEE Transactions on Vehicular Technology

interference networks. In both works, no QoS constraints were imposed, so they are applicable to best effort traffics. Besides, since in both works the EE was defined as the sum of per-user EE instead of the ratio of sum rate to total power consumption, these strategies cannot be applied to downlink cellular networks.

In this paper, we study energy-efficient precoding design for downlink MIMO cellular networks with coordinated beamforming, where each coordinative BS and each user are equipped with multiple antennas. This is a kind of MIMO interference network. Specifically, we strive to optimize the precoding to maximize the EE of the network under the minimal data rate requirement of each user, where both the transmit power and circuit power consumptions are taken into account. Because both the objective function and constraints are non-convex, the optimization problem is non-convex. An iterative algorithm is proposed to find a solution of the problem. Simulation results show that the EE achieved by the proposed precoder is close to an upper bound derived from interference-free assumption and higher than that achieved by the transmit power minimization precoder in a wide range of data rate requirements.

The main contributions of this paper are summarized as follows:

1) Differing from [20] and [21], we consider minimal data rate constraint of each user for a downlink cellular network. As a result, the proposed precoder can accommodate both real time and best effort traffics by setting different values of the data rate requirement.
2) We find a solution of the original non-convex problem by constructing a convex subset of the constraints and a quasi-concave lower bound of the objective function. Then we propose an iterative algorithm to maximize the lower bound within the convex subset, and prove that the algorithm can converge to a stationary point of the original problem. We use linear interference alignment (IA) to initialize the iterative algorithm, and prove that it provides a feasible solution and is approximately optimal in EE when interference is strong. By judiciously designing the initial value, the algorithm yields a near-optimal solution.

The rest of the paper is organized as follows. In Section II, the system model is introduced. In Section III, the problem of the energy-efficient precoder design is formulated and transformed to another optimization problem. Then, an iterative algorithm is developed to find the solution of the problem, the convergence of the algorithm is proved, and the initialization and feasibility are addressed. Simulation results are shown in Section IV, where the impact of user locations, QoS requirements, and the numbers of antennas are analyzed. Finally, Sections V concludes the paper.

Notation: The superscript $(\cdot)^H$ denotes the Hermitian transpose of a matrix. The $n \times n$ identity matrix is denoted by $\mathbf{I}_n$. The complex Gaussian distribution is denoted by $\mathcal{CN}(\cdot,\cdot)$. The symbols $\mathbb{E}(\cdot)$, $\mathrm{Tr}(\cdot)$, and $\det(\cdot)$ denote expectation, trace, and determinant operators, respectively. The symbol $\nabla f(\cdot)$ denotes the gradient of the function $f(\cdot)$.

II. System Model

Consider a K-cell downlink MIMO interference network, where each BS serves multiple users. The k-th BS, which conveys data to I_k users, is equipped with M_k antennas. The i-th user in cell k, user i_k, is equipped with N_{i_k} antennas. We assume that every BS has perfect channel information from itself to all users in the K cells, and the channels are shared among the coordinated BSs but the data for the users is not shared.

The signal received at user i_k can be expressed as

$$\mathbf{y}_{i_k} = \mathbf{H}_{i_k k}\mathbf{V}_{i_k}\mathbf{x}_{i_k} + \underbrace{\sum_{m=1, m\neq i}^{I_k} \mathbf{H}_{i_k k}\mathbf{V}_{m_k}\mathbf{x}_{m_k}}_{\text{multi-user interference}} + \underbrace{\sum_{j\neq k, j=1}^{K}\sum_{l=1}^{I_j}\mathbf{H}_{i_k j}\mathbf{V}_{l_j}\mathbf{x}_{l_j}}_{\text{inter-cell interference}} + \mathbf{n}_{i_k}, \quad \forall i_k \in \mathcal{I} \tag{1}$$

where $\mathbf{x}_{i_k}$ is the transmitted symbol vector to user i_k conveying d_{i_k} data streams with $\mathbb{E}[\mathbf{x}_{i_k}\mathbf{x}_{i_k}^H] = \mathbf{I}_{d_{i_k}}$, $\mathbf{V}_{i_k}$ is the $M_k \times d_{i_k}$ precoding matrix including beamforming and transmit power implicitly, $\mathbf{H}_{i_k j}$ is the $N_{i_k} \times M_j$ channel matrix from BS j to user i_k, $\mathcal{I} = \{i_k | i \in \{1,2,\ldots,I_k\}, \quad k \in \{1,2,\ldots,K\}\}$ is the set of all users, and $\mathbf{n}_{i_k}$ is an additive white Gaussian noise vector subject to $\mathcal{CN}(0, \sigma_{i_k}^2 \mathbf{I}_{N_{i_k}})$.

With the linear precoding at the BSs and maximal likelihood detector at each user, the achievable rate of user i_k can be expressed as [22],

$$R_{i_k} = \log_2 \det\left(\mathbf{I}_{N_{i_k}} + \mathbf{H}_{i_k k}\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H\mathbf{H}_{i_k k}^H\mathbf{J}_{i_k}^{-1}\right) \tag{2}$$

where $\mathbf{J}_{i_k} = \sum_{(l,j)\neq(i,k)}\mathbf{H}_{i_k j}\mathbf{V}_{l_j}\mathbf{V}_{l_j}^H\mathbf{H}_{i_k j}^H + \sigma_{i_k}^2\mathbf{I}_{N_{i_k}}$ is the interference-plus-noise covariance matrix of user i_k.

For notational simplicity, we use $\mathbf{V}$ to represent $\{\mathbf{V}_{i_k}\}_{i_k\in\mathcal{I}}$ which are all the precoding matrices for the users in K cells.

III. Problem Formulation and Optimization Algorithm

In this section, we start by formulating the problem to optimize the EE of the network under the minimal data rate constraint of each user. Considering that the original problem is non-convex, we convert it into a new optimization problem by constructing a convex subset of the constraints and a quasi-concave lower bound of the objective function, which is then solved by an iterative algorithm. We proceed to prove that the iterative algorithm converges to a stationary point of the original problem, and address the initial value selection.

A. Problem Formulation

The EE of the network is defined as the ratio of the sum rate to the overall power consumption [2], which is

$$\eta(\mathbf{V}) = \frac{\sum_{k=1}^{K}\sum_{i=1}^{I_k} R_{i_k}}{\sum_{k=1}^{K}\sum_{i=1}^{I_k}\rho\mathrm{Tr}(\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H) + \sum_{k=1}^{K} M_k P_c + K P_o} \tag{3}$$

where ρ is the reciprocal of the power amplifier efficiency, $\mathrm{Tr}(\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H) = P_{i_k}$ is the transmit power for user i_k, P_c is the circuit power consumed at each antenna consisting of the bandpass filters, duplexers, and other radio frequency circuits, and P_o is the power consumption at each BS consisting of power supply and cooling.

We strive to design the precoding matrices for all users that maximize the EE of the network under the QoS constraint imposed by each user. From (2) and (3), the optimization problem can be formulated as

$$\max_{\mathbf{V}} \frac{\sum_{k=1}^{K}\sum_{i=1}^{I_k} \log_2 \det\left(\mathbf{I}_{N_{i_k}} + \mathbf{H}_{i_k k}\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H\mathbf{H}_{i_k k}^H\mathbf{J}_{i_k}^{-1}\right)}{\sum_{k=1}^{K}\sum_{i=1}^{I_k}\rho\mathrm{Tr}(\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H) + \sum_{k=1}^{K} M_k P_c + K P_o} \tag{4a}$$

$$\mathbf{s.t.}\ \log_2 \det\left(\mathbf{I}_{N_{i_k}} + \mathbf{H}_{i_k k}\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H\mathbf{H}_{i_k k}^H\mathbf{J}_{i_k}^{-1}\right) \geq r_{i_k},\ \forall i_k \in \mathcal{I} \tag{4b}$$

where r_{i_k} is the minimal data rate requirement of user i_k.

Remark 1: We do not impose transmit power constraints on the optimization problem. In this way, we can observe the impact of data rate requirement of each user on the achieved EE of the network in a wide range. In fact, with extra convex power constraints, the corresponding problem can be solved as well using an approach similar to what proposed as follows.

B. Problem Transformation

Optimization problem (4) is non-convex, because the objective function is non-concave over $\mathbf{V}$ and the constraints are not convex [23]. To find a solution of this problem, we construct a convex subset of the original set of non-convex constraint (4b) and a lower bound of the EE that is a quasi-concave function of $\mathbf{V}$.

Note that the non-convexity of the problem comes from the non-convexity of the data rate in (2), and the data rate is achievable by maximal likelihood detector at each user. Based on this observation, we find a lower bound of the data rate achieved by an optimal linear detector, which becomes a concave function of $\mathbf{V}$.

