Energy-Efficient Resource Allocation in SDMA Systems with Large Numbers of Base Station Antennas

Derrick Wing Kwan Ng*, Ernest S. Lo†, and Robert Schober*
*The University of British Columbia, Canada
†Centre Tecnològic de Telecomunicacions de Catalunya - Hong Kong (CTTC-HK)

***Abstract*— In this paper, resource allocation for energy efficient communication in space division multiple access (SDMA) downlink networks with large numbers of transmit antennas is studied. The considered problem is modeled as a non-convex optimization problem which takes into account the circuit power consumption and a minimum required data rate. By exploiting the properties of fractional programming, the considered non-convex optimization problem in fractional form is transformed into an equivalent optimization problem in subtractive form, which enables the derivation of an efficient iterative resource allocation algorithm. The optimal power allocation solution for each iteration is derived based on a low complexity user selection policy for maximization of the energy efficiency of data transmission (bit/Joule delivered to the users). Simulation results illustrate that the proposed iterative resource allocation algorithm converges in a small number of iterations and unveil the trade-off between energy efficiency and the number of antennas.**

I. Introduction

Recently, an increasing interest in multi-media services such as video conferencing and online high definition (HD) video streaming has led to a tremendous demand for high data rate communications with certain guaranteed quality of service (QoS) properties such as a minimum required data rate. Multiple-input multiple-output (MIMO) technology with large numbers of antennas is considered as a viable solution for addressing this issue [1]-[3]. In [1], the authors investigated the uplink sum capacity (bit/s/Hz) of cellular networks for unlimited numbers of antennas at both the base station (BS) and the users. In [2], high throughputs were shown in both uplink and downlink for a time-division duplex multi-cell system which employs multiple BSs equipped with large numbers of antennas. In [3], the authors studied the asymptotic performance of linear receivers for large numbers of transmit and receive antenna pairs. A substantial capacity gain and a better interference management were observed with MIMO compared to single antenna systems in all studies [1]-[3].Yet, the advantages of MIMO do not come for free. The extra power consumption in antenna circuitries and advanced signal processing algorithms have significant financial implications for service providers, which has been largely overlooked in the literature so far. As a result, energy efficient system designs, which adopt *energy efficiency* (bit-per-Joule) as the performance metric, have recently drawn much attention in both industry and academia.

This work has been supported in part by the Natural Science and Engineering Council of Canada (NSERC) Strategic Project Grant STPGP 396545.

In [4], a power loading algorithm is designed to minimize the energy-per-goodbit in MIMO systems. In [5], an energy efficient power allocation algorithm was studied for multi-carrier systems and frequency selective channels. In [6], a risk-return model was proposed as a performance metric for energy-efficient sub-optimal power allocation in cognitive radio multi-carrier systems. However, these works considered single user systems and the energy efficient resource allocation algorithms proposed in [4]-[6] may not be applicable in multi-user systems.

In this paper, we address the above issues. For this purpose, we formulate the resource allocation problem for energy efficient communication in space division multiple access (SDMA) systems as an optimization problem. By exploiting the properties of fractional programming, the considered non-convex optimization problem in fractional form is transformed into an equivalent optimization problem in subtractive form with a tractable solution, which can be found with an iterative algorithm. In each iteration, a closed-form power allocation solution and a low complexity user selection policy are computed for maximization of the network energy efficiency.

II. SDMA Downlink Network Model

We consider an SDMA network which consists of a BS with N_T antennas and K mobile users equipped with a single antenna. The channel gains are assumed to be time-invariant (slow fading) and known at the BS. The downlink received symbol at user $k \in \{1, \dots, K\}$ is given by

$$y_k = \sqrt{P_k l_k g_k}\mathbf{h}_k^T \mathbf{f}_k x_k + I_k + z_k, \quad (1)$$
$$I_k = \sum_{j \neq k} \mathbf{h}_k^T \mathbf{f}_j x_j s_j \sqrt{P_j l_k g_k} \quad (2)$$

