Energy-Efficient Uplink Multi-User MIMO

Guowang Miao [♭]

Abstract—This paper addresses optimal energy-efficient design for uplink (UL) MU-MIMO in a single cell environment. The energy efficiency is measured by throughput per Joule, while both RF transmission power and device electronic circuit power are considered. We define the energy efficiency (EE) capacity for UL MU-MIMO and study the power allocation that achieves this capacity. First we assume all users consume a fixed amount of circuit power and show that user antennas should be used only when the corresponding spatial channels are sufficiently good and using them improves the overall network EE. Mobile devices may have improved circuit management capability and turn off circuit operations when some antennas are not used to reduce circuit power consumption. Therefore we further study energy-efficient UL MU-MIMO with improved circuit management and show that some antennas should not be used even when their channel states are good because turning them on consumes too much circuit power. Based on theoretical analysis, we further develop low-complexity yet globally optimal energy-efficient power allocation algorithms that converge to the optimum exponentially. Simulation results are provided to demonstrate the significant gain in network energy efficiency.

Index Terms– energy efficiency, multi-user MIMO, power allocation, SDMA

I. INTRODUCTION

Energy efficiency is increasingly important for mobile devices because battery technology has not kept up with the growing demand of ubiquitous multimedia communications [1], [2]. In addition to energy saving, energy-efficient communications have the benefit of reducing interference to other co-channel users as well as lessening environmental impacts, e.g., heat dissipation and electronic pollution. Therefore, recent research has focused on energy-efficient wireless communication techniques [3]–[9]. When the transmission bandwidth approaches infinity, the minimum received signal energy per bit for reliable communication over additive white Gaussian noise (AWGN) channels approaches -1.59 dB [3]. For band-limited transmission, the lowest order modulation should be used [4]. The investigation in [3], [4] does not account for additional circuit power that is consumed by electronic operations that can not be avoided. Considering the circuit power, the approaches in [3], [4] are no longer optimal. Energy dissipation of both transmitter circuits and radio-frequency (RF) output is investigated in [10], where the modulation level is adapted to minimize the energy consumption based on simulation observations. In [6]–[9], [11], optimal energy-efficient orthogonal frequency-division multiple access (OFDMA) is designed to balance the circuit power consumption as well as the transmission power consumption on all OFDM subchannels. Furthermore, it is shown in [12], [13] that energy-efficient power control in multi-cell networks improves not only energy efficiency but also spectral efficiency uniformly for all users because of the conservative nature of power optimization, which reduces other-cell interference to improve the overall network throughput.

On the other hand, multiple-input multiple-output (MIMO) has been a key technology for wireless systems because of its potential to achieve high capacity, increased diversity, and interference suppression [14]. In a multi-user scenario, multi-user multiple-input multiple-output (MU-MIMO) systems can provide a substantial gain in networks by allowing multiple users to communicate in the same frequency and time slots [15], [16]. MU-MIMO takes the advantage of both high capacity achieved with MIMO processing and the benefits of space-division multiple access and has been accepted by major wireless standards like IEEE 802.16m [17] and 3GPP Long Term Evolution (LTE) [18]. Recently there has been some research interest in energy-efficient communications for MIMO systems [19]–[22]. For example, in [20], MIMO systems based on Alamouti diversity schemes are studied to improve the energy efficiency of sensor networks. A mechanism to switch between MIMO with two transmitter antennas and SIMO to conserve mobile terminals energy is proposed in [19]. A low-complexity energy-efficient and reconfigurable reduced dimension maximum likelihood MIMO detector is proposed in [22]. However there is very limited research studying energy-efficient MU-MIMO and its optimal power allocation. The relationships between power allocated on each antenna, channel states, antenna circuit power consumption, and antennas that should be turned on, remain unclear for energy-efficient MIMO and MU-MIMO communications. This motivates the work in this paper.

In this paper, we address the energy-efficient design of uplink (UL) MU-MIMO in a single cell environment. We account for both circuit and transmission powers when designing power allocation schemes and emphasize energy efficiency over peak rates or throughput. The proposed scheme balances the energy consumption of circuit operations and RF transmissions of all users to achieve the maximum network energy efficiency, which is defined as the number of bits transmitted per Joule of energy across the whole network. We first assume that all users consume a fixed amount of circuit power in addition to the RF power and demonstrate the existence of a unique globally optimal power allocation that achieves the energy efficiency capacity. We also provide a one-dimensional iterative algorithm to obtain this optimum. In practice users may have improved circuit management capability and turn off part of

[♭] G. Miao is with the School of Information and Communication Technology, Royal Institute of Technology, Stockholm, Sweden (Email: guowang@kth.se). Part of the work was done when the author was still with Dallas Telecom Lab of Samsung Telecom America.
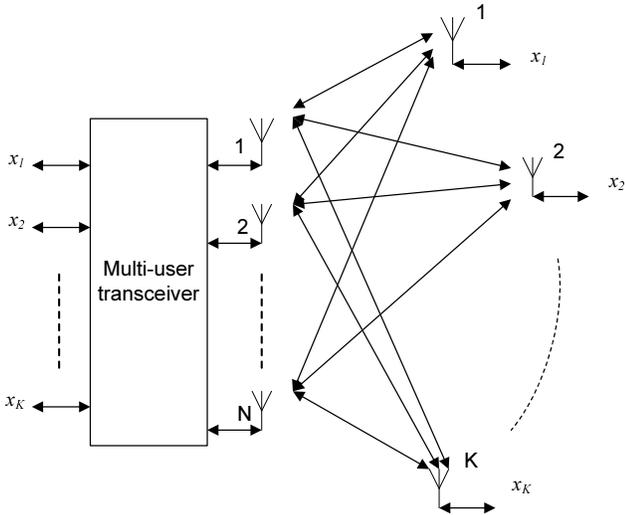
Fig. 1: System Diagram of a Multi-User System

AP is given by

$$\mathbf{y} = \mathbf{H} \cdot \mathbf{Q} \cdot \mathbf{P} \cdot \mathbf{x} + \mathbf{n} = \sum_{i=1}^{K} \mathbf{H}_i \cdot \mathbf{Q}_i \cdot \mathbf{P}_i \cdot \mathbf{x}_i + \mathbf{n}, \quad (1)$$

where $\mathbf{y} = [y_1, y_2, ..., y_N]^T$. $\mathbf{x}_i = [x_{i1}, x_{i2}, ..., x_{ik_i}]^T$, consists of transmitted signals of User i and $E[|x_{ij}|^2] = 1$, where E is the expectation. Here $[]^T$ is the transpose of a vector. $\mathbf{P}_i = \text{diag}\{\sqrt{p_{i1}}, \sqrt{p_{i2}}, ..., \sqrt{p_{ik_i}}\}$ is the power allocation matrix of User i. $\mathbf{Q}_i$ is the precoding matrix of User i. $\mathbf{H}_i$ is the $N \times k_i$ channel matrix of User i and is assumed to have rank k_i, which is generally true in a rich-scattering environment. $\mathbf{n}$ is the length-N noise vector, which is Gaussian distributed with a zero mean and a covariance matrix $\sigma^2 \mathbf{I}_N$, where $\mathbf{I}_N$ is the identity matrix of size N.

$$\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_K]^T,$$

$$\mathbf{P} = \text{diag}\{\mathbf{P}_1, \mathbf{P}_2, ..., \mathbf{P}_K\},$$

$$\mathbf{Q} = \text{diag}\{\mathbf{Q}_1, \mathbf{Q}_2, ..., \mathbf{Q}_K\},$$

and

$$\mathbf{H} = [\mathbf{H}_1, \mathbf{H}_2, ..., \mathbf{H}_K].$$

With a linear detector, the decision vector for the transmitted symbols is

$$\hat{\mathbf{x}} = \mathbf{w} \cdot \mathbf{y} = \mathbf{w} \cdot \mathbf{H} \cdot \mathbf{Q} \cdot \mathbf{P} \cdot \mathbf{x} + \mathbf{w} \cdot \mathbf{n}. \quad (2)$$

Using singular value decomposition (SVD),

$$\mathbf{H}_i = \mathbf{U}_i \begin{bmatrix} \mathbf{\Lambda}_i \\ \mathbf{0} \end{bmatrix} \mathbf{V}_i^H = [\dot{\mathbf{U}}_i \ddot{\mathbf{U}}_i] \begin{bmatrix} \mathbf{\Lambda}_i \\ \mathbf{0} \end{bmatrix} \mathbf{V}_i^H = \dot{\mathbf{U}}_i \mathbf{\Lambda}_i \mathbf{V}_i^H, \quad (3)$$

where $\mathbf{U}_i$ and $\mathbf{V}_i$ are $N \times N$ and $k_i \times k_i$ unitary matrices and $[]^H$ is the Hermitian transpose. $\dot{\mathbf{U}}_i$ consists of the first k_i columns of $\mathbf{U}_i$.

$$\mathbf{\Lambda}_i = \text{diag}\{\lambda_{i1}, \lambda_{i2}, ..., \lambda_{ik_i}\}$$

where $\lambda_{ij} \geq 0$.