Define a function of precoding matrices as follows,

$$f_{i_k}(\mathbf{V}, \mathbf{U}_{i_k}, \mathbf{Q}_{i_k}) \triangleq \log_2 e \cdot (\ln\det(\mathbf{Q}_{i_k}) - \mathrm{Tr}(\mathbf{Q}_{i_k}\mathbf{E}_{i_k}) + d_{i_k}) \tag{5}$$

where $\mathbf{U}_{i_k}$ is a detection matrix at user i_k, $\mathbf{Q}_{i_k}$ is an auxiliary positive definite matrix, e is the base of natural logarithms, and $\mathbf{E}_{i_k}$ is the mean-square-error (MSE) covariance matrix, which is

$$\begin{aligned}\mathbf{E}_{i_k} = &(\mathbf{I}_{d_{i_k}} - \mathbf{U}_{i_k}^H\mathbf{H}_{i_k k}\mathbf{V}_{i_k})(\mathbf{I}_{d_{i_k}} - \mathbf{U}_{i_k}^H\mathbf{H}_{i_k k}\mathbf{V}_{i_k})^H \\ &+ \sum_{(l,j)\neq(i,k)} \mathbf{U}_{i_k}^H\mathbf{H}_{i_k j}\mathbf{V}_{l_j}\mathbf{V}_{l_j}^H\mathbf{H}_{i_k j}^H\mathbf{U}_{i_k} + \sigma_{i_k}^2\mathbf{U}_{i_k}^H\mathbf{U}_{i_k}\end{aligned} \tag{6}$$

Theorem 1: $f_{i_k}(\mathbf{V}, \mathbf{U}_{i_k}, \mathbf{Q}_{i_k})$ is a lower bound of the data rate R_{i_k} in (2) and is concave over each group of the matrices $\mathbf{V}, \mathbf{U}_{i_k}, \mathbf{Q}_{i_k}$ when the other two are fixed.

Proof: See Appendix A. ■

With the help of Theorem 1, we can construct a convex subset of constraint (4b) for any given groups of matrices $\mathbf{U}_{i_k}$ and $\mathbf{Q}_{i_k}$ as follows,

$$f_{i_k}(\mathbf{V}, \mathbf{U}_{i_k}, \mathbf{Q}_{i_k}) \geq r_{i_k},\ \forall i_k \in \mathcal{I} \tag{7}$$

With Theorem 1 we can also find a lower bound of the objective function in (4a) for any given groups of matrices $\mathbf{U}_{i_k}$ and $\mathbf{Q}_{i_k}$ as follows,

$$L(\mathbf{V}) = \frac{\sum_{k=1}^{K}\sum_{i=1}^{I_k} f_{i_k}(\mathbf{V}, \mathbf{U}_{i_k}, \mathbf{Q}_{i_k})}{\sum_{k=1}^{K}\sum_{i=1}^{I_k}\rho\mathrm{Tr}(\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H) + \sum_{k=1}^{K} M_k P_c + K P_o} \tag{8}$$

which is tight when $\mathbf{U}_{i_k} = \mathbf{U}_{i_k}^*$ and $\mathbf{Q}_{i_k} = \mathbf{Q}_{i_k}^*$ as shown in Appendix A, where with $\mathbf{U}_{i_k}^*$ and $\mathbf{Q}_{i_k}^*$ the function in (5) is maximized.

The denominator in (8) is convex quadratic and the numerator is concave quadratic over $\mathbf{V}$. Therefore, all the superlevel sets of (8)

$$\Theta_\alpha = \{\mathbf{V} \in \mathbf{dom}\ L | L(\mathbf{V}) \geq \alpha\} \tag{9}$$

are convex quadratic for any real number α. That is to say, $L(\mathbf{V})$ is quasi-concave over $\mathbf{V}$ with fixed $\mathbf{U}$ and $\mathbf{Q}$ [23], where $\mathbf{U}$ is short for $\{\mathbf{U}_{i_k}\}_{i_k\in\mathcal{I}}$ and $\mathbf{Q}$ is short for $\{\mathbf{Q}_{i_k}\}_{i_k\in\mathcal{I}}$.

Then, a new optimization problem that maximizes the EE lower bound within a convex subset of constraint (4b) can be formulated as follows,

$$\max_{\mathbf{V},\mathbf{U},\mathbf{Q}} \frac{\sum_{k=1}^{K}\sum_{i=1}^{I_k} f_{i_k}(\mathbf{V}, \mathbf{U}_{i_k}, \mathbf{Q}_{i_k})}{\sum_{k=1}^{K}\sum_{i=1}^{I_k}\rho\mathrm{Tr}(\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H) + \sum_{k=1}^{K} M_k P_c + K P_o} \tag{10a}$$

$$\mathbf{s.t.}\quad f_{i_k}(\mathbf{V}, \mathbf{U}_{i_k}, \mathbf{Q}_{i_k}) \geq r_{i_k},\ \forall i_k \in \mathcal{I} \tag{10b}$$

from which we can at least find a local optimum of the original problem (4) as shown in the following subsection.

C. Iterative Algorithm

Since the objective function in (10a) is concave over each group of the matrices $\mathbf{U}$ and $\mathbf{Q}$, and quasi-concave over $\mathbf{V}$ when the other two are fixed, we use the block coordinate descent method [24] to solve problem (10), i.e., fix two of the three matrices to update the third. The matrices $\mathbf{U}$ and $\mathbf{Q}$ can be updated with (A.2) and (A.4) in Appendix A, respectively. Since the objective function is quasi-concave over $\mathbf{V}$ with fixed $\mathbf{U}$ and $\mathbf{Q}$, the update of $\mathbf{V}$ can be obtained by a bisection searching algorithm [23].

Specifically, with fixed $\mathbf{U}$ and $\mathbf{Q}$ the optimal precoding matrices can be found from checking the following feasibility problem,

$$\begin{aligned}&\text{Find}\quad \mathbf{V} \\ &\mathbf{s.t.}\quad \sum_{k=1}^{K}\sum_{i=1}^{I_k} f_{i_k}(\mathbf{V}, \mathbf{U}_{i_k}, \mathbf{Q}_{i_k}) \\ &\qquad - \alpha\left(\sum_{k=1}^{K}\sum_{i=1}^{I_k}\rho\mathrm{Tr}(\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H) + \sum_{k=1}^{K} M_k P_c + K P_o\right) \geq 0\end{aligned} \tag{11a}$$

$$f_{i_k}(\mathbf{V}, \mathbf{U}_{i_k}, \mathbf{Q}_{i_k}) \geq r_{i_k},\ \forall i_k \in \mathcal{I} \tag{11b}$$

where (11a) is one of the superlevel sets Θ_α in (9), which is a convex inequality when α is a positive real number.

This is a convex feasibility problem since all the inequalities in (11a) and (11b) are convex. Let $L(\mathbf{V}^*)$ denote the optimal value of the objective function of optimization problem (10) with fixed $\mathbf{U}$ and $\mathbf{Q}$. By checking the feasibility of problem (11), we can identify whether $L(\mathbf{V}^*)$ is larger or less than a given value α. If problem (11) is feasible, i.e., $L(\mathbf{V}^*) \geq \alpha$, in the next iteration the value of α should be increased. Otherwise, $L(\mathbf{V}^*) \leq \alpha$, in the next iteration the value of α should be reduced. The searching procedure will be terminated when the increment of α is small enough.

The iterative algorithm is summarized in Table I, which converges to a stationary point of the original problem (4), as proved in Appendix B.

TABLE I
THE ITERATIVE ALGORITHM

Initialization of $\mathbf{V}$.
repeat
1. Update $\mathbf{U}$ with (A.2);
2. Update $\mathbf{Q}$ with (A.4);
3. Update $\mathbf{V}$ with the bisection algorithm:
   given $\theta \leq L(\mathbf{V}^*)$, $\tau \geq L(\mathbf{V}^*)$, tolerance $\varepsilon > 0$.
   repeat
   (1). $\alpha := (\theta + \tau)/2$;
   (2). Solve the convex feasibility problem (11);
   (3). **if** (11) is feasible, $\theta := \alpha$;   **else** $\tau := \alpha$.
   until $\tau - \theta \leq \varepsilon$
until convergence

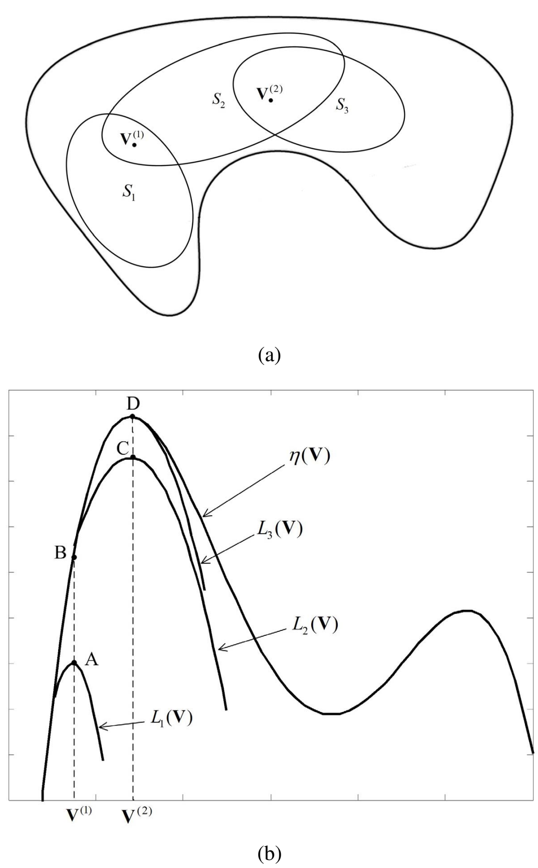


Fig. 1. The change of (a) convex subset and (b) quasi-concave lower bound with iterations.