where x_k and P_k are the transmitted symbol and power for the link from the BS to user k, respectively. I_k is the received interference at user k due to spatial reuse for multiple users access. $\mathbf{f}_k \in \mathbb{C}^{N_T \times 1}$ is the precoding vector. $\mathbb{C}^{N \times M}$ is the space of all $N \times M$ matrices with complex entries and $[\cdot]^T$ denotes the transpose operation. $s_j \in \{0, 1\}$ is the user selection indicator. $\mathbf{h}_k \in \mathbb{C}^{N_T \times 1}$ is the vector of small scale fading coefficients between the BS and user k. l_k and g_k represent the path loss and shadowing between the BS and user k. z_k is additive white Gaussian noise (AWGN) with zero mean and power spectral density N_0.

III. Resource Allocation and Scheduling

In this section, we introduce the adopted system performance metric and formulate the corresponding resource allocation problem.

A. Instantaneous Channel Capacity

In this subsection, we define the adopted system performance measure. Given perfect channel state information (CSI) at the receiver, the maximum channel capacity between the BS and user k with channel bandwidth W is given by

$$C_k = W\log_2\Big(1+\Gamma_k\Big), \tag{3}$$

$$\Gamma_k = \frac{P_k l_k g_k |\mathbf{h}_k^T\mathbf{f}_k|^2}{WN_0+|I_k|^2}, \tag{4}$$

where Γ_k is the received signal-to-interference-plus-noise ratio (SINR) at user k and $|\cdot|$ denotes the absolute value of a complex-valued scalar.

The *weighted system capacity* is defined as the total average number of bit successfully delivered to the K mobile users and is given by

$$U(\mathcal{P},\mathcal{F},\mathcal{S}) = \sum_{k=1}^{K} w_k s_k C_k, \tag{5}$$

where $\mathcal{P}$, $\mathcal{F}$, and $\mathcal{S}$ are the power, precoding coefficient, and user selection policies, respectively. w_k is a positive constant provided by the upper layers, which allows the resource allocator to prioritize different users for the sake of fairness. On the other hand, we model the power dissipation in the system as the sum of two dynamic terms which can be expressed as [4]

$$U_{TP}(\mathcal{P},\mathcal{F},\mathcal{S}) = N_T\times P_C + \sum_{k=1}^{K}\varepsilon P_k s_k + P_0, \tag{6}$$

where P_C is a constant circuit chain power consumption required in each transmit antenna. P_0 is the basic power consumed independent of the number of transmit antennas. The first term in (6) represents the total *power consumption* of all antennas which includes the power dissipations in the transmit filter, mixer, frequency synthesizer, and digital-to-analog converter. The second term denotes the total power consumption of the power amplifier at the BS. $\varepsilon \geq 1$ is a constant which accounts for the inefficiency in the power amplifier. For example, if $\varepsilon = 5$, it means that for every 10 Watt of power radiated in the RF, 50 Watt are consumed in the power amplifier and the power efficiency is $\frac{1}{\varepsilon}=\frac{1}{5}=20\%$.

Hence, the *energy efficiency* of the considered system is defined as the total average number of bits/Joule

$$U_{eff}(\mathcal{P},\mathcal{F},\mathcal{S}) = \frac{U(\mathcal{P},\mathcal{F},\mathcal{S})}{U_{TP}(\mathcal{P},\mathcal{F},\mathcal{S})}. \tag{7}$$

B. Optimization Problem Formulation

The optimal resource allocation policies $(\mathcal{P}^*,\mathcal{F}^*,\mathcal{S}^*)$ can be obtained by solving

$$\begin{aligned} &\max_{\mathcal{P},\mathcal{F},\mathcal{S}}\; U_{eff}(\mathcal{P},\mathcal{F},\mathcal{S}) \\ &\text{s.t. C1: }\sum_{k=1}^{K} s_k C_k \geq r, \\ &\quad \text{C2: }\sum_{k=1}^{K} P_k s_k \leq P_T, \quad \text{C3: }\sum_{k=1}^{K} s_k \leq N_T, \\ &\quad \text{C4: } P_k \geq 0,\ \forall k, \qquad \text{C5: } s_k = \{0,1\},\ \forall k. \end{aligned} \tag{8}$$

Here, C1 specifies the minimum system data rate requirement r. C2 is a transmit power constraint for the BS in the downlink. C3 puts a limit to the number of users which can be served by the BS in each transmission. C5 is imposed to guarantee each user can at most use the channel once. In other words, the BS is not allowed to multiplex different messages to the same user, since this would require a sophisticated receiver at the user, such as a successive interference cancellation receiver, for recovering all messages. C4 are the boundary constraints for the power allocation variables.