With local channel knowledge $\mathbf{H}_i$, User i sets the precoding matrix $\mathbf{Q}_i = \mathbf{V}_i$. Define

$$\mathbf{U} = [\dot{\mathbf{U}}_1, \dot{\mathbf{U}}_2, ..., \dot{\mathbf{U}}_K]$$

and

$$\mathbf{\Lambda} = \text{diag}\{\mathbf{\Lambda}_1, \mathbf{\Lambda}_2, ..., \mathbf{\Lambda}_K\}.$$

It is easy to see the decision vector at the AP is

$$\hat{\mathbf{x}} = \mathbf{w} \cdot \mathbf{U} \cdot \mathbf{\Lambda} \cdot \mathbf{P} \cdot \mathbf{x} + \mathbf{w} \cdot \mathbf{n}. \quad (4)$$

There are many ways of designing the linear receiver $\mathbf{w}$. Since receiver design is not the focus of this paper, we use the zero-forcing (ZF) receiver [24], i.e.,

$$\mathbf{w} = (\mathbf{U}^H \mathbf{U})^{-1} \mathbf{U}^H, \quad (5)$$

in its simplicity. Note that the restriction on $\sum_{i=1}^{K} k_i \leq N$ is needed for the existence of the ZF receiver. The decision vector is

$$\hat{\mathbf{x}} = \mathbf{\Lambda} \cdot \mathbf{P} \cdot \mathbf{x} + \hat{\mathbf{n}}, \quad (6)$$

the circuit operations when some antennas are not used to reduce the circuit power consumption. Therefore we further study energy-efficient UL MU-MIMO with improved circuit management. We show that in this case, the problem is non-concave and multiple local maximums may exist. However, we will develop algorithms that converge to the global optimum.

The rest of the paper is organized as follows. In Section II, we formulate the problem and define energy-efficient UL MU-MIMO. In Section III and IV, we investigate optimal conditions for energy-efficient transmission and develop an algorithm to obtain the globally optimal solution. In Section V, we study energy-efficient MU-MIMO with improved circuit management. Simulation results are provided in Section VI to demonstrate the performance improvement. Finally, we conclude the paper in Section VII.

II. ENERGY-EFFICIENT MU-MIMO

In this section, we introduce energy-efficient MU-MIMO. Throughout the paper, matrices are shown with capital boldface letters, vectors with lowercase boldface, and scalars with either upper or lowercase letters without boldface.

Consider a MU-MIMO system, as illustrated in Fig. 1, where one access point (AP) is serving K users that desire best-effort data service, e.g. file transfer and email, and have no data rate requirements. Both the AP and all users desire energy-efficient communications. The AP has N antennas. User i has k_i antennas and $\sum_{i=1}^{K} k_i \leq N$. Assume block fading [23], that is, the channel state remains constant during each data frame. The channel state information (CSI) between the AP and users is predetermined earlier through either training pilots as in a time-division duplex system or a feedback channel as in a frequency-division duplex system. Each user has its CSI while the AP has CSI of all users. Signaling overhead and incomplete channel state information will result in performance loss and the study on its impact is beyond the scope of this paper. In a flat-fading propagation environment, the received signal at the

where $\hat{\mathbf{n}} = (\mathbf{U}^H\mathbf{U})^{-1}\mathbf{U}^H \cdot \mathbf{n}$, which is also Gaussian distributed with a zero mean and a covariance matrix

$$E[\hat{\mathbf{n}}\hat{\mathbf{n}}^H] = \sigma^2[(\mathbf{U}^H\mathbf{U})^{-1}]^H, \tag{7}$$

with all elements in the diagonal being σ^2.

From (6), the transmissions of different users are uncoupled. The AP can detect each symbol independently and the achieved *signal-to-noise ratio* (SNR) of all the symbols for User i is

$$\eta_i = \left[\frac{p_{i1}\lambda_{i1}^2}{\sigma^2}, \frac{p_{i2}\lambda_{i2}^2}{\sigma^2}, ..., \frac{p_{ik_i}\lambda_{ik_i}^2}{\sigma^2} \right]^T. \tag{8}$$

Given the transceiver structure and the channel state, each user determines the optimal data rate and power on each antenna. Define the data rate vector of User i to be, $\mathbf{R}_i = [r_{i1}, r_{i2}, ..., r_{ik_i}]$. Correspondingly, the overall data rate is

$$R_i = \sum_{k=1}^{k_i} r_{ik}. \tag{9}$$

Define B as the system bandwidth. The achievable data rate r_{ik} is determined by [25]

$$r_{ik} = B\log_2(1 + \frac{\eta_{ik}}{\Gamma}), \tag{10}$$

where $\eta_{ik} = \frac{p_{ik}\lambda_{ik}^2}{\sigma^2}$ and Γ is the SNR gap that defines the gap between the channel capacity and a practical coding and modulation scheme, and other implementation factors. For instance, for a coded *quadrature amplitude modulation* (QAM) system, the gap is given by [25]

$$\Gamma = 10^{((9.8 + \gamma_m - \gamma_c)/10)}, \tag{11}$$

where γ_m is the system design margin and γ_c is the coding gain in dB. The SNR gap may also capture other effects like the several dBs performance loss because of using the simple zero-forcing receiver. If Shannon capacity [26] is achieved, $\Gamma = 1$. Define the overall transmission power of User i to be P_{Ti} such that

$$P_{Ti} = \frac{\sum_{k=1}^{k_i} p_{ik}}{\zeta}, \tag{12}$$

where $\zeta \in [0, 1]$ is the power amplifier efficiency. We assume all users are operating in the linear ranges of their power amplifiers and do not assume power constraints for any user. This is because energy-efficient communications are different from spectral-efficient communications. With spectral-efficient communications, users want to use as much power as possible to achieve high throughput and it is necessary to consider power constraints for power allocation [27]. But with energy efficient communications, all users tend to use as little power as possible. Therefore the chance that the power allocated exceeds the amplifier linear range is very limited and there is no need to consider power constraints particularly.

In addition to transmission power, mobile devices also incur additional circuit power consumption owing to inevitable electronic operations which are relatively independent of the radio frequency (RF) transmission [10], [20]. Define the circuit power of User i as P_{Ci}. The overall power consumption of User i will then be

$$P_i = P_{Ci} + P_{Ti}. \tag{13}$$

The AP also consumes electronic circuit energy to receive and decode signals. Define the receiver circuit power as P_r. Similar to the circuit power, P_r models the average energy consumption of AP device electronics, such as mixers, filters, and analog-to-digital converters. For readers not interested in the receiver circuit power, they can assume $P_r = 0$. It is desirable to maximize the amount of data sent with a given amount of energy. The amount of energy $\triangle e$ consumed in a small duration, $\triangle t$, is

$$\triangle e = \triangle t \left(\alpha \sum_i P_i + \beta P_r \right),$$

where the weights $\alpha \in [0, 1]$ and $\beta \in [0, 1]$ characterize the priorities of transmitter and receiver power consumptions. For example $\alpha = 1$ and $\beta = 0$ indicate that the receiver power consumption is not considered. The MU-MIMO system wants to send a maximum amount of data by choosing the optimal transmission power allocation to maximize

$$\frac{\sum_i R_i \triangle t}{\triangle e}, \tag{14}$$

which is equivalent to maximizing

$$U(\mathbf{P}) = \frac{\sum_i R_i}{\alpha \sum_i P_i + \beta P_r}, \tag{15}$$

U is the total number of bits sent per *Joule* of energy consumption. U is called the energy efficiency of MU-MIMO. The unit of the energy efficiency is bits per Joule, which has been frequently used in literature for energy-efficient communications [3], [4], [28]–[30]. At metric (15), the energy used for sending each information bit is minimized. The energy efficiency capacity of MU-MIMO is defined as

$$U^* = \max_{\mathbf{P}} \frac{\sum_i R_i}{\alpha \sum_i (P_{Ti} + P_{Ci}) + \beta P_r}, \tag{16}$$

and the optimal energy-efficient power allocation achieving the energy efficiency capacity is

$$\mathbf{P}^* = \arg\max_{\mathbf{P}} U = \arg\max_{\mathbf{P}} \frac{\sum_i R_i}{\alpha \sum_i (P_{Ti} + P_{Ci}) + \beta P_r}. \tag{17}$$

When $K = 1$, (16) and (17) give the energy efficiency capacity and the optimal power allocation for a point-to-point MIMO system. Therefore the results in this paper are also applicable to MIMO systems.

III. PRINCIPLES OF ENERGY-EFFICIENT MU-MIMO POWER ALLOCATION

In the following, we demonstrate that a unique globally optimal power allocation always exists and gives the necessary and sufficient conditions for a power allocation scheme to

achieve the energy efficiency capacity. It is proved in Appendix I that U has the following properties.

Lemma 1. *U is strictly quasi-concave in $\mathbf{P}$.*

For a strictly quasi-concave function, if a local maximum exists, it is also globally optimal [31]. Hence, a unique globally optimal power allocation always exists and is summarized in Theorem 1 according to the proof in Appendix I.

Theorem 1. *There exists a unique globally optimal energy-efficient power allocation $\mathbf{P}^*$ that achieves the energy efficiency capacity, where p_{ik}^* is given by*

$$p_{ik}^* = \begin{cases} \frac{B\zeta}{\alpha U^* \ln 2} - \frac{\Gamma\sigma^2}{\lambda_{ik}^2} & \text{if } \frac{B\zeta\lambda_{ik}^2}{\alpha\Gamma\sigma^2 \ln 2} > U^*, \\ 0 & \text{otherwise,} \end{cases} \quad (18)$$

Correspondingly, the energy efficiency capacity is

$$U^* = U(\mathbf{P}^*). \quad (19)$$

Theorem 1 says that the the kth antenna of User i should be used only when the corresponding spatial channel, characterized by λ_{ik}^2, is sufficiently good such that using it improves the overall network energy efficiency. The power allocation is indeed a weighted water-filling solution and different users may have different water levels. The water level is determined by the energy efficiency capacity. The relative difference of power allocation of different users on different antennas depends on the channel gains of those spatial channels.

Based on Theorem 1, we have the following basic properties of power allocation.