To better understand the proposed iterative algorithm, we illustrate its procedure in Fig. 1, which shows the change of the convex subset and the quasi-concave EE lower bound with iterations. Denote $\mathbf{U}^{(1)}$ and $\mathbf{Q}^{(1)}$ as the matrices in the first iteration. By maximizing the lower bound $L_1(\mathbf{V})$ within the subset $S_1 = \{\mathbf{V} | f_{i_k}(\mathbf{V}, \mathbf{U}_{i_k}^{(1)}, \mathbf{Q}_{i_k}^{(1)}) \geq r_{i_k}, \ i_k \in \mathcal{I}\}$, point A$(\mathbf{V}^{(1)}, \mathbf{U}^{(1)}, \mathbf{Q}^{(1)})$ is obtained. In the next iteration, updating $\mathbf{U}$ and $\mathbf{Q}$, we obtain point B$(\mathbf{V}^{(1)}, \mathbf{U}^{(2)}, \mathbf{Q}^{(2)})$ on the curve of the original non-concave function $\eta(\mathbf{V})$. Then, $L_2(\mathbf{V})$, the lower bound through point B within the subset $S_2 = \{\mathbf{V} | f_{i_k}(\mathbf{V}, \mathbf{U}_{i_k}^{(2)}, \mathbf{Q}_{i_k}^{(2)}) \geq r_{i_k}, \ i_k \in \mathcal{I}\}$ is maximized, and we obtain point C$(\mathbf{V}^{(2)}, \mathbf{U}^{(2)}, \mathbf{Q}^{(2)})$. Since $\mathbf{U}^{(2)}$ and $\mathbf{Q}^{(2)}$ are given by (A.2) and (A.4) when $\mathbf{V}^{(1)}$ is used, we have $f_{i_k}(\mathbf{V}^{(1)}, \mathbf{U}_{i_k}^{(2)}, \mathbf{Q}_{i_k}^{(2)}) \geq f_{i_k}(\mathbf{V}^{(1)}, \mathbf{U}_{i_k}^{(1)}, \mathbf{Q}_{i_k}^{(1)}) \geq r_{i_k}$. Therefore, $\mathbf{V}^{(1)}$ must be located in the intersection set between S_1 and S_2, and the solution $\mathbf{V}^{(2)}$ in S_2 provides higher EE than $\mathbf{V}^{(1)}$. Similarly, $\mathbf{V}^{(2)}$ must be located in the intersection set between S_2 and S_3, and $\mathbf{V}^{(3)}$ provides higher EE than $\mathbf{V}^{(2)}$. We can see that the subsets cover more and more regions of the original set of constraint (4b) and the gap between the lower bound of the EE and the original objective function diminishes with iterations. If $\mathbf{V}^{(2)}$ is a stationary point of $\eta(\mathbf{V})$ as shown in Fig. 1(b), by maximizing the lower bound $L_3(\mathbf{V})$ within the subset S_3, the iterative algorithm will converge to the point D$(\mathbf{V}^{(2)}, \mathbf{U}^{(3)}, \mathbf{Q}^{(3)})$.

Remark 2: In [8] a weighted minimum MSE (WMMSE) algorithm was proposed. Although the idea of constructing a concave lower bound of data rate expression is similar to that deriving the WMMSE, the major difference of the proposed algorithm from the WMMSE lies in the constraints. In our problem, the QoS constraints are non-convex, hence we construct a convex subset that varies in every iteration, while in the WMMSE the power constraints are convex and remain unchanged with iterations. Besides, our objective function is EE that is quasi-concave even after transformation, while in [8] the objective function is SE that is concave after transformation.

D. Initialization and Feasibility

Because the objective function of problem (4) is non-concave, there might be several local optima in the feasible region. This indicates that the solution of the proposed algorithm depends on the initial value. A usual way to increase the possibility of finding the globally optimal solution is to iterate from many initial values and then choose the best solution, which however leads to heavy computational burden.

Besides, the QoS constraint (4b) is not always feasible due to the interference especially when the QoS targets are stringent. When every user only has a single antenna, a necessary condition for the feasibility of QoS constraints was presented in [25], which however is not applicable for our network where the users have multiple antennas.

In this paper, we employ an alternative way to initialize the algorithm. We use a feasible precoder that is EE-optimal in special case as the initial value. When proving the convergency of the proposed algorithm, we have shown that if the initial value of the precoder ensures the feasibility of the QoS constraints, the algorithm can always yield a feasible solution.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2014.2343255, IEEE Transactions on Vehicular Technology

When each user is equipped with more than one antenna, IA can achieve higher degrees of freedom (DoF) [26], [27]. By judiciously designing linear beamformer and detector for each BS and each user in the network, interference can be thoroughly eliminated by linear IA [26], [28], which provides high SE in interference-limited environments. The analysis of linear IA feasibility for MIMO interference channel is an important issue, which has drawn significant attention recently [27], [29]–[31]. For the given system setting, i.e., the number of cells, the number of users in each cell, and the numbers of antennas at each BS and each user, the maximal number of data streams can be derived from the necessary and sufficient conditions for the linear IA feasibility, i.e., the conditions for interference-free transmission. With the help of the feasibility analysis in the literature, the following theorem shows that linear IA is a feasible solution to optimization problem (4).[2]

Theorem 2: Given the data rate requirement r_{i_k} of user i_k, problem (4) is feasible *if* linear IA is employed, where the maximal number of data streams $d^*_{i_k}$ for user i_k supported by IA can be found from the necessary and sufficient conditions for the IA feasibility [27].

Proof: See Appendix C. ∎

In fact, IA is able to maximize EE when the ICI is strong, as indicated in the following theorem.

Theorem 3: Linear IA is approximately EE-optimal among all the linear beamforming matrices when the users are located at "exact cell-edge",[3] and the data rate requirement of each user is high.

Proof: See Appendix D. ∎

The two theorems suggest that we can use an energy-efficient IA precoder as the initial value, which is a linear IA beamformer combined with the optimal transmit power maximizing the network EE under the QoS constraints.

The IA beamformer can be obtained from [26] and [28]. From the data rate achieved by linear IA shown in (C.1), the optimal transmit power can be found from the following problem

$$\max_{P_{i_kj}} \frac{\sum_{k=1}^{K}\sum_{i=1}^{I_k}\sum_{j=1}^{d^*_{i_k}} \log_2\left(1+\frac{P_{i_kj}\lambda_{i_kj}}{\sigma^2_{i_k}}\right)}{\sum_{k=1}^{K}\sum_{i=1}^{I_k}\sum_{j=1}^{d^*_{i_k}} \rho P_{i_kj} + \sum_{k=1}^{K} M_k P_c + K P_o} \tag{12a}$$

$$\textbf{s.t.}\ \log_2\left(1+\frac{P_{i_kj}\lambda_{i_kj}}{\sigma^2_{i_k}}\right) \geq \frac{r_{i_k}}{d^*_{i_k}},\quad \forall j=1,...,d^*_{i_k},\quad \forall i_k \in \mathcal{I} \tag{12b}$$

where P_{i_kj} is the transmit power for the j-th data stream of user i_k, and λ_{i_kj} is define in (C.1). For simplicity, different data streams of each user are set with equal data rate constraints so that the constraints are linear. This does not affect the solution to the original problem, since the optimized transmit power is used to initialize the precoding matrices for the iterative algorithm.