IV. Solution of the Optimization Problem

The objective function in (8) is a non-convex function and a brute force approach is required for obtaining a global optimal solution. However, such a method has exponential complexity with respect to (w.r.t.) the number of users which is computationally infeasible even for small size systems. In order to derive an efficient resource allocation algorithm, we introduce the following transformation.

A. Transformation of the Objective Function

The objective function in (8) can be classified as s nonlinear fractional program [7]. For the sake of notational simplicity, we define Θ as the set of feasible solutions of the optimization problem in (8). Without loss of generality, we define the maximum energy efficiency q^* of the considered system as

$$\begin{aligned} q^* &= \frac{U(\mathcal{P}^*,\mathcal{F}^*,\mathcal{S}^*)}{U_{TP}(\mathcal{P}^*,\mathcal{F}^*,\mathcal{S}^*)} \\ &= \max_{\mathcal{P},\mathcal{F},\mathcal{S}} \frac{U(\mathcal{P},\mathcal{F},\mathcal{S})}{U_{TP}(\mathcal{P},\mathcal{F},\mathcal{S})},\ \forall\{\mathcal{P},\mathcal{F},\mathcal{S}\}\in\Theta. \end{aligned} \tag{9}$$

We are now ready to introduce the following Theorem.

Theorem 1: The maximum energy efficiency q^* is achieved *if and only if*

$$\begin{aligned} &\max_{\mathcal{P},\mathcal{F},\mathcal{S}} \quad U(\mathcal{P},\mathcal{F},\mathcal{S}) - q^* U_{TP}(\mathcal{P},\mathcal{F},\mathcal{S}) \\ &\quad = \quad U(\mathcal{P}^*,\mathcal{F}^*,\mathcal{S}^*) - q^* U_{TP}(\mathcal{P}^*,\mathcal{F}^*,\mathcal{S}^*) = 0, \end{aligned} \tag{10}$$

for $U(\mathcal{P},\mathcal{F},\mathcal{S})\geq 0$ and $U_{TP}(\mathcal{P},\mathcal{F},\mathcal{S})>0$.

Proof: Please refer to Appendix A for a proof of Theorem 1.

Theorem 1 reveals that for an optimization problem with an objective function in fractional form, there exists an equivalent[1] objective function in subtractive form, e.g. $U(\mathcal{P},\mathcal{F},\mathcal{S}) - q^* U_{TP}(\mathcal{P},\mathcal{F},\mathcal{S})$ in the considered case. As a result, we can focus on the equivalent objective function in the rest of the paper.

[1] Here, "equivalent" means both problem formulations will lead to the same resource allocation policies.

TABLE I

ITERATIVE RESOURCE ALLOCATION ALGORITHM.

Algorithm 1 Iterative Resource Allocation Algorithm

Initialize the maximum number of iterations L_{max} and the maximum tolerance ϵ
Set maximum energy efficiency $q = 0$ and iteration index $n = 0$
repeat {Main Loop}
  Solve the inner loop problem in (11) for a given q and obtain resource allocation policies $\{\mathcal{P}', \mathcal{F}', \mathcal{S}'\}$
  if $U(\mathcal{P}', \mathcal{F}', \mathcal{S}') - qU_{TP}(\mathcal{P}', \mathcal{F}', \mathcal{S}') < \epsilon$ **then**
    Convergence $=$ **true**
    return $\{\mathcal{P}^*, \mathcal{F}^*, \mathcal{S}^*\} = \{\mathcal{P}', \mathcal{F}', \mathcal{S}'\}$ and $q^* = \frac{U(\mathcal{P}', \mathcal{F}', \mathcal{S}')}{U_{TP}(\mathcal{P}', \mathcal{F}', \mathcal{S}')}$
  else
    Set $q = \frac{U(\mathcal{P}', \mathcal{F}', \mathcal{S}')}{U_{TP}(\mathcal{P}', \mathcal{F}', \mathcal{S}')}$ and $n = n + 1$
    Convergence $=$ **false**
  end if
until Convergence $=$ **true** or $n = L_{max}$