Proposition 1. *The energy efficiency capacity decreases strictly, while the optimal allocated power on each spatial channel, if nonzero, increases strictly with the circuit power of any user.*

Proof: Define $\mathbf{P}^*$ to be the optimal power allocation given a set of circuit power conditions $\{P_{C_i}\}$. The achieved energy efficiency is U^*. Suppose any P_{C_i} decreases a certain amount to $P_{C_i} - \Delta P_C$. With the same power allocation $\mathbf{P}^*$, the energy efficiency is higher than U^*. Hence the energy efficiency capacity increases. Therefore energy efficiency capacity decreases strictly with circuit power. Furthermore, according to (18), the optimal power on each spatial channel, if nonzero, decreases strictly with energy efficiency capacity. The proposition follows. ∎

The main intuition behind Proposition 1 is that as circuit power increases, higher power should be allocated to achieve higher data rate such that each information bit can be transmitted faster and less circuit energy is consumed. Similarly we have Proposition 2.

Proposition 2. *When receiving power is considered ($\beta > 0$), the energy efficiency capacity decreases strictly, while the optimal allocated power on each spatial channel, if nonzero, increases strictly with the receiving power.*

IV. A ONE-DIMENSIONAL LOW-COMPLEXITY ALGORITHM

Theorem 1 provides the necessary and sufficient condition for a power allocation to be the unique and globally optimum one. However, it is difficult to directly solve the joint nonlinear equations in Theorem 1. Therefore, we develop an iterative method to search for the optimal $\mathbf{P}^*$ based on the analysis of the optimal power allocation in Theorem 1.

Define $p_{ik}(\mu) = \left[\mu - \frac{\Gamma\sigma^2}{\lambda_{ik}^2}\right]^+$, where $[x]^+ = \max(x, 0)$, and the corresponding power allocation matrix to be $\mathbf{P}(\mu)$. Clearly when $\mu = \frac{B\zeta}{\alpha U^* \ln 2}$, $\mathbf{P}(\mu) = \mathbf{P}^*$. Define

$$f(\mu) = U(\mathbf{P}(\mu)) \quad (20)$$

and it is easy to see that the optimal μ^* that maximizes $f(\mu)$ equals $\frac{B\zeta}{\alpha U^* \ln 2}$. Therefore we only need to find μ^* such that

$$\mu^* = \arg\max_\mu f(\mu). \quad (21)$$

As shown in Appendix II, when $f(\mu) > 0$, $f(\mu)$ is strictly quasi-concave in μ. Hence a unique globally optimal μ^* exists such that for any $\mu < \mu^*$, $f'(\mu) > 0$, and for any $\mu > \mu^*$, $f'(\mu) < 0$. Assume $\mu_1 \le \mu^* \le \mu_2$. To determine μ^*, let $\hat{\mu} = \frac{\mu_1 + \mu_2}{2}$. If $f'(\mu)|_{\hat{\mu}} = 0$, μ^* is found. If $f'(\mu)|_{\hat{\mu}} < 0$, then $\mu_1 < \mu^* < \hat{\mu}$ and replace μ_2 with $\hat{\mu}$; otherwise, replace μ_1 with $\hat{\mu}$. This iteration continues until $\mu_2 - \mu_1$ is sufficiently small to meet the convergence requirement. This energy-efficient MU-MIMO power allocation (EMMPA) algorithm is summarized in Table I.

Algorithm *Energy-Efficient MU-MIMO Power Allocation*
1. $\mu_1 \leftarrow \min_{i,j} \frac{\Gamma\sigma^2}{\lambda_{ij}^2}$
($*$ μ^* is above $\min_{i,j} \frac{\Gamma\sigma^2}{\lambda_{ij}^2}$; otherwise $p_{ik} = 0$ for all i, k $*$)
2. $\alpha \leftarrow$ a value above 1, e.g. 10; $\mu_2 \leftarrow \mu_1 * \alpha$
3.   **while** $f'(\mu_2) > 0$
4.   **do** $\mu_1 \leftarrow \mu_2$, $\mu_2 \leftarrow \mu_2 * \alpha$
5.   **while** no convergence
($*$ search the optimum iteratively $*$)
6.   **do** $\mu \leftarrow \frac{\mu_2 + \mu_1}{2}$;
7.   **if** $f'(\mu) > 0$
8.   **then** $\mu_1 \leftarrow \mu$;
9.   **else** $\mu_2 \leftarrow \mu$
10. **return** μ and $p_{ik} \leftarrow \left[\mu - \frac{\Gamma\sigma^2}{\lambda_{ik}^2}\right]^+$.

TABLE I: Energy-Efficient MU-MIMO Power Allocation

The global convergence to the optimal allocation of EMMPA is guaranteed by the strict quasi-concavity of $f(\mu)$ [32] and the convergence rate in Proposition 3 can be easily proven.

Proposition 3. *EMMPA converges to the globally optimal μ^*. Any μ, which satisfies $|\mu - \mu^*| \le \epsilon$, can be found within at most $\lceil \log_2(\frac{(\alpha-1)\mu^*}{\epsilon} - 1)\rceil$ iterations.*

The EMMPA algorithm should be implemented at the AP. Each user needs to report its circuit power to the AP before the communications. This is a one-time report and the signaling

overhead is negligible. After running EMMPA, the AP only needs to broadcast μ^* and all users can determine their optimal power allocations according to (18).

V. ENERGY-EFFICIENT MU-MIMO WITH IMPROVED CIRCUIT MANAGEMENT

In the above we have assumed User i consumes a fixed amount of circuit power, P_{C_i}, regardless of how many antennas are used. However, according to Theorem 1, the power allocated on some antennas may be zero. User i can turn off these antennas to reduce circuit energy consumption. With the improved circuit management, circuit power is a function of the set of antennas that are turned on, e.g., a function of the number of antennas turned on. In the following, for notation simplicity, assign the circuit power of User i to $P_{C_i}(k_i^o)$, where k_i^o is the number of antennas that have positive power allocation. $P_{C_i}(k_i^o)$ is increasing in k_i^o. Please note the result in this section also applies in generic cases such that P_{C_i} is a function of the set of antennas that are turned on. A simple example is

$$P_{C_i}(k_i^o) = k_i^o P_\alpha + I(k_i^o)P_\beta, \tag{22}$$

where P_α is the extra antenna-related circuit power consumption when one more antenna is turned on and P_β is the power consumption of circuit components that are independent of the number of antennas turned on. When $k_i^o = 0$, the user can be turned off completely to avoid any circuit power consumption and $P_{C_i}(0) = 0$. The indicator function $I(A)$ is defined as

$$I(A) = \begin{cases} 1 & \text{if } A > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{23}$$

In this scenario, the energy efficiency capacity is given by

$$U^* = \max_{\mathbf{P}} \tilde{U}(\mathbf{P}) = \max_{\mathbf{P}} \frac{\sum_i R_i}{\alpha \sum_i (P_{Ti} + P_{Ci}(k_i^o)) + \beta P_r}, \tag{24}$$

and the optimal energy-efficient power allocation achieving the energy efficiency capacity is

$$\mathbf{P}^* = \arg\max_{\mathbf{P}} \tilde{U}(\mathbf{P}) = \arg\max_{\mathbf{P}} \frac{\sum_i R_i}{\alpha \sum_i (P_{Ti} + P_{Ci}(k_i^o)) + \beta P_r}, \tag{25}$$

where $k_i^o = \sum_k I(p_{ik})$.

A. Principles of Energy-Efficient Power Allocation

With improved circuit management, the energy efficiency function $\tilde{U}$ is no longer continuous or quasi-concave in $\mathbf{P}$. Theorem 1 is not appropriate in characterizing the globally optimal power allocation. Observe Antenna j of User i and define it to be Antenna (i,j). Assume the power on all other antennas have been optimally allocated and define $\mathbf{P}_{ij}^{(o)}(p_{ij})$ to be the power allocation that equals the optimal power allocation except that the power on Antenna (i,j) is p_{ij}. Antenna (i,j) may have two states, on or off. If it is on, the energy efficiency is

$$\tilde{U}\left(\mathbf{P}_{ij}^{(o)}(p_{ij})\right) = \frac{R_{ij}^o + B\log_2\left(1 + \frac{p_{ij}\lambda_{ij}^2}{\Gamma\sigma^2}\right)}{P_{ij}^o + p_{ij}\frac{\alpha}{\zeta} + P_{Ci}(k_i^o)}, \tag{26}$$

where $R_{ij}^o = \sum_{\{u,k:u\neq i,k\neq j\}} B\log_2\left(1 + \frac{p_{uk}^*\lambda_{uk}^2}{\Gamma\sigma^2}\right)$ and $P_{ij}^o = \alpha\sum_{u\neq i} P_{Ci}(k_u^o) + \beta P_r + \frac{\alpha}{\zeta}\sum_{\{u,k:u\neq i,k\neq j\}} p_{uk}^*$. Similar to the proof in Appendix 1, $\tilde{U}$ is also strictly quasi-concave in p_{ij}. Hence there is a unique p_{ij} that maximizes $\tilde{U}$ if $p_{ij} > 0$, i.e., Antenna (i,j) should be turned on. In the following we study the condition that Antenna (i,j) should be turned on. The partial derivative of $\tilde{U}$ with respect to p_{ij} is

$$\frac{\partial \tilde{U}}{\partial p_{ij}} = \frac{f(p_{ij})}{(1 + \frac{p_{ij}\lambda_{ij}^2}{\Gamma\sigma^2})(\frac{\alpha}{\zeta}p_{ij} + P_{ij}^o + P_{Ci}(k_i^o))^2}, \tag{27}$$

where $f(p_{ij}) = \frac{B}{\ln 2}\frac{\lambda_{ij}^2}{\Gamma\sigma^2}(\frac{\alpha}{\zeta}p_{ij} + P_{ij}^o + P_{Ci}(k_i^o)) - \frac{\alpha}{\zeta}R_{ij}^o(1 + \frac{p_{ij}\lambda_{ij}^2}{\Gamma\sigma^2}) - B\frac{\alpha}{\zeta}(1 + \frac{p_{ij}\lambda_{ij}^2}{\Gamma\sigma^2})\log_2(1 + \frac{p_{ij}\lambda_{ij}^2}{\Gamma\sigma^2})$. Because $\tilde{U}$ is strictly quasi-concave, if there exists a $p_{ij} > 0$ such that $\frac{\partial \tilde{U}}{\partial p_{ij}} = 0$, it is unique. This further indicates that if $p_{ij} > 0$, there is only one p_{ij} such that $f(p_{ij}) = 0$. In Appendix III, we prove that only when

$$\frac{\lambda_{ij}^2}{\sigma^2} > \frac{R_{ij}^o\alpha\Gamma\ln 2}{(P_{ij}^o + P_{Ci}(k_i^o))B\zeta}, \tag{28}$$

where k_i^o is the number of antennas when Antenna (i,j) is turned on, will there exist a $p_{ij} > 0$ such that $f(p_{ij}) = 0$. This is also the condition for Antenna (i,j) to be turned on.