Since the objective function is pseudo-concave [32], the optimal transmit power can be found from Karush-Kuhn-Tucker conditions, which can be expressed as follows,

$$-\frac{1}{P_s^2}\left(\frac{g_{i_kj}P_s}{(1+g_{i_kj}P_{i_kj})\ln 2} - \rho R_s\right) - \mu_{i_kj}g_{i_kj} = 0 \tag{13a}$$

$$2^{\frac{r_{i_k}}{d^*_{i_k}}} - 1 - P_{i_kj}g_{i_kj} \leq 0 \tag{13b}$$

$$\mu_{i_kj}\left(2^{\frac{r_{i_k}}{d^*_{i_k}}} - 1 - P_{i_kj}g_{i_kj}\right) = 0 \tag{13c}$$

$$\forall j=1,...,d^*_{i_k},\quad \forall i_k \in \mathcal{I}$$

where

$$g_{i_kj} = \frac{\lambda_{i_kj}}{\sigma^2_{i_k}}$$

$$P_s = \sum_{k=1}^{K}\sum_{i=1}^{I_k}\sum_{j=1}^{d^*_{i_k}} \rho P_{i_kj} + \sum_{k=1}^{K} M_k P_c + K P_o$$

$$R_s = \sum_{k=1}^{K}\sum_{i=1}^{I_k}\sum_{j=1}^{d^*_{i_k}} \log_2\left(1+\frac{P_{i_kj}\lambda_{i_kj}}{\sigma^2_{i_k}}\right)$$

and $\mu_{i_kj}, j \in \{1,\ldots,d^*_{i_k}\},\ i_k \in \mathcal{I}$ are the non-negative Lagrange multipliers.

When $\mu_{i_kj}=0$, we have $P_{i_kj} = \left(\frac{P_s g_{i_kj}}{\rho R_s \ln 2} - 1\right)\frac{1}{g_{i_kj}}$ from (13a). When $\mu_{i_kj} > 0$, we have $P_{i_kj} = \left(2^{\frac{r_{i_k}}{d^*_{i_k}}} - 1\right)\frac{1}{g_{i_kj}}$ from (13c). Since $P_{i_kj} \geq \left(2^{\frac{r_{i_k}}{d^*_{i_k}}} - 1\right)\frac{1}{g_{i_kj}}$ from (13b), the optimal transmit power can be obtained as

$$P^*_{i_kj} = \left(\max\left\{2^{\frac{r_{i_k}}{d^*_{i_k}}},\ \frac{P_s g_{i_kj}}{\rho R_s \ln 2}\right\} - 1\right)\frac{1}{g_{i_kj}} \tag{14}$$

Then, the energy-efficient IA precoder for user i_k is

$$\mathbf{V}^{\mathrm{IA}}_{i_k} = \mathbf{W}_{i_k} \cdot diag\left\{\sqrt{P^*_{i_k1}},\cdots,\sqrt{P^*_{i_kd^*_{i_k}}}\right\} \tag{15}$$

where $\mathbf{W}_{i_k}$ is the $M_k \times d^*_{i_k}$ linear IA beamforming matrix, whose explicit and iterative solutions can be found from [26] and [28], respectively.

Theorem 2 suggests that using $\mathbf{W}_{i_k}$ is only a sufficient condition for the feasibility of the QoS constraints, where the maximal number of data streams is restricted to $d^*_{i_k}$ in order to ensure interference-free transmission. Moreover, Theorem 3 implies that it is EE-optimal only when the interference is strong. In practice, interference is not always strong. This suggests that some antennas can be exploited to transmit more data streams even when these data streams are subject to weak interference. In other words, the number of data streams d_{i_k} transmitted to user i_k can exceed $d^*_{i_k}$. To achieve a trade-off between using the antennas for signal transmission and interference mitigation, we initialize the precoding matrix with

[2] In practice, if the QoS requirements of some users are too strict to be supported by the coordinated BSs even with the maximal transmit power, these users can access to different frequency or time resources.

[3] For the users located at the exact cell-edge, the average gains of the channels from all coordinated BSs to a user are equal. This is only a mathematical definition, while such users may rarely appear in practical systems with more than three coordinative BSs.

$d_{i_k} = \min(M_k, N_{i_k})$ columns vectors, which is the maximal number of data streams being able to be transmitted without interference. Then, the $M_k \times d_{i_k}$ initial precoding matrix can be written as

$$\mathbf{V}_{i_k}^{\text{Ini}} = \left[\mathbf{V}_{i_k}^{\text{IA}}, \mathbf{V}_{i_k}^{+}\right], \forall i_k \in \mathcal{I} \tag{16}$$

where $\mathbf{V}_{i_k}^{+}$ is an $M_k \times (d_{i_k} - d_{i_k}^{*})$ matrix tending to be a zero matrix so that the data rate constraints are still satisfied.

In particular, we set $\mathbf{V}_{i_k}^{+}$ to be $\epsilon \cdot \bar{\mathbf{W}}_{i_k}$, where $\bar{\mathbf{W}}_{i_k}$ is an $M_k \times (d_{i_k} - d_{i_k}^{*})$ matrix whose columns are the orthonormal bases of the null space of $\mathbf{W}_{i_k}$, and ϵ is an infinitesimal number that tends to 0.

After iterations with the initial value $\mathbf{V}_{i_k}^{\text{Ini}}$, $\mathbf{V}_{i_k}$ will change adaptively according to the interference strength. If the interference is strong, the number of data streams with non-trivial power will still tend to be $d_{i_k}^{*}$ ($\mathbf{V}_{i_k}^{+}$ still tends to be a zero matrix). Otherwise, the number of data streams with non-trivial power will correspondingly increase.

IV. Simulation Results

In this section, we evaluate the performance of the proposed energy-efficient coordinated precoder.

The simulation set-up is as follows. We consider a coordinated cluster of three cells, and the cell radius is 500 m. In order to observe the impact of data rate requirement of each user on the achieved EE in a wide range, in the simulation each BS serves one user in the same time-frequency resource, although the proposed algorithm is still applicable when each BS serves more users. Each user has the same data rate requirement, i.e., $r_{i_k} = r$, $\forall i_k \in \mathcal{I}$. The numbers of antennas of each BS and each user are set to be equal, which are $M_k = N_{i_k} = 2$, $\forall i_k \in \mathcal{I}$. The path loss follows 3GPP channel model [33], and the small-scale channel is subject to Rayleigh fading. We set the noise variance $\sigma_{i_k}^2$ as -76 dBm, where the inter-cluster interference is taken into account. The circuit power consumption for each antenna P_c is set as 17.6 W, the constant power consumption P_o for each BS (with only one sector) is set as 43.3 W, and the power amplifier efficiency $1/\rho$ = 0.311. These power assumption parameters come from [34] for a macro BS. All the simulation results are obtained by averaging over 100 channel realizations.

Unless otherwise specified, this set-up will be used for the following simulations.

A. Convergence Rate of the Iterative Algorithm

In Fig. 2, we evaluate the convergence of the proposed iterative algorithm under a single channel realization. To show the impact of data rate requirements on the convergence rate, the distance between each BS and its serving user, d, is fixed to be 300 m. Note that the impact of increasing the distance is similar to that of requiring higher data rate, because both result in a higher transmit power. We can see that the algorithm converges faster with lower data rate requirements.

In the following, all results are obtained after the iterative algorithm converges.

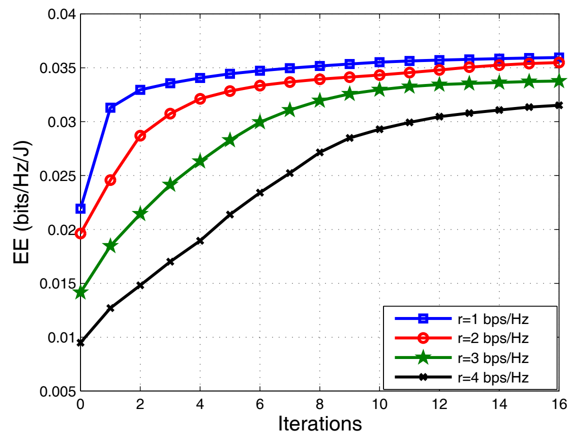


Fig. 2. Convergence of the algorithm, d = 300 m.

B. Impact of User Location

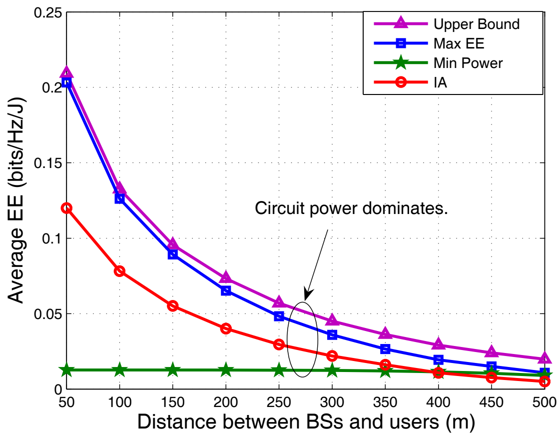
We first show the impact of the user location on average EE achieved by the proposed algorithm (with legend "Max EE"). For comparison, we provide the average EE of a scheme that minimizes the overall transmit power under the QoS constraint (4b) (with legend "Min Power"). To observe the contributions of the beamformer and transmit power separately, we show the EE of the energy-efficient IA precoder $\mathbf{V}_{i_k}^{\text{IA}}$ in (15) (with legend "IA").