B. Iterative Algorithm for Energy Efficiency Maximization

In this section, we propose an iterative algorithm (known as the Dinkelbach method [7]) for solving (8) with an equivalent objective function. The proposed algorithm is summarized in Table I and the convergence to optimal energy efficiency is guaranteed.

Proof: Please refer to Appendix B for the proof of convergence.

As shown in Table I, in each iteration in the main loop, we solve the following optimization problem for a given parameter q:

$$\max_{\mathcal{P}, \mathcal{F}, \mathcal{S}} \quad U(\mathcal{P}, \mathcal{F}, \mathcal{S}) - qU_{TP}(\mathcal{P}, \mathcal{F}, \mathcal{S}) \quad \text{s.t. C1, C2, C3, C4, C5.} \tag{11}$$

The transformed problem is a mixed combinatorial and non-convex optimization problem. The non-convex nature comes from the power allocation variables and precoding coefficients. The multiuser interference appears in the denominator of the capacity equation in (3) which couples the power allocation variables. On the other hand, the combinatorial nature comes from the integer constraint for user selection. To obtain an optimal solution, an exhaustive search is needed with complexity $\sum_{a=1}^{N_T} \binom{K}{a}$ for $K \geq N_T$ or $\sum_{g=1}^{K} \binom{N_T}{g}$ for $N_T > K$, which is computationally infeasible for $N_T, K \gg 1$. In order to derive an efficient resource allocation algorithm, we solve the above problem in two steps by fixing the resource allocation policies $\{\mathcal{F}, \mathcal{S}\}$. In the first step, we employ a low complexity sub-optimal user selection scheme. Then, in the second step, we use the closed-form power allocation for a given selected user set $\mathcal{S}$ with zero-forcing beamforming (ZFBF) precoding. Note that by fixing resource allocation policies $\{\mathcal{F}, \mathcal{S}\}$, Algorithm 1 in Table I converges to a sub-optimal solution since only the power allocation is optimized for energy efficiency maximization.

Step 1 (δ-Orthogonal User Selection): We propose an efficient user selection algorithm. Without loss of generality, we define a column vector $\mathbf{\Upsilon}_k = \mathbf{h}_k \sqrt{g_k l_k}$. Let $\Delta\Big(\mathbf{\Upsilon}_k, \mathbf{\Upsilon}_j\Big) = \frac{|\mathbf{\Upsilon}_j^\dagger \mathbf{\Upsilon}_k|}{\|\mathbf{\Upsilon}_j\| \|\mathbf{\Upsilon}_k\|}$ where $\|\cdot\|$ denotes the Euclidean norm of a vector. Then, a $\delta-$ orthogonal user set, $\mathcal{S}_\perp$, is given by

$$\begin{aligned} \mathcal{S}_\perp \;=\; & \Big\{k, j \Big| k, j = \{1, \ldots, K\}, k = \arg\max_t \|\mathbf{\Upsilon}_t\|^2, \\ & \Delta\Big(\mathbf{\Upsilon}_k, \mathbf{\Upsilon}_j\Big) \leq \delta \times w_j, \forall j \neq k\Big\}, \end{aligned} \tag{12}$$