Readily, by setting $\frac{\partial \tilde{U}}{\partial p_{ij}}$ to be zero, we have the following necessary condition of globally optimal energy-efficient power allocation.

Theorem 2. *With improved circuit management, the optimal energy-efficient power allocation $\mathbf{P}^*$ achieving the energy efficiency capacity satisfies, for antennas that are turned on,*

$$p_{ij}^* = \frac{B\zeta}{\alpha U^*\ln 2} - \frac{\Gamma\sigma^2}{\lambda_{ij}^2} \tag{29}$$

and these antennas have channel conditions

$$\frac{\lambda_{ij}^2}{\sigma^2\Gamma} > \frac{R_{ij}^o\alpha\ln 2}{(P_{ij}^o + P_{Ci}(k_i^o))B\zeta}, \tag{30}$$

where k_i^o is the number of antennas of User i when Antenna (i,j) is turned on. Correspondingly, the energy efficiency capacity is

$$U^* = \tilde{U}(\mathbf{P}^*). \tag{31}$$

According to Theorem 2, whether or not Antenna (i,j) should be turned on is determined by multiple factors. $\frac{\lambda_{ij}^2}{\sigma^2\Gamma}$ characterizes the channel condition of Antenna (i,j) and determines the effective receiver SNR once the power is allocated. If it is above the threshold $\frac{R_{ij}^o\alpha\ln 2}{(P_{ij}^o + P_{Ci}(k_i^o))B\zeta}$, Antenna (i,j) should be used since using it improves the overall network energy efficiency. The threshold decreases with the power amplifier efficiency, ζ, indicating that with improved power amplifier efficiency, worse channel conditions can be used for data transmission on Antenna (i,j) to achieve energy

efficiency capacity. The threshold is also determined by the states of all other antennas. The energy efficiency achieved by all other antennas is $\frac{R_{ij}^o}{P_{ij}^o + P_{Ci}(k_i^o)}$ assuming Antenna (i,j) has an infinite small amount of power allocated. The threshold increases with $\frac{R_{ij}^o}{P_{ij}^o + P_{Ci}(k_i^o)}$, which indicates that if the energy efficiency achieved by all other antennas increases, the channel state for Antenna (i,j) should also be better such that turning it on improves the overall network energy efficiency.

Similar to the proof of Proposition 1, we have the following property of energy-efficient MU-MIMO with improved circuit power management.

Proposition 4. *With improved circuit management, the energy efficiency capacity decreases strictly and the optimal allocated power on each spatial channel, if nonzero, increases strictly with the circuit power of any antenna that is on. If receiving power is considered ($\beta > 0$), the energy efficiency capacity decreases strictly while the optimal allocated power on each spatial channel, if nonzero, increases strictly with the receiving power.*

B. Algorithm Development

Different from Theorem 1, Theorem 2 only gives the necessary conditions of globally optimal energy efficient power allocation. With improved circuit power management, there may be multiple power allocation schemes that satisfy Theorem 2 because there may be multiple local maximums of the energy efficiency function $\tilde{U}$. An example is given in Fig. 2, where we assume one user with two antennas is communicating to the AP, i.e., a MIMO system. The circuit power of the user is assumed to be $P_{C_1} = k_1^o + I(k_1^o)$. Observing Fig. 2, $\tilde{U}$ has three local maximums, each of which satisfies Theorem 2. When both antennas are turned on, there is a unique power allocation that maximizes $\tilde{U}$. When the state of one antenna switches from on to off, i.e., p_{11} or p_{12} goes to zero, the energy efficiency $\tilde{U}$ increases abruptly because of the reduction of circuit power. In this case, we need to compare the three local maximums to determine the energy efficiency capacity and the optimal power allocation. Therefore, to find the optimal power allocation, we can use EMMPA in Table I to determine the corresponding optimal power allocation and the maximum energy efficiency achieved when Antenna $(1,1)$ is on, $(1,2)$ is on, or both are on. The antenna configuration and power allocation that achieve the highest energy efficiency is the optimal one.

1) An Exhaustive Search Algorithm: For a generic case, define the antenna configuration to be a binary vector, $\mathbf{c}$, of length $\sum_i k_i$, in which 1 indicates the corresponding antenna is on and 0 otherwise. Given a certain antenna configuration, the circuit power of all users is determined and the power allocation is the same as that in Section III. So EMMPA can be used to determine the corresponding optimal power allocation for that antenna configuration. Therefore we need to determine the globally optimal antenna configuration and use EMMPA to find the corresponding power allocation. One simple approach is that we can exhaustively search all antenna configurations and use EMMPA to determine the maximum energy efficiency achieved for each configuration. The antennas configuration and corresponding power allocation that achieves the highest energy efficiency is the globally optimal one. We call this the exhaustive search power allocation (ESPA) algorithm and summarize it in Table II.

Algorithm *Exhaustive search power allocation*
1. $U_{max} \leftarrow 0$; $\mathbf{P}^* \leftarrow \mathbf{0}$;
2.   **for** all antenna configurations
3.   Calculate the circuit power for each user;
4.   Use EMMPA to find the optimal μ, power allocation $\mathbf{P}$, and the EE U;
5.   **if** $U > U_{max}$ and all antennas turned on have positive power allocation
6. $\mathbf{c} \leftarrow$ current antenna configuration;
7. $\mu^* = \mu$, $U_{max} \leftarrow U$, and $\mathbf{P}^* \leftarrow \mathbf{P}$;
8.   **return** $\mathbf{c}$, μ^*, $\mathbf{P}^*$, and U_{max}.
TABLE II: ESPA for MU-MIMO with Improved Circuit Management

The complexity of ESPA grows exponentially with the total number of antennas of all users and based on Proposition 3, it can be easily shown that the convergence rate is characterized by Proposition 5.

Proposition 5. *ESPA converges to the globally optimal power allocation. The optimal antenna configuration, as well as the power allocation for antennas turned on, $p_{ik} = \left[\mu - \frac{\Gamma\sigma^2}{\lambda_{ik}^2} \right]^+$, where μ satisfies $\left| \mu - \frac{B\zeta}{\alpha U^*} \right| \leq \epsilon$, can be found within at most $\left\lceil \log_2 \left(\frac{(\alpha-1)\mu^*}{\epsilon} - 1 \right) \right\rceil (2^{\sum_i k_i} - 1)$ iterations.*

2) A Quadratic-Complexity Algorithm: For a small number of users and antennas, ESPA is effective in finding the globally optimal solution. When there are many users in the system and each has many antennas, ESPA has high complexity. In the following, we further develop a low-complexity algorithm. This algorithm consists of two steps. In the first step, the linear-complexity EMMPA is used to filter out antennas that are a subset of antennas that should be turned off in the globally optimal antenna configuration. These antennas are turned off because of their poor channel conditions. In the second step, the remaining antennas are examined and some may be turned off to achieve higher energy efficiency because they consume relatively large amounts of circuit power. We will show that while the proposed algorithm has quadratic complexity, it is also globally optimal.

In the first step, assume all antennas are turned on and the circuit power, $P_{C_i}(k_i^o)$, of all users can be determined. According to Theorem 1, antennas with channel conditions $\frac{B\zeta\lambda_{ik}^2}{\alpha\Gamma\sigma^2 \ln 2} \leq U^*$ will have zero power allocation and we define these antennas by set $S^{(1)}$. If any antenna is turned off, the circuit power of the corresponding user is reduced. According to Proposition 1, the optimal power allocated on each spatial channel, if nonzero, will be reduced. Therefore, any antenna in $S^{(1)}$ will not be allocated positive power if any antenna is turned off. This indicates that $S^{(1)}$ belongs to the set of
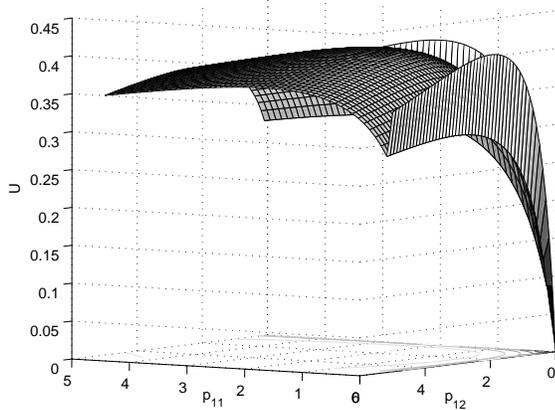
Fig. 2: An example of non-quasi-concave energy efficiency function $\tilde{U}$.

antennas that should be turned off in the globally optimal antenna configuration. $S^{(1)}$ can be determined by EMMPA. In the second round, we turn off all antennas in $S^{(1)}$ and calculate the circuit power, $P_{C_i}(k_i^o)$, of all users. Then we use EMMPA again to determine $S^{(2)}$, the set of antennas that should be turned off in this round. Similarly $S^{(2)}$ also belongs to the set of antennas that should be turned off in the globally optimal antenna configuration. In addition, $S^{(2)}$ is a super set of $S^{(1)}$. We iterate this process until reaching a round when all antennas turned on have positive power allocation. This Iterative EMMPA algorithm is summarized in Table III, whose property is given in Proposition 6 based on the above analysis.