To evaluate the performance of different precoders, the maximal likelihood detector is employed at each user for different schemes, therefore we use the corresponding achievable rate for the precoders in all the following simulations.

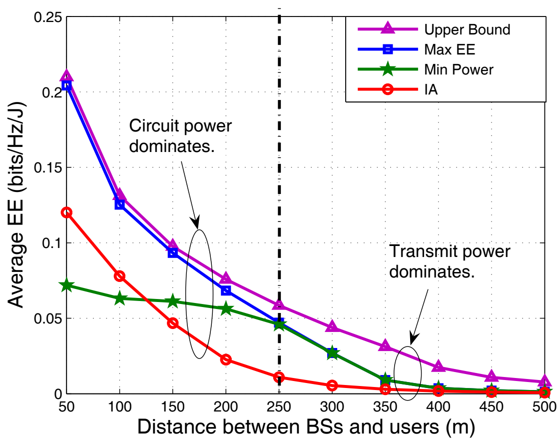
Since the globally optimal solution of the original problem (4) is hard to obtain, we simulate an ideal case with no interference among the links (with legend "Upper Bound"). Maximal ratio transmission and maximal ratio combining are applied at the transceivers, and the transmit power of each data stream is optimized to maximize the EE with the same QoS constraints. This can serve as a performance upper bound of the EE. Note that only in the simulation to show this "Upper Bound", the interference is not taken into account.

Fig. 3 shows that the EE of the proposed algorithm is close to the "Upper Bound" and is much higher than the "Min Power" precoder, especially for cell-center users where the circuit power consumption dominates. This is because the proposed precoder transmits with higher data rate than the required value r, but the "Min Power" precoder always transmits with r in order to minimize the transmit power. The IA precoder is inferior to the proposed precoder because it only optimizes the transmit power while the proposed precoder jointly optimizes the beamformer and transmit power. Moreover, the number of data streams of IA precoder is fixed as $d_{i_k}^{*}$. By contrast, the number of data streams of our precoder is adaptive to the interference strength. When the users are located in cell-edge, the proposed precoder performs similarly to the "Min Power" precoder because the transmit power consumption becomes dominant. In this case the proposed precoder performs closely to the IA precoder because all the interferences are strong. This validates Theorem 3 even with data rate requirement of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2014.2343255, IEEE Transactions on Vehicular Technology



(a)



(b)

Fig. 3. Average EE versus d: (a) $r = 1$ bps/Hz; (b) $r = 5$ bps/Hz.

5 bps/Hz, which is not very high.

C. Impact of Data Rate Requirement

In the sequel, we analyze the impact of the QoS requirement on the average EE of the proposed precoder. Again, we compare with the "Min Power" precoder. In order to understand where the EE gain of the proposed precoder comes from, we also show the corresponding sum rate and total transmit power. The value of $r = 0$ bps/Hz corresponds to the best effort traffic.

Fig. 4 shows that the proposed precoder achieves much higher EE than the "Min Power" precoder in a wide region of the data rate requirement especially when the user is closer to its master BS. When the value of r is small, the circuit power consumption becomes dominant. In this scenario, the proposed precoder transmits with higher data rate than the required value of r (see Fig. 5(a)), which yields higher EE, but the "Min Power" precoder always transmits with the minimal data rate requirement no matter how far the users are from the BSs. When the value of r grows, the two precoders achieve the same EE because the transmit power becomes dominant so that maximizing the EE is equivalent to minimizing the transmit power (see Fig. 5(b)).

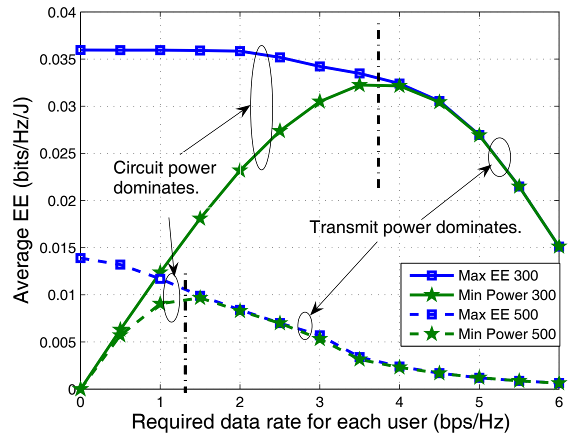


Fig. 4. Average EE versus r, d = 300 m or 500 m.

In high throughput cellular networks, SE is another important factor in the system design [5], [35]. By comparing Fig. 4 with Fig. 5(a), we can see that the proposed precoder achieves both higher EE and higher SE (i.e., sum rate) than the "Min Power" precoder when the required data rate of each user is low, and in a wide range the EE and the SE supported by the proposed precoder change little with the required data rate. The two precoders perform the same in both SE and EE when the required data rate of each user is high.

D. Impact of Antenna Number

Since increasing the number of antennas will consume more circuit power but save more transmit power at the same time, its impact on EE is unclear. Fig. 6 shows the average EE of the proposed precoder under different numbers of antennas (with legend "Max EE (M_k, N_{i_k})"). As a comparison, the performance of the "Min Power" precoder is shown as well. Considering that the spatial multiplexing gain is limited to $\min(M_k, N_{i_k})$, we also show the configurations of $M_k = N_{i_k} = 8$, where each user consumes more circuit power that is not counted into the EE of the downlink network.

As shown in Fig. 6, "Max EE (2,2)" is superior to "Max EE (8,2)" for cell-center users but is inferior to "Max EE (8,2)" for cell-edge users. This is because when the users are located in cell-center where the circuit power consumption dominates, "Max EE (8,2)" consumes more circuit power. When the users are in cell-edge where the transmit power consumption dominates, though "Max EE (8,2)" increases the circuit power, it saves transmit power yielding a higher EE. The "Max EE (8,8)" system always performs the best, because when the circuit power consumption dominates it can transmit with higher data rate and when the transmit power consumption is dominant it can save the transmit power.

By contrast, the "Min Power" precoder behaves quite differently. When the circuit power consumption dominates, "Min Power (2,2)" is the best because the increased number of antennas only increases the circuit power but does not provide higher data rate. When the transmit power consumption is dominant, the two precoders perform similarly.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2014.2343255, IEEE Transactions on Vehicular Technology

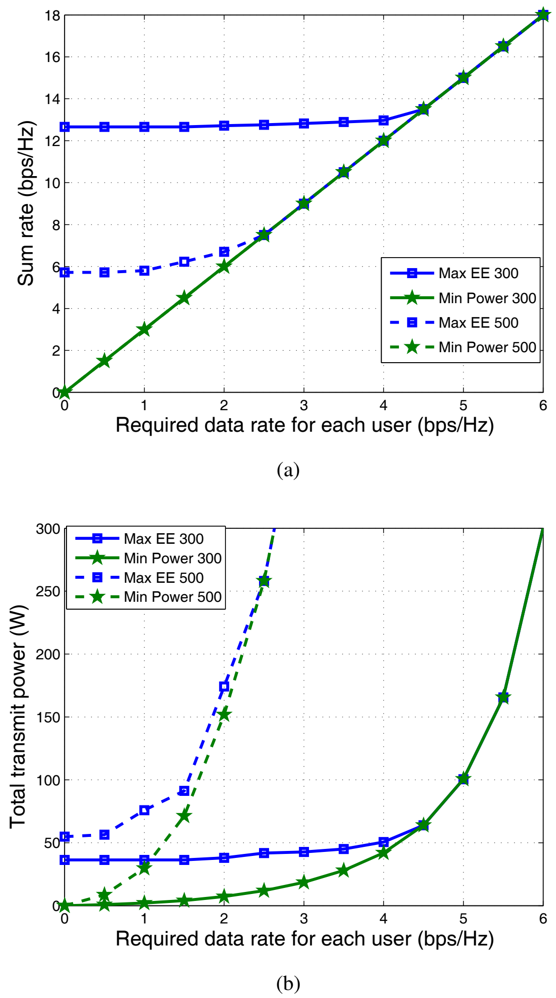


Fig. 5. (a) Sum rate and (b) total transmit power, d = 300 m or 500 m.

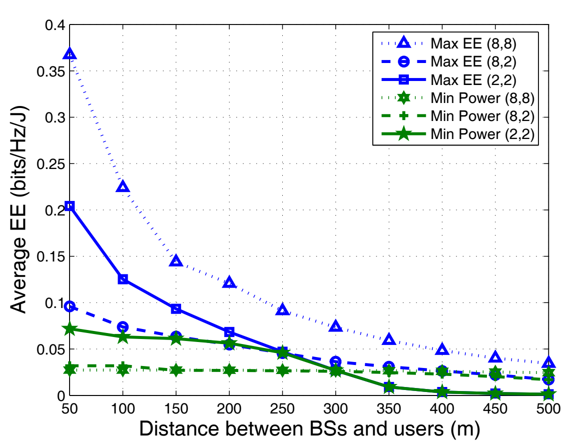


Fig. 6. Average EE under different antenna configurations, $r = 5$ bps/Hz.