where δ and $[\cdot]^\dagger$ represent a threshold for measuring orthogonality and the conjugate transpose operation, respectively. $k = \arg\max_t \|\mathbf{\Upsilon}_t\|^2$ represents the user who has the largest channel gain for joint BS transmission and is able to tolerate strong interference due to spatial reuse. In other words, we first select the strongest user and then perform user selection by selecting at most the $N_T - 1$ elements in set $S_\perp$ with small values of $\Delta\Big(\mathbf{\Upsilon}_k, \mathbf{\Upsilon}_j\Big)$, since those users introduce less interference to the strongest user. Note that the search space of user selection decreases from $\sum_{a=1}^{N_T} \binom{K}{a}$ and $\sum_{g=1}^{K} \binom{N_T}{g}$ to $2K - 1$ for $K \geq N_T$ and $N_T > K$, respectively. Note that a user with higher value of w_k (priority) has a higher chance of being selected. Note that although the proposed algorithm can only guarantee a δ *orthogonality* between the strongest user and each other selected user, it has been shown that the proposed scheme performs well with the following zero-forcing beamforming scheme [8].

Step 2 (Zero-Forcing Beamformimg): The considered system can be categorized as a MIMO broadcast channel. Hence, dirty paper coding (DPC) is optimal in achieving the multiuser broadcast capacity region. Yet, DPC requires a very high complexity which is considered impractical. On the contrary, although ZFBF is a suboptimal precoding scheme, it has been considered as a practical precoding solution due to its linear complexity and promising performance. Therefore, we focus on ZFBF in the rest of the paper.

Since ZFBF is used for transmission, the capacity equation in (3) can be rewritten as

$$C_k = W \log_2 \Big(1 + \Gamma_k\Big) \text{ with } \Gamma_k = \frac{P_k l_k g_k |\mathbf{h}_k^T \mathbf{f}_k|^2}{W N_0}. \tag{13}$$

On the other hand, without loss of generality, we assume that user 1 to user k are selected by searching the orthogonal set $\mathcal{S}_\perp$. Define the set of selected users as $\widetilde{\mathcal{S}}_\perp \subset \mathcal{S}_\perp$ where $|\widetilde{\mathcal{S}}_\perp| \leq N_T$. Then, we define a super channel matrix $\mathbf{H} \in \mathbb{C}^{|\widetilde{\mathcal{S}}_\perp| \times N_T}$ such that

$$\mathbf{H}^T = \big[\mathbf{\Upsilon}_1 \, \mathbf{\Upsilon}_2 \, \ldots \, \mathbf{\Upsilon}_k\big]. \tag{14}$$

The corresponding ZFBF super matrix $\mathbf{B} \in \mathbb{C}^{N_T \times |\widetilde{\mathcal{S}}_\perp|}$ can be calculated in the BS and is given by

$$\mathbf{B} = \mathbf{H}^\dagger \Big(\mathbf{H}\mathbf{H}^\dagger\Big)^{-1} \mathbf{D}, \tag{15}$$

where $\mathbf{D} \in \mathbb{C}^{|\widetilde{\mathcal{S}}_\perp| \times |\widetilde{\mathcal{S}}_\perp|}$ is a diagonal matrix with diagonal elements $\gamma_k = 1 / \sqrt{\Big[\Big(\mathbf{H}\mathbf{H}^\dagger\Big)^{-1}\Big]_{k,k}} = \sqrt{l_k g_k} \Big|\mathbf{h}_k^T \mathbf{f}_k\Big|$. Here, operator $\Big[\cdot\Big]_{a,b}$ refers to the element in row a and column b of a matrix. Note that γ_k represents the equivalent channel gain between the BS and user k. Hence, the ZFBF vector $\mathbf{f}_k$ for user k is given by

$$\mathbf{f}_k = \Big[\mathbf{B}\Big]_{:,k}, \tag{16}$$

where operator $\Big[\cdot\Big]_{:,b}$ refers to column b of a matrix.

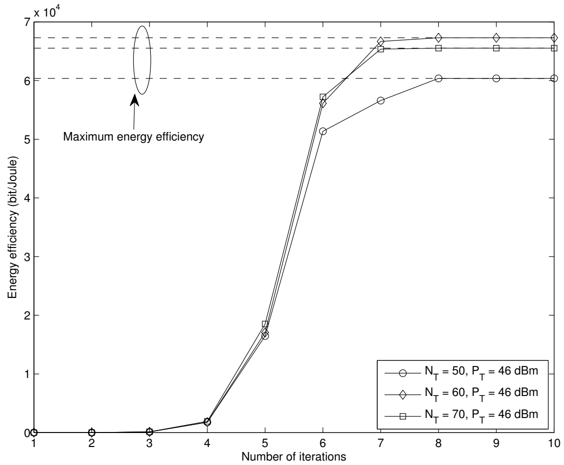

Fig. 1. Energy efficiency (bit-per-Joule) versus number of iterations with different numbers of antennas N_T and $K = 60$ users for maximum transmit power $P_T = 46$ dBm. The dashed lines represent the maximum achievable energy efficiencies for different cases.