Algorithm *Iterative EMMPA*
1. Let $S^{(0)}$ be an empty set.
2. Assume all antennas are turned on and calculate the circuit power, $P_{C_i}(k_i^o)$, of all users.
3. Use EMMPA to determine $S^{(1)}$ and the corresponding optimal μ^* and $\mathbf{P}^*$. $m \leftarrow 1$.
4. **while** $S^{(m)}$ differs from $S^{(m-1)}$
5.   **do** Turn off all antennas in $S^{(m)}$ and calculate the circuit power, $P_{C_i}(k_i^o)$, of all users.
6. $m \leftarrow m + 1$.
7.   Use EMMPA to determine $S^{(m)}$ and the corresponding optimal μ^* and $\mathbf{P}^*$.
8. **return** $S^{(m)}$, μ^*, and $\mathbf{P}^*$.

TABLE III: Iterative EMMPA

Proposition 6. *The output of the iterative EMMPA algorithm, $S^{(m)}$, is a subset of antennas that should be turned off in the globally optimal antenna configuration. $S^{(m)}$ can be found within at most $\sum_i k_i - 1$ rounds.*

We can look at the iterative EMMPA algorithm from another perspective. In each round, some additional antennas are turned off to achieve higher energy efficiency. Therefore, U^* increases

in each round. More antennas may be turned off in each round because although in the first several rounds, their channel conditions may not fulfill $\frac{B\zeta\lambda_{ik}^2}{\alpha\Gamma\sigma^2\ln 2} \leq U^*$, they fulfill this condition in a later round owing to the increase of U^*. Finally, only the antennas with the best channel conditions are kept on for data transmission.

The iterative EMMPA algorithm finds a subset of antennas that should be turned off in the globally optimal antenna configuration. However some more antennas may still need to be turned off to achieve the energy efficiency capacity. An example has been given in Fig. 2. In this example, both antennas have good channel states and the iterative EMMPA algorithm determines that both antennas should be turned on. However, turning both antennas on may not be globally optimal because one of them may consume too much circuit power. Therefore, the remaining antennas that are kept on by the iterative EMMPA algorithm should be further examined.

In the second step, we determine which remaining antennas should be turned off. Define $\overline{S}^{(0)}$ to be the set of the remaining antennas and $U^{(0)}$ to be the highest energy efficiency achieved when all the remaining antennas are turned on. Define $U_{i,j}^{(0)}$ to be the highest energy efficiency achieved when only Antenna (i,j) in $\overline{S}^{(0)}$ is turned off. If turning off any antenna in $\overline{S}^{(0)}$ will not improve the energy efficiency, i.e.,

$$\max_{(i,j)\in\overline{S}^{(0)}} U_{i,j}^{(0)} \leq U^{(0)}, \tag{32}$$

no antennas in $\overline{S}^{(0)}$ should be turned off and the selection process is done. Otherwise, we turn off the antenna that results in the highest energy efficiency, i.e., Antenna (k,l), where

$$(k,l) = \arg \max_{(i,j)\in\overline{S}^{(0)}} U_{i,j}^{(0)}. \tag{33}$$

Then a higher energy efficiency $U_{k,l}^{(0)}$ is achieved. In the second round, let $U^{(1)} = U_{k,l}^{(0)}$. Define the set of remaining antennas that are still on to be $\overline{S}^{(1)}$. The above selection process can be repeated until in round m, no antennas should be turned off, i.e.,

$$\max_{(i,j)\in\overline{S}^{(m)}} U_{i,j}^{(m)} \leq U^{(m)}. \tag{34}$$

Note that m is always smaller than the number of antennas in $\overline{S}^{(0)}$ because in each round, one antenna will be turned off and finally, at least one antenna in $\overline{S}^{(0)}$ must be kept on to achieve non-zero energy efficiency.

The whole algorithm for energy-efficient MU-MIMO with improved circuit management is named improved EMMPA and summarized in Table IV. The global optimality of improved EMMPA is verified later in Section VI.

The complexity of the improved EMMPA grows quadratically with the total number of antennas on the user side and based on Proposition 3, it can be easily shown that the convergence rate is characterized by Proposition 7.

Proposition 7. *The output of the improved EMMPA is obtained*

Algorithm *Improved EMMPA*
1. Use iterative EMMPA to determine $\overline{S}^{(0)}$ and $U^{(0)}$.
2. $m \leftarrow 0$ and $U_{max} \leftarrow 0$.
3. **repeat**
4.   **for** Antenna (i,j) in $\overline{S}^{(m)}$
5.   **do** Turn on only antennas in $\overline{S}^{(m)}$ excluding (i,j) and calculate circuit power of all users.
6.   Use EMMPA to determine $U_{i,j}^{(m)}$ and the corresponding μ and $\mathbf{P}$.
7.   **if** $U_{i,j}^{(m)} > U_{max}$
8.   **then** $U_{max} \leftarrow U_{i,j}^{(m)}$ and $(k,l) \leftarrow (i,j)$. $\mu^* \leftarrow \mu$, $\mathbf{P}^* \leftarrow \mathbf{P}$.
9.   **if** $U_{max} > U^{(m)}$
10.   **then** $\overline{S}^{(m+1)} \leftarrow \overline{S}^{(m)}$ excluding (k,l)
11. $U^{(m+1)} \leftarrow U_{max}$.
12. $m \leftarrow m+1$;
13. **until** $U_{max} \leq U^{(m-1)}$.
14. **return** $\overline{S}^{(m-1)}$, μ^*, and $\mathbf{P}^*$.

TABLE IV: Improved EMMPA

within at most

$$\left\lceil \log_2(\frac{(\alpha-1)\mu^*}{\epsilon} - 1) \right\rceil \frac{1}{2}(\sum_i k_i - 1)\sum_i k_i$$

iterations.

Similar to EMMPA, the algorithms proposed in this section should also be implemented at the AP. The AP needs to broadcast both the antenna configuration and the μ^* and all users can determine their optimal power allocations according to (29).

VI. SIMULATION RESULTS FOR ENERGY-EFFICIENT MU-MIMO

The proposed energy-efficient MU-MIMO can be applied in different types of wireless networks to improve the network energy efficiency. In this section, we provide simulation results for a single-cell cellular network to demonstrate the performance of energy-efficient MU-MIMO. System parameters are listed in Table V. In each trial, users are dropped uniformly within 250 meters from the AP. The performance below is the average over all trials.

A. Performance of Energy-Efficient MU-MIMO without Improved Circuit Management

First we consider the case that no circuit management is used and each user consumes a fixed amount, $P_{C_i} = 100$ mW, of circuit power. Fig. 3 gives the average energy efficiency capacity when there are two users in the network and each user has 1, 2, 3, or 4 antennas. The average energy efficiency capacity is an average of multiple user droppings and channel realizations. The number of AP antennas is varied to observe its impact on energy efficiency capacity. Here we assume the same P_r regardless of the number of AP antennas so that

only the impact of receiver diversity on the system energy efficiency can be observed. On the other hand, Fig. 4 compares the average energy efficiency capacity when the AP has 64 antennas. We can see that without circuit management, more users and more antennas always help improve the energy efficiency capacity of MU-MIMO. Fig. 5 compares the energy efficiency of EMMPA and that of the fixed power allocation (FPA). With the fixed power allocation, each user employs a fixed amount of transmission power, given by the value in the x axis, and allocates it equally on all spatial channels. Two scenarios are considered. In one scenario, there are four users in the network, each with two antennas, and the AP has eight antennas. In the second scenario, there is only one user with four antennas in the network and the AP has eight antennas. As shown in Fig. 5, significant gain in energy efficiency can be observed by using EMMPA. The gain is even larger when there are multiple users because EMMPA effectively exploits multi-user diversity in the network to improve energy efficiency.
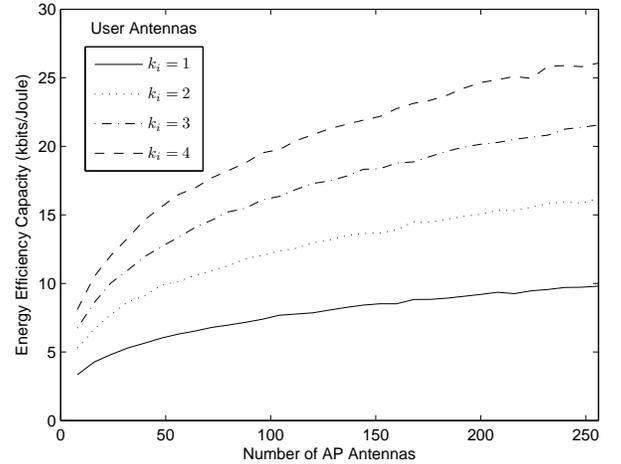

Fig. 3: Relationship between energy efficiency capacity, transmitter antennas, and receive antennas.

B. Performance of Energy-Efficient MU-MIMO with Improved Circuit Management

In the following, we consider energy-efficient MU-MIMO with improved circuit management and assume $P_{C_i}(k_i^o) = P_\alpha k_i^o + P_\beta I(k_i^o)$ mW for all users.