V. Conclusion

We studied energy-efficient precoder that maximizes the EE of multi-cell multi-antenna networks with coordinated beamforming under the minimal data rate constraint imposed by each user, which reflects the QoS provision for best effort, and real and non-real time services. To find a solution of the non-convex optimization problem, we constructed a convex subset of the original constraint set and a quasi-concave lower bound of the EE. An iterative algorithm was proposed to maximize the lower bound of the EE within the convex subset, which was proved to converge to at least a local optimal solution of the original problem, and the initial value of the algorithm was carefully designed. Simulation results showed that the proposed precoder performs closely to an upper bound derived from interference-free assumption, and achieves much higher EE than the precoder that minimizes the transmit power and the energy-efficient IA precoder when the circuit power consumption is dominant.

Appendix A
Proof of Theorem 1

First, we show that $f_{i_k}(\mathbf{V}, \mathbf{U}_{i_k}, \mathbf{Q}_{i_k})$ is concave over $\mathbf{V}, \mathbf{U}_{i_k}, \mathbf{Q}_{i_k}$, respectively. By substituting (6) into (5), we have

$$\begin{aligned}&\ln 2 \cdot f_{i_k}(\mathbf{V}, \mathbf{U}_{i_k}, \mathbf{Q}_{i_k})\\&= \ln\det(\mathbf{Q}_{i_k}) + d_{i_k} - \mathrm{Tr}(\sigma_{i_k}^2 \mathbf{Q}_{i_k} \mathbf{U}_{i_k}^H \mathbf{U}_{i_k})\\&\quad - \mathrm{Tr}(\mathbf{Q}_{i_k}(\mathbf{I}_{d_{i_k}} - \mathbf{U}_{i_k}^H \mathbf{H}_{i_k k} \mathbf{V}_{i_k})(\mathbf{I}_{d_{i_k}} - \mathbf{U}_{i_k}^H \mathbf{H}_{i_k k} \mathbf{V}_{i_k})^H)\\&\quad - \sum_{(l,j)\neq(i,k)} \mathrm{Tr}(\mathbf{Q}_{i_k} \mathbf{U}_{i_k}^H \mathbf{H}_{i_k j} \mathbf{V}_{l_j} \mathbf{V}_{l_j}^H \mathbf{H}_{i_k j}^H \mathbf{U}_{i_k}) \qquad \text{(A.1)}\end{aligned}$$

From the last two terms in (A.1), which are traces of quadratic terms of $\mathbf{V}$, we can see that $f_{i_k}(\mathbf{V}, \mathbf{U}_{i_k}, \mathbf{Q}_{i_k})$ is concave quadratic over $\mathbf{V}$ when $\mathbf{U}_{i_k}$ and $\mathbf{Q}_{i_k}$ are fixed. Similarly, with fixed $\mathbf{V}$ and $\mathbf{Q}_{i_k}$, it is concave quadratic over $\mathbf{U}_{i_k}$. With fixed $\mathbf{V}$ and $\mathbf{U}_{i_k}$, the MSE covariance matrix $\mathbf{E}_{i_k}$ is fixed, and $f_{i_k}(\mathbf{V}, \mathbf{U}_{i_k}, \mathbf{Q}_{i_k}) = \log_2 e \cdot (\ln\det(\mathbf{Q}_{i_k}) - \mathrm{Tr}(\mathbf{Q}_{i_k}\mathbf{E}_{i_k}) + d_{i_k})$ is concave over $\mathbf{Q}_{i_k}$ due to the concavity of $\ln\det(\mathbf{Q}_{i_k})$ and the linearity of $\mathrm{Tr}(\mathbf{Q}_{i_k}\mathbf{E}_{i_k})$.

Then, we prove that $f_{i_k}(\mathbf{V}, \mathbf{U}_{i_k}, \mathbf{Q}_{i_k})$ is a lower bound of R_{i_k} by showing that R_{i_k} is the maximum of $f_{i_k}(\mathbf{V}, \mathbf{U}_{i_k}, \mathbf{Q}_{i_k})$.

Given the precoding matrices $\mathbf{V}$, the trace of $\mathbf{E}_{i_k}$ is minimized with an MMSE detector

$$\mathbf{U}_{i_k}^* = \left(\sum_{j=1}^{K} \sum_{l=1}^{I_l} \mathbf{H}_{i_k j} \mathbf{V}_{l_j} \mathbf{V}_{l_j}^H \mathbf{H}_{i_k j}^H + \sigma_{i_k}^2 \mathbf{I}_{N_{i_k}} \right)^{-1} \mathbf{H}_{i_k k} \mathbf{V}_{i_k} \qquad \text{(A.2)}$$

By substituting $\mathbf{U}_{i_k}^*$ into (5), we obtain

$$f_{i_k}(\mathbf{V}, \mathbf{U}_{i_k}^*, \mathbf{Q}_{i_k}) = \log_2 e \cdot (\ln\det(\mathbf{Q}_{i_k}) - \mathrm{Tr}(\mathbf{Q}_{i_k}\mathbf{E}_{i_k}^*) + d_{i_k}) \qquad \text{(A.3)}$$

where $\mathbf{E}_{i_k}^*$ is the MSE covariance matrix when $\mathbf{U}_{i_k}^*$ is used.

By taking the gradient of $f_{i_k}(\mathbf{V}, \mathbf{U}_{i_k}^*, \mathbf{Q}_{i_k})$ over $\mathbf{Q}_{i_k}$ and setting the gradient as $\mathbf{0}$, we obtain the optimal auxiliary matrix that maximizes the function in (A.3), which is

$$\mathbf{Q}_{i_k}^* = (\mathbf{E}_{i_k}^*)^{-1} \qquad \text{(A.4)}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2014.2343255, IEEE Transactions on Vehicular Technology

With $\mathbf{U}_{i_k}^*$ and $\mathbf{Q}_{i_k}^*$, the maximum of the function in (5) is achieved, which is

$$f_{i_k}(\mathbf{V}, \mathbf{U}_{i_k}^*, \mathbf{Q}_{i_k}^*) = \log_2 \det\left((\mathbf{E}_{i_k}^*)^{-1}\right) \tag{A.5}$$

where $\mathbf{U}_{i_k}^*$ and $\mathbf{Q}_{i_k}^*$ are functions of $\mathbf{V}$ as shown in (A.2) and (A.4). Though when $\mathbf{U}_{i_k}$ and $\mathbf{Q}_{i_k}$ are fixed matrices, we have proved that $f_{i_k}(\mathbf{V}, \mathbf{U}_{i_k}, \mathbf{Q}_{i_k})$ is concave over $\mathbf{V}$, when $\mathbf{U}_{i_k}^*$ and $\mathbf{Q}_{i_k}^*$ are functions of $\mathbf{V}$, $f_{i_k}(\mathbf{V}, \mathbf{U}_{i_k}^*, \mathbf{Q}_{i_k}^*)$ is not guaranteed to be concave over $\mathbf{V}$.

According to the relation between the MSE and the achievable rate [8], the expression in (A.5) is equal to the data rate in (2). This suggests that R_{i_k} is the maximum value of $f_{i_k}(\mathbf{V}, \mathbf{U}_{i_k}, \mathbf{Q}_{i_k})$. In other words, $f_{i_k}(\mathbf{V}, \mathbf{U}_{i_k}, \mathbf{Q}_{i_k})$ is a lower bound of the data rate R_{i_k}.