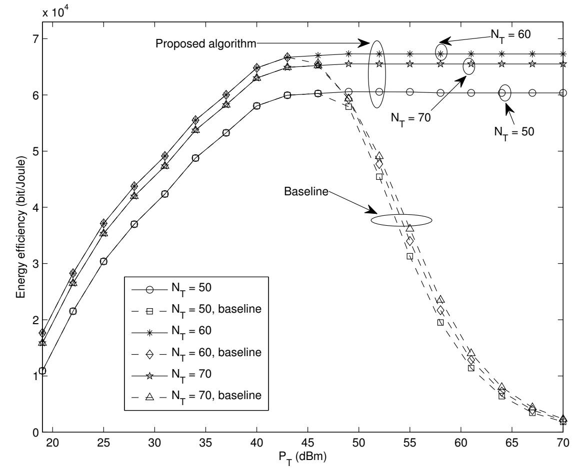

Fig. 2. Energy efficiency (bit-per-Joule) versus maximum transmit power, P_T, for $K = 60$ users.

Power Allocation Solution: The power optimization variables are concave w.r.t. the objective function for a given user selection set and a given ZFBF precoding coefficient set. So standard optimization techniques are applicable in solving the optimal power allocation. For this purpose, we first need the Lagrangian function of (11). Upon rearranging terms, the Lagrangian can be written as

$$\begin{aligned}\mathcal{L}(\lambda,\zeta) &= \sum_{k\in\tilde{\mathcal{S}}_\perp}(w_k+\zeta)C_k - \lambda\Big(\sum_{k\in\tilde{\mathcal{S}}_\perp}P_k - P_T\Big) - r\zeta \\ &\quad -q\Big(N_T\times P_C + \sum_{k\in\tilde{\mathcal{S}}_\perp}\varepsilon P_k + P_0\Big), \qquad (17)\end{aligned}$$

where λ and ζ are the Lagrange multipliers chosen to satisfy the BS power constraint C2 and data rate requirement C1 in (8), respectively. Note that the boundary constraint C4 will be absorbed into the Karush-Kuhn-Tucker (KKT) conditions when deriving the optimal solution in the following. So, by KKT conditions, the closed-form power allocation for the BS to serve user k for a given parameter q is obtained as

$$\frac{\partial\mathcal{L}(\lambda,\zeta)}{\partial P_k} = 0, \qquad (18)$$

$$\Rightarrow P_k = \left[\frac{W(w_k+\zeta)}{(q\varepsilon+\lambda)\ln(2)} - \frac{N_0W}{|\gamma_k|^2}\right]^+, \qquad (19)$$

where $[x]^+ = \max\{0, x\}$. It can be observed that variable w_k (provided by the MAC layer) affects the power allocation by changing the water-level. In other words, high priority users will be allocated higher transmit powers, compared to low priority users. The optimal values of λ and ζ can be easily found by using numerical methods such as the gradient method or the bisection method, due to the concavity of the transformed problem with respect to the power allocation variables.