First we verify the global optimality of improved EMMPA. Assume $P_\beta = 20$ mW and P_α is varied to observe the impact of circuit power consumption of individual antennas on the suboptimality gap of improved EMMPA and iterative EMMPA algorithms. Fig. 6 gives the normalized energy efficiency of improved and iterative EMMPA when the AP has 16 antennas and each user has 4 antennas. The average energy efficiency achieved by improved and iterative EMMPA is normalized by that of ESPA, which is globally optimal. From Fig. 6, improved EMMPA performs exactly the same as ESPA. In addition to the same average energy efficiency, improved EMMPA always
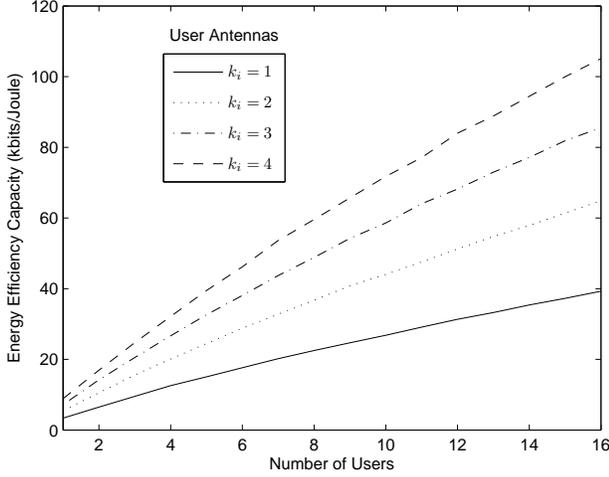
8

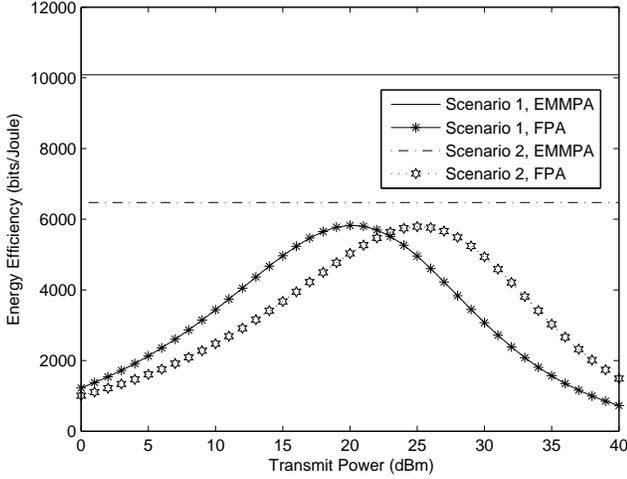Fig. 4: Relationship between energy efficiency capacity, users, and transmitter antennas.

Fig. 5: Comparison between EMMPA and Fixed Power Allocation (Scenario 1: $N = 8$, $K = 4$, $k_i = 2$; Scenario 2: $N = 8$, $K = 1$, $k_i = 4$).

obtains the same instantaneous power allocation and energy efficiency as ESPA according to our observation of simulation results. Therefore, improved EMMPA is also globally optimal. When P_α is small, iterative EMMPA performs very closely to the global optimum because channel state is the main factor that determines which antennas should be turned on. When P_α is larger, the impact of individual antenna circuit power consumption grows and since iterative EMMPA does not consider this impact, it has larger performance loss. Therefore improved EMMPA is needed to take this impact into account to achieve the globally optimal performance.

Fig. 7 compares the average computing time of improved EMMPA, iterative EMMPA, and ESPA with the same simula-

tion setting as that in Fig. 6. When K increases, the computing time of ESPA grows exponentially, more than ten times when K is increased by one. On the other hand, the computing time of iterative and improved EMMPA grows very slightly. The individual antenna circuit power consumption P_α has impact on the computing time of improved EMMPA because it determines on average how many additional rounds in the second step are needed to identify antennas that should be turned off because of high antenna circuit power consumption.

Fig. 8 gives the average energy efficiency capacity when the AP has 16 antennas while each user has two antennas. The number of users is varied from one to eight. $P_\beta = 20$ mW and P_α is varied to observe the impact of antenna circuit power on energy efficiency capacity. From Fig. 8 we can see more users always help improve the network energy efficiency because of increased multi-user diversity. Fig. 9 illustrates the average energy efficiency capacity when the AP has 16 antennas and two users are accessing the AP. The number of antennas of each user is varied from one to eight and $P_\beta = 20$ mW. We can see that when individual antennas consume a small amount of circuit power, i.e., P_α is small, more antennas always improve the network energy efficiency. However when individual antennas consume too much circuit power and P_α is large, e.g., $P_\alpha = 40$ dBm, more antennas do not help improve the network energy efficiency because most antennas should be turned off in this case to reduce circuit power consumption.
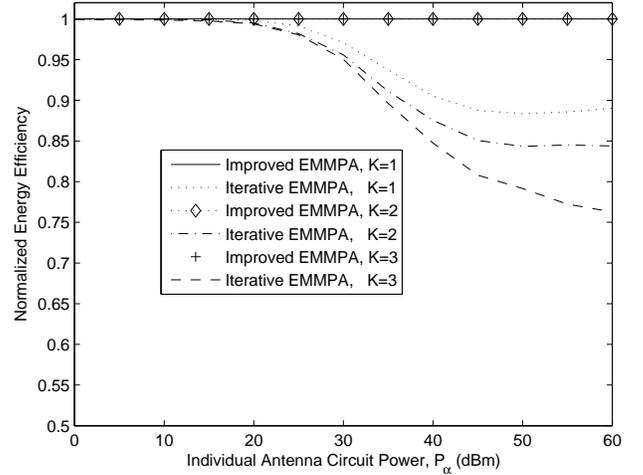

Fig. 6: Suboptimality gap of improved EMMPA and iterative EMMPA ($N = 16$ and $k_i = 4$).

Fig. 10 compares the average energy efficiency of improved EMMPA, iterative EMMPA, and FPA in the same two scenarios as in Fig. 5. We can see iterative EMMPA achieves very close performance to that of improved EMMPA. This indicates that with $P_{C_i}(k_i^o) = 20k_i^o + 40I(k_i^o)$, the impact of circuit power consumed by different antennas on energy efficient MU-MIMO is negligible and the channel states mainly determine which antennas should be turned on. However, when the amount of circuit power consumed by each
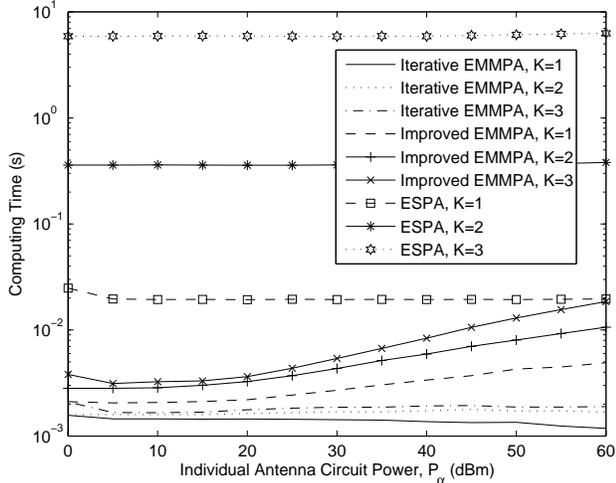
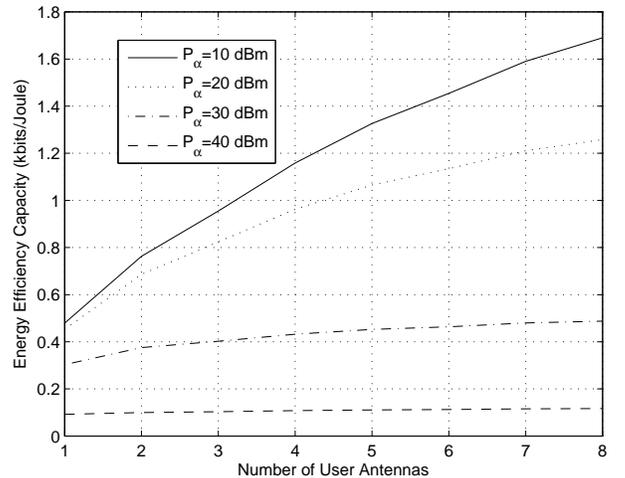Fig. 7: Complexity comparison of improved EMMPA, iterative EMMPA, and ESPA ($N = 16$ and $k_i = 4$).

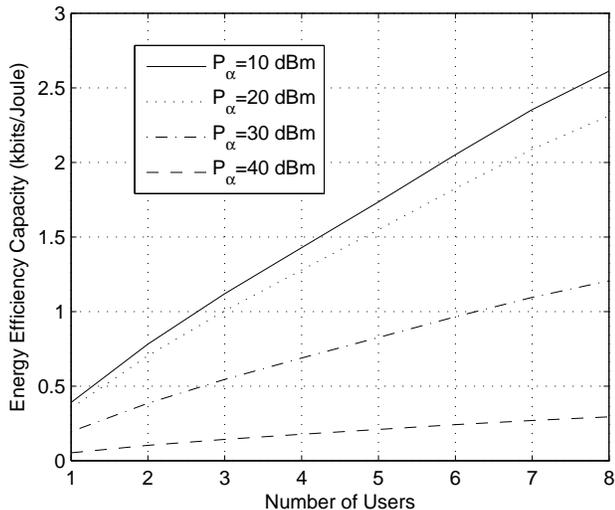Fig. 9: Relationship between energy efficiency capacity and user antennas ($N = 16$ and $K = 2$).

Fig. 8: Relationship between energy efficiency capacity and number of users ($N = 16$ and $k_i = 2$).