Appendix B
Proof of Convergence

Define the objective function at iteration $n = 1, 2, ...$ as

$$\begin{aligned}&L(\mathbf{V}^{(n)}, \mathbf{U}^{(n)}, \mathbf{Q}^{(n)})\\ =&\frac{\sum_{k=1}^{K}\sum_{i=1}^{I_k} f_{i_k}(\mathbf{V}^{(n)}, \mathbf{U}_{i_k}^{(n)}, \mathbf{Q}_{i_k}^{(n)})}{\sum_{k=1}^{K}\sum_{i=1}^{I_k} \rho \mathrm{Tr}(\mathbf{V}_{i_k}^{(n)}\mathbf{V}_{i_k}^{(n)H}) + \sum_{k=1}^{K} M_k P_c + K P_o}\end{aligned} \tag{B.1}$$

Updating $\mathbf{U}^{(n)}$ with the MMSE detector in (A.2), we obtain $L(\mathbf{V}^{(n-1)}, \mathbf{U}^{(n)}, \mathbf{Q}^{(n-1)}) \geq L(\mathbf{V}^{(n-1)}, \mathbf{U}^{(n-1)}, \mathbf{Q}^{(n-1)})$, and updating $\mathbf{Q}^{(n)}$ with the optimal auxiliary matrix in (A.4), we obtain $L(\mathbf{V}^{(n-1)}, \mathbf{U}^{(n)}, \mathbf{Q}^{(n)}) \geq L(\mathbf{V}^{(n-1)}, \mathbf{U}^{(n)}, \mathbf{Q}^{(n-1)})$. Updating $\mathbf{V}^{(n)}$ using the bisection algorithm in Table I, we obtain $L(\mathbf{V}^{(n)}, \mathbf{U}^{(n)}, \mathbf{Q}^{(n)}) \geq L(\mathbf{V}^{(n-1)}, \mathbf{U}^{(n)}, \mathbf{Q}^{(n)})$ with the constraints $f_{i_k}(\mathbf{V}^{(n)}, \mathbf{U}_{i_k}^{(n)}, \mathbf{Q}_{i_k}^{(n)}) \geq r_{i_k}, \forall i_k \in \mathcal{I}$. That is to say, $L(\mathbf{V}^{(n)}, \mathbf{U}^{(n)}, \mathbf{Q}^{(n)})$ monotonically increases with n and the constraints of problem (10) are always satisfied. Since the EE is bounded, according to the monotone bounded theorem, $L(\mathbf{V}^{(n)}, \mathbf{U}^{(n)}, \mathbf{Q}^{(n)})$ will converge after a number of iterations. By setting the initial value of precoder as $\mathbf{V}_{i_k}^{\mathrm{Ini}}$ in (16), we can ensure that in all iterations of the algorithm we can obtain a solution satisfying the QoS constraints.

Since the block coordinate descent method [24] is applied to solve problem (10), the iterative algorithm can converge to the stationary point of problem (10) [36]. In the sequel, we show that if $(\mathbf{V}^*, \mathbf{U}^*, \mathbf{Q}^*)$ is the stationary point of problem (10), $\mathbf{V}^*$ will be a stationary point of problem (4).

If $(\mathbf{V}^*, \mathbf{U}^*, \mathbf{Q}^*)$ is a stationary point of problem (10), it should satisfy the following stationarity condition

$$\mathrm{Tr}(\nabla_{\mathbf{V}} L(\mathbf{V}^*, \mathbf{U}^*, \mathbf{Q}^*)^H(\mathbf{V} - \mathbf{V}^*)) \leq 0 \tag{B.2}$$

Denote the (p, q)th element of $\mathbf{V}_{l_j}$ as $v_{l_j pq}$. From (8) and (3), we have

$$\begin{aligned}&\frac{\partial L(\mathbf{V}^*, \mathbf{U}^*, \mathbf{Q}^*)}{\partial v_{l_j pq}}\\ =&\frac{1}{P(\mathbf{V}^*)}\sum_{k=1}^{K}\sum_{i=1}^{I_k}\frac{\partial f_{i_k}(\mathbf{V}^*, \mathbf{U}_{i_k}^*, \mathbf{Q}_{i_k}^*)}{\partial v_{l_j pq}}\\ &-\frac{1}{P(\mathbf{V}^*)^2}\frac{\partial P(\mathbf{V}^*)}{\partial v_{l_j pq}}\sum_{k=1}^{K}\sum_{i=1}^{I_k} f_{i_k}(\mathbf{V}^*, \mathbf{U}_{i_k}^*, \mathbf{Q}_{i_k}^*)\\ =&\frac{1}{P(\mathbf{V}^*)}\sum_{k=1}^{K}\sum_{i=1}^{I_k}\frac{\partial R_{i_k}(\mathbf{V}^*)}{\partial v_{l_j pq}}\\ &-\frac{1}{P(\mathbf{V}^*)^2}\frac{\partial P(\mathbf{V}^*)}{\partial v_{l_j pq}}\sum_{k=1}^{K}\sum_{i=1}^{I_k} R_{i_k}(\mathbf{V}^*)\\ =&\frac{\partial \eta(\mathbf{V}^*)}{\partial v_{l_j pq}}\end{aligned} \tag{B.3}$$

where $P(\mathbf{V}^*)$ is the dominator of (8) with precoder $\mathbf{V}^*$, the first and the last equalities are from the chain rule and the second equality comes from (A.5).

Further considering (B.2), we know that

$$\mathrm{Tr}(\nabla_{\mathbf{V}} \eta(\mathbf{V}^*)^H(\mathbf{V} - \mathbf{V}^*)) \leq 0$$

which means that $\mathbf{V}^*$ is a stationary point of the original problem (4).

Appendix C
Proof of Theorem 2

According to the IA feasibility analysis [27], each user can receive at most $d_{i_k}^*$ interference-free data streams with linear IA, where $d_{i_k}^*$ can be found from the necessary and sufficient conditions for the IA feasibility (Theorem 2 in [27]).

The achievable data rate of user i_k with linear IA can be expressed as

$$\begin{aligned}R_{i_k}^{\mathrm{IA}} &= \log_2 \det(\mathbf{I}_{d_{i_k}^*} + \frac{1}{\sigma_{i_k}^2}\mathbf{D}_{i_k}^H \mathbf{H}_{i_k k}\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H \mathbf{H}_{i_k k}^H \mathbf{D}_{i_k})\\ &= \sum_{j=1}^{d_{i_k}^*}\log_2(1 + \frac{P_{i_k j}\lambda_{i_k j}}{\sigma_{i_k}^2})\end{aligned} \tag{C.1}$$

where $\mathbf{V}_{i_k} = \mathbf{W}_{i_k} \cdot diag\left\{\sqrt{P_{i_k 1}}, \cdots, \sqrt{P_{i_k d_{i_k}^*}}\right\}$ includes the beamforming and transmit power, $\mathbf{W}_{i_k}$ is the beamforming matrix and $\mathbf{D}_{i_k}$ is the detection matrix for user i_k corresponding to linear IA, whose explicit expressions and iterative solutions can be respectively found from [26] and [28], $P_{i_k j}$ is the transmit power of the j-th data stream of user i_k, and $\lambda_{i_k j}$ is the j-th eigenvalue of the matrix $\mathbf{D}_{i_k}^H \mathbf{H}_{i_k k}\mathbf{W}_{i_k}\mathbf{W}_{i_k}^H \mathbf{H}_{i_k k}^H \mathbf{D}_{i_k}$.

It is shown from (C.1) that $R_{i_k}^{\mathrm{IA}}$ only depends on the user i_k's own precoder $\mathbf{V}_{i_k}$. This suggests that without transmit power constraints, $\{\mathbf{V}_{i_k} | R_{i_k}^{\mathrm{IA}} \geq r_{i_k}\}_{i_k \in \mathcal{I}}$ is always a feasible set. Since the data rate expression R_{i_k} in (4) is achieved by an optimal non-linear detector, maximal likelihood detector [22], we have

$$R_{i_k} \geq R_{i_k}^{\mathrm{IA}} \geq r_{i_k}, \quad \forall i_k \in \mathcal{I} \tag{C.2}$$

Thus, the optimization problem in (4) is feasible.

APPENDIX D
PROOF OF THEOREM 3

When the average channel gains of all the links between the coordinated BSs and each user are equal, linear IA can achieve the maximal DoF of the network. Then, the achievable rate of user i_k with linear IA can be expressed as [26]

$$R_{i_k} = d^*_{i_k}\log_2(P_t) + o(\log_2(P_t))$$

where $d^*_{i_k}$ is the maximal achievable DoF of user i_k and P_t is the total transmit power of the coordinated BSs.

When the data rate requirement of each user r_{i_k} is high, the required transmit power P_t will be high as well. In this case, the term $o(\log_2(P_t))$ becomes negligible compared to $\log_2(P_t)$, and the minimal transmit power required to satisfy the minimal data rate requirement of the user can be approximated as

$$P_t \approx 2^{\frac{r_{i_k}}{d^*_{i_k}}}$$

Since $d^*_{i_k}$ is the maximal DoF of each user, the linear IA can achieve the minimal total transmit power among all linear transceivers given the value of r_{i_k}.

In this case, the EE can be approximated as

$$\frac{d^*_{\Sigma}\log_2(P_t)}{\rho P_t + \sum_{k=1}^{K} M_k P_c + K P_o}$$

which is maximal because P_t is minimal when the data rate requirement is a given value, where d^*_{Σ} is the total DoF of the network.

That is to say, linear IA precoder is approximately EE-optimal among the linear transceivers when users are located in cell-edge and the data rate requirement of each user is high.

REFERENCES

[1] Y. Li, Y. Tian, and C. Yang, "Energy-efficient coordinated beamforming with individual data rate constraints," in *IEEE PIMRC*, 2013.