V. Results

In this section, we evaluate the system performance through simulations. A single cell with a radius of 1 km is considered. The carrier center frequency is 2.5 GHz and the bandwidth is $W = 200$ kHz. We assume a noise power of $N_0W = -123$ dBm and $w_k = 1\,\forall k$. The K desired users are uniformly distributed between the reference distance and the cell boundary at 1 km. The 3rd Generation Partnership Project (3GPP) path loss model is used with a reference distance of $d_0 = 35$ m and log-normal shadowing with a standard deviation of 8 dB. The small scale fading coefficients of the BS-to-user link are modeled as independent and identically distributed (i.i.d.) Rayleigh random variables with zero means and unit variances. The average system capacity is obtained by counting the number of packets successfully decoded by the users averaged over both the macroscopic and microscopic fading. We assume a static circuit power consumption of $P_0 = 40$ dBm [9], a minimum data rate requirement[2] of $r = 21.1$ Megabit/s, and an orthogonality parameter of $\delta = 0.1$. $P_C = 41$ dBm denotes the power dissipation incurred by each antenna for transmission [10]. In practice, the value of P_C depends on the application-specific integrated circuit (ASIC) and the chosen implementation algorithms. On the other hand, we assume a power efficiency of 20% in the RF power amplifier. i.e., $\varepsilon = \frac{1}{0.2} = 5$. Note that if the resource allocator is unable to guarantee the minimum data rate r in a time slot, we set the energy efficiency and average capacity in that particular time slot to zero to account for the corresponding failure.

A. Convergence of Iterative Algorithm

Figure 1 illustrates the evolution of the proposed iterative algorithm for different numbers of transmit antennas N_T, a maximum transmit power of $P_T = 46$ dBm at the BS, and $K = 60$ users. The results in Figure 1 were averaged over 100000 independent adaptation processes where each adaptation process involves different realizations for path loss, shadowing, and multipath fading. It can be observed that the iterative algorithm converges to the optimal value[3] within 10 iterations for all considered numbers of transmit antennas. In

[2]Note that 21.1 Megabit/s is the *maximum* data rate for category 13 in 3GPP Release 7.

[3]Here, the optimality is with respect to the optimization of power allocation given a selected user set and ZFBF transmission.

other words, the maximum system energy efficiency can be achieved within a few iterations on average.

B. Energy Efficiency, Average Capacity, and Average Total Power Consumption versus Transmit Power

Figure 2 illustrates the energy efficiency versus the total transmit power for $K = 60$ users. The number of iterations for the proposed iterative resource allocation algorithm is 10. It can be observed that energy efficiency first increases and then decreases with an increasing number of antennas. This is because when the number of antennas is small, i.e., $N_T < K$, a large throughput gain can be achieved by multiplexing the messages of more users in the same channel via additional antennas. However, when the number of antennas is large enough, i.e., $N_T > K$, the performance gain due to the users multiplexing is saturated. In the meantime, the power consumption per antenna increases linearly with the number of antennas, which degrades the energy efficiency. Besides, the energy efficiency of the proposed algorithm approaches a constant value in the high transmit power regime, since the resource allocator is not willing to consume more power when the maximum energy efficiency is achieved.

Figure 2 also contains the energy efficiency of a baseline resource allocation scheme. For the baseline scheme, we maximize the average system capacity (bit/s/Hz) with constraints C1-C5 in (8), instead of the energy efficiency. The optimal resource allocation polices for the baseline scheme can be obtained by using the traditional water-filling approach. It can be observed that the proposed algorithm provides a significant performance gain in terms of energy efficiency over the baseline scheme, especially in the high transmit power regime. This is because the latter scheme uses excess power to increase the system capacity by sacrificing the energy efficiency.

Figure 3 shows the average capacity versus maximum transmit power P_T for $K = 60$ users. We compare again the system performance of the proposed algorithm with the the baseline resource allocator. The number of iterations in the proposed algorithm is set to 10. It can be observed that the average capacity of the proposed algorithm approaches a constant in the high transmit power regime. This is because the proposed algorithm clips the transmit power at the BS in order to maximize the system energy efficiency. We note that, as expected, the baseline scheme resource allocator achieves a higher average capacity than the proposed algorithm in the high transmit power regime, since the baseline scheme is always transmitting with full power. However, the superior average capacity of the baseline scheme comes at the expense of low energy efficiencies. On the other hand, an increasing number of antennas in the baseline scheme benefits the average capacity due to an improved beamforming gain.