TABLE V: Simulation Parameters

Carrier frequency	1.5 GHz
System bandwidth	10 kHz
Thermal noise power, N_o	-141 dBW/MHz
User antenna height	1.6 m
BS antenna height	40 m
Environment	Macro cell in urban area
Receiver power, P_r	1000 mW
Propagation Model	Okumura-Hata model
Shadowing	10 dB lognormal
Fading	Rayleigh flat fading
Power amplifier efficiency, ζ	0.5
SNR gap, Γ	0 dB
α	1
β	1

antenna increases, the impact would grow and the performance difference between improved EMMPA and iterative EMMPA may not be negligible, as shown in Fig. 6. Compared to FPA, significant gain of energy efficiency can be observed. For example, in scenario 1, more than 100% gain can be observed, indicating half energy is needed to transmit the same amount of information bits using energy-efficient MU-MIMO.

VII. CONCLUSION

In this paper, we have investigated the optimal energy-efficient MU-MIMO. Both electronic circuit and RF transmission power consumptions have been considered. We have first analyzed an MU-MIMO system based on distributed SVD decomposition of the channels of all users and derived the achieved SNR conditions for all users. Then we have proposed the concept of energy-efficient MU-MIMO and defined the energy efficiency capacity for MU-MIMO. We have demonstrated the existence of a uniquely globally optimal power allocation that could achieve this energy efficiency capacity. The optimal power allocation is shown to be a dynamic water-filling approach where the water level is determined by the energy efficiency capacity. A one-dimensional low-complexity algorithm has been developed to obtain the globally optimal power allocation and this algorithm converges to the global optimum at an exponential speed. Furthermore, we have studied energy-efficient MU-MIMO with improved circuit management where users can choose to turn off electronic circuit operations when some antennas are not used. Our result showed that some antennas should not be used even when they have good channel states because turning them on consumes
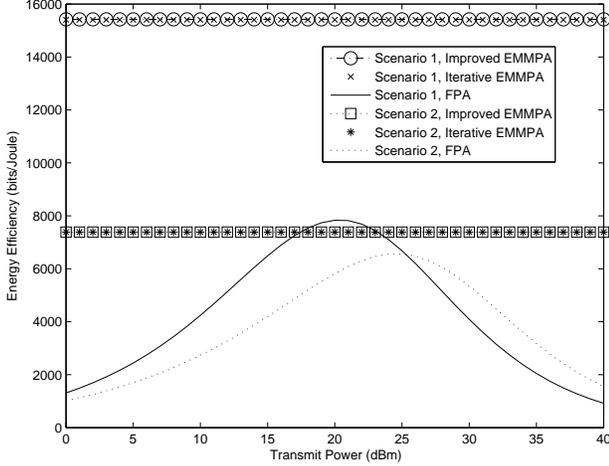
Fig. 10: Comparison between Improved EMMPA, Iterative EMMPA, and FPA (Scenario 1: $N = 8$, $K = 4$, $k_i = 2$; Scenario 2: $N = 8$, $K = 1$, $k_i = 4$).

too much circuit power. We have developed globally optimal algorithms to determine the set of antennas that should be kept on and the corresponding power allocation. Comprehensive simulation results have been provided to demonstrate the algorithm performance and the significant gain in energy efficiency for a cellular network.

This paper has focused on narrow-band flat-fading channels. However, the solutions, especially the methodology, can be easily adapted in broadband channels. For example, in LTE, MU-MIMO can be applied in each resource block (RB) to multiplex the transmission of multiple users in the network [33]. Here each RB has a small number of, e.g. 12, adjacent subcarriers and therefore experiences flat fading. We can apply the proposed power allocations schemes in each RB to improve the network energy efficiency. Some improvements may be needed. For example, as the data transmission of a user may occur in multiple RBs, the circuit power in the algorithms should be replaced by the total circuit power divided by the number of RBs used, i.e. the average circuit power per RB. The detailed discussion is beyond the scope of this paper and would be the future work.

VIII. ACKNOWLEDGEMENT

Proof: According to Section II,

$$
\begin{aligned}
U(\mathbf{P}) &= \frac{\sum_i R_i}{\alpha \sum_i (P_{Ti} + P_{Ci}) + \beta P_r} \\
&= \frac{\sum_{i=1}^{K} \sum_{k=1}^{k_i} B \log_2 \left(1 + \frac{p_{ik}\lambda_{ik}^2}{\Gamma\sigma^2}\right)}{\alpha \sum_{i=1}^{K} P_{Ci} + \beta P_r + \frac{\alpha}{\zeta} \sum_{i=1}^{K} \sum_{k=1}^{k_i} p_{ik}}.
\end{aligned}
\tag{I.35}
$$

Define the upper contour sets of $U(\mathbf{P})$ as

$$
S_\mu = \{\mathbf{P} \succeq \mathbf{0} | U(\mathbf{P}) \geq \mu\},
\tag{I.36}
$$

where symbol $\succeq$ denotes matrix inequality and $\mathbf{P} \succeq \mathbf{0}$ means each element of $\mathbf{P}$ is nonnegative. According to Proposition C.9 of [31], $U(\mathbf{P})$ is strictly quasi-concave if and only if S_μ is strictly convex for any real number μ. When $\mu < 0$, no points exist on the contour $U(\mathbf{P}) = \mu$. When $\mu = 0$, only $\mathbf{0}$ is on the contour $U(\mathbf{0}) = \mu$. Hence, S_μ is strictly convex when $\mu \leq 0$. Now we investigate the case when $\mu > 0$. S_μ is equivalent to

$$
\begin{aligned}
S_\mu = \Bigg\{ &\mathbf{P} \succeq \mathbf{0} | \mu\alpha \sum_{i=1}^{K} P_{Ci} + \mu\beta P_r + \frac{\mu\alpha}{\zeta} \sum_{i=1}^{K} \sum_{k=1}^{k_i} p_{ik} \\
&- \sum_{i=1}^{K} \sum_{k=1}^{k_i} B \log_2 \left(1 + \frac{p_{ik}\lambda_{ik}^2}{\Gamma\sigma^2}\right) \leq 0 \Bigg\}.
\end{aligned}
\tag{I.37}
$$

Clearly $\mu\alpha \sum_{i=1}^{K} P_{Ci} + \mu\beta P_r + \frac{\mu\alpha}{\zeta} \sum_{i=1}^{K} \sum_{k=1}^{k_i} p_{ik} - \sum_{i=1}^{K} \sum_{k=1}^{k_i} B \log_2 \left(1 + \frac{p_{ik}\lambda_{ik}^2}{\Gamma\sigma^2}\right)$ is strictly convex in $\mathbf{P}$ as its Hessian is positive definite. Therefore S_μ is strictly convex and we have the strict quasi-concavity of $U(\mathbf{P})$.

It is easy to see that at the local maximum of $U(\mathbf{P})$, which is also the global maximum because of the strict quasi-concavity, all $p_{ik} < \infty$ as otherwise $U(\mathbf{P}) = 0$. Hence at the local maximum, p_{ik} is positive or $p_{ik} = 0$. If p_{ik} is positive, it can be obtained by setting the partial derivative of $U(\mathbf{P})$ with respect to p_{ik} to be zero, i.e.,

$$
\left. \frac{\partial U(\mathbf{P})}{\partial p_{ik}} \right|_{\mathbf{P} = \mathbf{P}^*} = 0,
\tag{I.38}
$$

and we have

$$
p_{ik}^* = \frac{B\zeta}{\alpha U^* \ln 2} - \frac{\Gamma\sigma^2}{\lambda_{ik}^2},
\tag{I.39}
$$

Hence, the unique optimal energy-efficient power allocation is given by

$$
p_{ik}^* = \begin{cases} \frac{B\zeta}{\alpha U^* \ln 2} - \frac{\Gamma\sigma^2}{\lambda_{ik}^2} & \text{if } \frac{B\zeta\lambda_{ik}^2}{\alpha\Gamma\sigma^2} > U^*, \\ 0 & \text{otherwise.} \end{cases}
\tag{I.40}
$$

Lemma 1 and Theorem 1 are readily obtained. ∎

APPENDIX II
PROOF OF QUASI-CONCAVITY OF $f(\mu)$

Proof: Similar to the proof in Appendix I, define the upper contour sets of $f(\mu)$ as

$$S_\nu = \{\mu > 0 | f(\mu) \geq \nu\}. \tag{II.41}$$

We need to show S_ν is strictly convex for any real number ν, which is obvious when $\nu \leq 0$. When $\nu > 0$, S_ν is equivalent to

$$S_\nu = \left\{ \mu > 0 \left| \nu\alpha \sum_{i=1}^{K} P_{Ci} + \nu\beta P_r + \frac{\nu\alpha}{\zeta} \sum_{i=1}^{K} \sum_{k=1}^{k_i} \left[\mu - \frac{\Gamma\sigma^2}{\lambda_{ik}^2}\right]^+ \right.\right.$$
$$\left.\left. - \sum_{i=1}^{K}\sum_{k=1}^{k_i} B\log_2\left(1 + \frac{\left[\mu - \frac{\Gamma\sigma^2}{\lambda_{ik}^2}\right]^+ \lambda_{ik}^2}{\Gamma\sigma^2}\right) \leq 0 \right\} \right. \tag{II.42}$$
$$\triangleq \{\nu > 0 | M(\mu) \leq 0\}.$$

The second order derivative of $M(\mu)$ is

$$M''(\mu) = -d^2 \sum_{i=1}^{K}\sum_{k=1}^{k_i} B\log_2\left(1 + \frac{\left[\mu - \frac{\Gamma\sigma^2}{\lambda_{ik}^2}\right]^+ \lambda_{ik}^2}{\Gamma\sigma^2}\right) \Big/ d\mu^2$$
$$= \sum_{i,k:\mu > \frac{\Gamma\sigma^2}{\lambda_{ik}^2}} \frac{B}{\mu^2}. \tag{II.43}$$