[2] G. Y. Li, Z. Xu, C. Xiong, C. Yang, S. Zhang, Y. Chen, and S. Xu, "Energy-efficient wireless communications: tutorial, survey, and open issues," *IEEE Wireless Commun. Mag.*, vol. 18, no. 6, pp. 28–35, 2011.

[3] I. Humar, X. Ge, L. Xiang, M. Jo, M. Chen, and J. Zhang, "Rethinking energy efficiency models of cellular networks with embodied energy," *IEEE Netw.*, vol. 25, no. 2, pp. 40–49, 2011.

[4] Z. Xu, C. Yang, G. Y. Li, S. Zhang, Y. Chen, and S. Xu, "Energy-efficient configuration of spatial and frequency resources in MIMO-OFDMA systems," *IEEE Trans. Commun.*, vol. 61, no. 2, pp. 564–575, 2013.

[5] X. Hong, Y. Jie, C.-X. Wang, J. Shi, and X. Ge, "Energy-spectral efficiency trade-off in virtual MIMO cellular systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 10, pp. 2128–2140, 2013.

[6] L. Venturino, N. Prasad, and X. Wang, "Coordinated linear beamforming in downlink multi-cell wireless networks," *IEEE Trans. Wireless Commun.*, vol. 9, no. 4, pp. 1451–1461, 2010.

[7] W. Yu, T. Kwon, and C. Shin, "Multicell coordination via joint scheduling, beamforming, and power spectrum adaptation," *IEEE Trans. Wireless Commun.*, vol. 12, no. 7, pp. 1–14, 2013.

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

[9] H. Dahrouj and W. Yu, "Coordinated beamforming for the multicell multi-antenna wireless system," *IEEE Trans. Wireless Commun.*, vol. 9, no. 5, pp. 1748–1759, 2010.

[10] A. Tolli, H. Pennanen, and P. Komulainen, "Decentralized minimum power multi-cell beamforming with limited backhaul signaling," *IEEE Trans. Wireless Commun.*, vol. 10, no. 2, pp. 570–580, 2011.

[11] R. Zakhour and S. V. Hanly, "Min-max power allocation in cellular networks with coordinated beamforming," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 287–302, 2013.

[12] A. Wiesel, Y. C. Eldar, and S. Shamai, "Linear precoding via conic optimization for fixed MIMO receivers," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 161–176, 2006.

[13] C.-F. Tsai, C.-J. Chang, F.-C. Ren, and C.-M. Yen, "Adaptive radio resource allocation for downlink OFDMA/SDMA systems with multimedia traffic," *IEEE Trans. Wireless Commun.*, vol. 7, no. 5, pp. 1734–1743, 2008.

[14] D. W. K. Ng, E. S. Lo, and R. Schober, "Energy-efficient resource allocation for secure OFDMA systems," *IEEE Trans. Veh. Technol.*, vol. 61, no. 6, pp. 2572–2585, 2012.

[15] L. Chen, Y. Yang, X. Chen, and G. Wei, "Energy-efficient link adaptation on Rayleigh fading channel for OSTBC MIMO system with imperfect CSIT," *IEEE Trans. Veh. Technol.*, vol. 62, no. 4, pp. 1577–1585, 2013.

[16] S. Akin and M. C. Gursoy, "On the throughput and energy efficiency of cognitive MIMO transmissions," *IEEE Trans. Veh. Technol.*, vol. 62, no. 7, pp. 3245–3260, 2013.

[17] J. Xu and L. Qiu, "Energy efficiency optimization for MIMO broadcast channels," *IEEE Trans. Wireless Commun.*, vol. 12, no. 2, pp. 690–701, 2013.

[18] L. Xiang, X. Ge, C.-X. Wang, F. Y. Li, and F. Reichert, "Energy efficiency evaluation of cellular networks based on spatial distributions of traffic load and power consumption," *IEEE Trans. Wireless Commun.*, vol. 12, no. 3, pp. 961–973, 2013.

[19] X. Ge, T. Han, Y. Zhang, G. Mao, C.-X. Wang, J. Zhang, B. Yang, and S. Pan, "Spectrum and energy efficiency evaluation of two-tier femtocell networks with partially open channels," *IEEE Trans. Veh. Technol.*, vol. 63, no. 3, pp. 1306–1319, 2014.

[20] G. Miao, N. Himayat, G. Y. Li, and S. Talwar, "Distributed interference-aware energy-efficient power optimization," *IEEE Trans. Wireless Commun.*, vol. 10, no. 4, pp. 1323–1333, 2011.

[21] C. Jiang and L. J. Cimini, "Energy-efficient transmission for MIMO interference channels," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2988–2999, 2013.

[22] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[23] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[24] D. Bertsekas, *Nonlinear Programming*. Belmont, MA: Athena Scientific, 1999.

[25] Z. Xiang, M. Tao, and X. Wang, "Coordinated multicast beamforming in multicell networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 1, pp. 12–21, 2013.

[26] V. R. Cadambe and S. A. Jafar, "Interference alignment and degrees of freedom of the K-user interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3425–3441, 2008.

[27] T. Liu and C. Yang, "On the feasibility of linear interference alignment for MIMO interference broadcast channels with constant coefficients," *IEEE Trans. Signal Process.*, vol. 61, no. 9, pp. 2178–2191, 2013.

[28] ——, "Interference alignment transceiver design for MIMO interference broadcast channels," in *IEEE WCNC*, 2012.

[29] C. M. Yetis, T. Gou, S. A. Jafar, and A. H. Kayran, "On feasibility of interference alignment in MIMO interference networks," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4771–4782, 2010.

[30] G. Bresler, D. Cartwright, and D. Tse, "Settling the feasibility of interference alignment for the MIMO interference channel: the symmetric square case," *arXiv:1104.0888v1 [cs.IT]*, 2011.

[31] M. Razaviyayn, G. Lyubeznik, and Z.-Q. Luo, "On the degrees of freedom achievable through interference alignment in a MIMO interference channel," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 812–821, 2012.

[32] M. Avriel, *Nonlinear Programming: Analysis and Methods*. Courier Dover Publications, 2003.

[33] 3GPP, "Further advancements for E-UTRA physical layer aspects," in *TR 36.814 V9.0.0*, 2010.

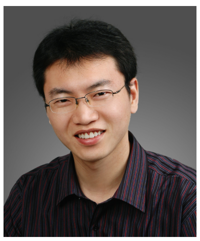
[34] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. A. Imran, D. Sabella, M. J. Gonzalez, O. Blume, and A. Fehske, "How much energy is needed to run a wireless network?" *IEEE Wireless Commun. Mag.*, vol. 18, no. 5, pp. 40–49, 2011.

[35] J. Jiang, M. Dianati, M. A. Imran, R. Tafazolli, and Y. Chen, "On the relation between energy efficiency and spectral efficiency of multiple-antenna systems," *IEEE Trans. Veh. Technol.*, vol. 62, no. 7, pp. 3463–3469, 2013.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2014.2343255, IEEE Transactions on Vehicular Technology

[36] C. A. Sagastizábal and M. V. Solodov, “Parallel variable distribution for constrained optimization,” *Computational Optimization and Applications*, vol. 22, no. 1, pp. 111–131, 2002.

Yang Li received the B.E degree in School of Electronics and Information Engineering from Beihang University (BUAA), Beijing, China in 2012. He is currently pursuing the M.E degree in School of Electronics and Information Engineering from BUAA. His research interests lie in cellular interference management and green radio.



Yafei Tian (S'05-M'08) received his B.S. degree in electronics engineering and Ph.D. degree in signal and information processing from Beihang University, Beijing, China, in 2001 and 2008, respectively. He is currently an Associate Professor with the School of Electronics and Information Engineering, Beihang University. He was a Visiting Scholar with the University of Southern California, Los Angeles, CA, USA, from 2010 to 2011. His research interests lie in heterogeneous cellular networks, energy efficient transmission and interference management.

Chenyang Yang (SM'08) received the M.S.E and Ph.D. degrees in electrical engineering from Beihang University (formerly Beijing University of Aeronautics and Astronautics), Beijing, China, in 1989 and 1997, respectively.

She is currently a Full Professor with the School of Electronics and Information Engineering, Beihang University. Her recent research interests include green radio, interference coordination and cooperative transmission for large scale networks.

Prof. Yang was the Chair of the IEEE Communications Society Beijing chapter from 2008 to 2012. She has served as Technical Program Committee Member for many IEEE conferences. She currently serves as an Associate editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, a guest editor of IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, an Associate Editor-in-Chief of the Chinese Journal of Communications, and an Associate Editor-in Chief of the Chinese Journal of Signal Processing. She was nominated as an Outstanding Young Professor of Beijing in 1995 and was supported by the First Teaching and Research Award Program for Outstanding Young Teachers of Higher Education Institutions by Ministry of Education (P.R.C. “TRAPOYT”) during 1999-2004.