At lease one antenna of one user should have positive power allocation as otherwise $f(\mu) = 0$, which is clearly not optimal. Therefore there exists at lease one i and k such that $\mu > \frac{\Gamma\sigma^2}{\lambda_{ik}^2}$. Hence $M''(\mu) > 0$ for all feasible μ of interest and S_ν is strictly convex and we have the strict quasi-concavity of $f(\mu)$ when $f(\mu) > 0$. ∎

APPENDIX III
CONDITION OF AN ANTENNA BEING TURNED ON

Proof: $\frac{\partial f(p_{ij})}{\partial p_{ij}} = -\frac{\lambda_{ij}^2}{\Gamma\sigma^2}\frac{\alpha}{\zeta}\left(R_{ij}^o + B\log_2(1 + \frac{p_{ij}\lambda_{ij}^2}{\Gamma\sigma^2})\right) < 0$. Therefore $f(p_{ij})$ is strictly decreasing. It is easy to see that as $p_{ij} \to +\infty$, $f(p_{ij}) \to -\infty$. If when $p_{ij} \to 0$, $f(p_{ij}) > 0$, then there exists a unique $p_{ij} > 0$ such that $f(p_{ij}) = 0$. Otherwise, Antenna (i,j) should be turned off.

$$\lim_{p_{ij} \to 0} f(p_{ij}) = \frac{B}{\ln 2}\frac{\lambda_{ij}^2}{\Gamma\sigma^2}(P_{ij}^o + P_{Ci}(k_i^o)) - \frac{\alpha}{\zeta}R_{ij}^o > 0.$$

Therefore, Antenna (i,j) should be turned on when $\frac{\lambda_{ij}^2}{\sigma^2} > \frac{R_{ij}^o\Gamma\alpha\ln 2}{(P_{ij}^o + P_{Ci}(k_i^o))B\zeta}$. ∎

REFERENCES

[1] K. Lahiri, A. Raghunathan, S. Dey, and D. Panigrahi, "Battery-driven system design: A new frontier in low power design," in *Proc. Intl. Conf. on VLSI Design*, Bangalore, India, Jan. 2002, pp. 261–267.

[2] G. W. Miao, N. Himayat, Y. Li, and A. Swami, "Cross-layer optimization for energy-efficient wireless communications: A survey," *Wireless Commun. and Mobile Computing.*, vol. 9, no. 4, pp. 529–542, Apr. 2009.

[3] S. Verdu, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory.*, vol. 48, no. 6, pp. 1319–1343, Jun. 2002.

[4] F. Meshkati, H. V. Poor, S. C. Schwartz, and N. B. Mandayam, "An energy-efficient approach to power control and receiver design in wireless networks," *IEEE Trans. Commun.*, vol. 5, no. 1, pp. 3306–3315, Nov. 2006.

[5] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-constrained modulation optimization," *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2349–2360, Sep. 2005.

[6] G. Miao, N. Himayat, Y. Li, and D. Bormann, "Energy-efficient design in wireless OFDMA," in *Proc. IEEE Conf. Commun. (ICC' 2008)*, 2008.

[7] G. W. Miao, N. Himayat, G. Y. Li, and S. Talwar, "Low-complexity energy-efficient OFDMA," in *Proc. IEEE Conf. Commun. (ICC' 2009)*, Jun. 2009, pp. 1–5.

[8] G. Miao, N. Himayat, and Y. Li, "Energy-efficient transmission in frequency-selective channels," in *Proc. IEEE Globecom 2008*, Nov. 2008, pp. 1–5.

[9] ——, "Energy-efficient link adaptation in frequency-selective channels," *IEEE Trans. Commun.*, vol. 58, no. 2, pp. 545–554, Feb. 2010.

[10] A. Y. Wang, S. Chao, C. G. Sodini, and A. P. Chandrakasan, "Energy efficient modulation and MAC for asymmetric RF microsensor system," in *Int. Symp. Low Power Electronics and Design*, Huntington Beach, CA, 2001, pp. 106–111.

[11] G. W. Miao, N. Himayat, G. Y. Li, and S. Talwar, "Low-complexity energy-efficient scheduling for uplink ofdma," *IEEE Trans. Commun.*, vol. 60, pp. 112–120, Jan. 2012.

[12] G. Miao, N. Himayat, Y. Li, A. T. Koc, and S. Talwar, "Interference-aware energy-efficient power optimization," in *Proc. IEEE ICC 2009*, Jun. 2008, pp. 1–5.

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

[14] A. J. Paulraj, D. A. Gore, R. U. Nabar, and H. Bolcskel, "An overview of MIMO communications - a key to gigabit wireless," *Proceedings of the IEEE.*, vol. 92, no. 2, pp. 198–218, Feb. 2002.

[15] Q.H. Spencer, A.L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Processing*, vol. 52, no. 2, pp. 461–471, Feb. 2004.

[16] A. Soysal and S. Ulukus, "Optimality of beamforming in fading MIMO multiple access channels," *IEEE Trans. Commun.*, vol. 57, no. 4, pp. 1171–1183, Apr. 2009.

[17] Q. Li, X. Lin, J. Zhang, and W. Roh, "Advancement of MIMO technology in wimax: from ieee 802.16d/e/j to 802.16m," *IEEE Commun. Magazine*, vol. 47, no. 6, pp. 100–107, Jun. 2009.

[18] J. Lee, J. Han, and J. Zhong, "MIMO technologies in 3GPP LTE and LTE-advnaced," *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, 2009.

[19] H. Kim, C. Chae, G. Veciana, and R. Heath, "A cross-layer approach to energy efficiency for adaptive MIMO systems exploiting spare capacity," *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 4264–4275, Aug. 2009.

[20] S. Cui A. Goldsmith, and A. Bahai, "Energy-efficiency of MIMO and cooperative MIMO techniques in sensor networks," in *IEEE J. Sel. Areas Commun.*, vol. 22, no. 6, Aug. 2004, pp. 1089–1098.

[21] N. Moezzi-Madani, T. Thorolfsson, J. Crop, P. Chiang, and W. Davis, "An energy-efficient 64-qam mimo detector for emerging wireless standards," in *Design, Automation & Test in Europe Conference & Exhibition (DATE), 2011*, Mar. 2011.

[22] E.P. Kim and N. R. Shanbhag, "An energy-efficient multiple-input multiple-output (mimo) detector architecture," in *2011 IEEE Workshop on Signal Processing Systems (SiPS)*, Oct. 2011.

[23] R. J. McEliece and W. E. Stark, "Channels with block interference," *IEEE Trans. Inf. Theory.*, vol. 30, no. 1, pp. 44–53, Jan. 1984.

[24] S. Verdu, *Multiuser Detection*. Cambridge University Press, 1998.

[25] J. M. Cioffi, *A Multicarrier Primer*. ANSI T1E1, 1999.

[26] C. E. Shannon, "Communication in the presence of noise," in *Proc. IRE*, vol. 37, Jan. 1949, pp. 10–21.

[27] W. Yu and T. Lan, "Transmitter optimization for the multi-antenna downlink with per-antenna power constraints," *IEEE Trans. Signal Processing*, vol. 55, no. 6, pp. 2646–2660, Jun. 2007.

[28] R. G. Gallager, "Power limited channels: Coding, multiaccess, and spread spectrum," in *Proc. Conf. Inform. Sci. and Syst.*, vol. 1, Mar. 1988.

[29] D. Goodman and N. Mandayam, "Power control for wireless data," *IEEE Personal Commun.*, vol. 7, no. 2, pp. 48–54, Apr. 2000.

[30] N. Feng, S. C. Mau, and N. B. Mandayam, "Pricing and power control for joint network-centric and user-centric radio resource management," *IEEE Trans. Commun.*, vol. 52, no. 9, pp. 1547–1557, Sep. 2004.

[31] E. Wolfstetter, *Topics in Microeconomics: Industrial Organization, Auctions, and Incentives*. Cambridge University Press, 1999.

[32] K. C. Kiwiel and K. Murty, "Convergence of the steepest descent method for minimizing quasiconvex functions," in *J. of Optimization Theory and Applications*, vol. 89, no. 1, Sep. 2005, pp. 221–226.

[33] Q. Li, G. Li, W. Lee, M. Lee, D. Mazzarese, B. Clerckx, and Z. Li, "Mimo techniques in wimax and lte: a feature overview," *IEEE Commun. Magazine*, vol. 48, no. 5, pp. 86–92, May 2010.

Guowang Miao (S'03, M'10) received a B.S. and a M.S. degree, in 2003 and 2006, in electronic engineering from Tsinghua University, Beijing, China, and a M.S. degree and a Ph.D. degree, both in 2009, in electrical and computer engineering from Georgia Institute of Technology, Atlanta, GA, USA. He joined Dallas Telecom Lab of Samsung, Texas, in Jan 2010 as a Senior Algorithms and Standards Engineer and worked on next-generation wireless communications technologies and 3GPP Long Term Evolution - Advanced (LTE-A) Standard, with a focus on both PHY and MAC layers. In 2011, he won an Individual Gold Award from Samsung Telecom America for his contribution in LTE-A standardization. Starting in Fall 2011, he is an assistant professor in the Department of Communications Systems, KTH - The Royal Institute of Technology, Stockholm, Sweden. His research interest is in the design and optimization of wireless communications and networking. He serves as a technical reviewer of more than twenty international journals and conferences. He has been a technical program committee member of many international conferences. He is also on the editorial board of several international journals. He was an exemplary reviewer for IEEE Communications Letters in 2011, representing fewer than 3% of all our reviewers.