Figure 4 depicts the average total power consumption, i.e., $\mathcal{E}\{U_{TP}(\mathcal{P}, \mathcal{F}, \mathcal{S})\}$, versus maximum transmit power P_T for the proposed algorithm and the baseline scheme with 10 iterations, where $\mathcal{E}\{\cdot\}$ denotes a statistical expectation. In the low transmit power regime, the baseline scheme consumes the same amount of average power as the proposed algorithm which suggests that full power transmission is optimal. However, as the maximum transmit power allowance P_T increases, the proposed algorithm stops to further consume more power since maximum energy efficiency is achieved. On the other hand, the baseline still consumes all the available power and results in a huge power consumption.

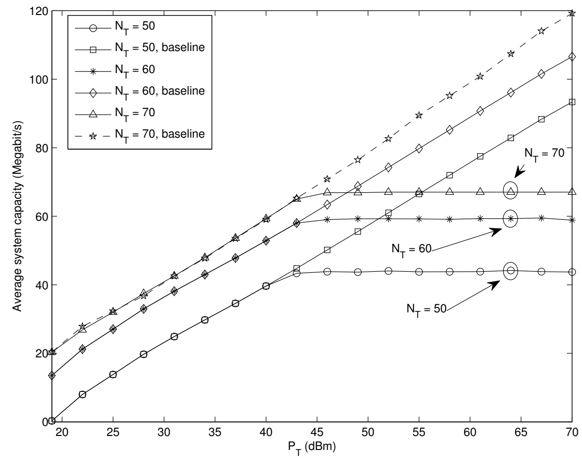

Fig. 3. Average capacity (Megabit/s) versus maximum transmit power, P_T, and $K = 60$ users for the proposed algorithm and the baseline.

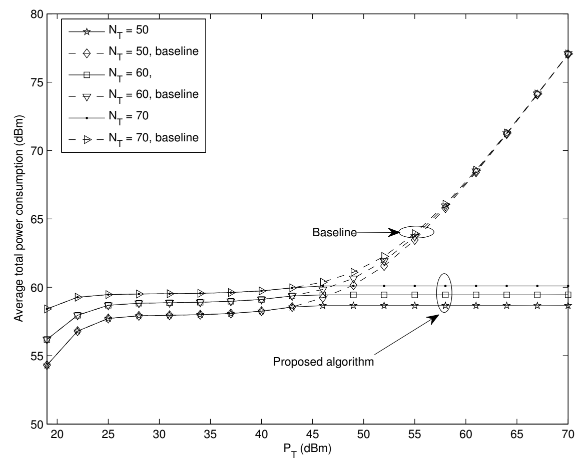

Fig. 4. Average total power consumption, $\mathcal{E}\{U_{TP}(\mathcal{P}, \mathcal{F}, \mathcal{S})\}$, versus maximum transmit power, P_T, and $K = 60$ users for the proposed algorithm and the baseline.

VI. Conclusion

In this paper, we formulated the resource allocation and scheduling design in SDMA networks as a non-convex and combinatorial optimization problem, in which the circuit power dissipation and the system data rate requirement were taken into consideration. By exploiting the properties of fractional programming, the considered problem was transformed into an equivalent problem with a tractable solution. An efficient iterative resource allocation algorithm with closed-form power allocation and low complexity user selection was derived for maximization of the energy efficiency. Simulation results did not only show that the proposed algorithm converges to the optimal solution within a small number of iterations, but demonstrated also the achievable maximum energy efficiency.

Interesting topics for future work include studying the impact of imperfect channel state information and optimizing the number of antennas.

APPENDIX

A. Proof of Theorem 1

We now prove the forward implication of Theorem 1 by following a similar approach as in [7]. Without loss of generality, we define q^* and $\{\mathcal{P}^*, \mathcal{F}^*, \mathcal{S}^*\} \in \Theta$ as the optimal energy efficiency and the optimal resource allocation policy of the *original objective function* in (8), respectively. Then, the optimal energy efficiency can be expressed as

$$q^* = \frac{U(\mathcal{P}^*, \mathcal{F}^*, \mathcal{S}^*)}{U_{TP}(\mathcal{P}^*, \mathcal{F}^*, \mathcal{S}^*)} \geq \frac{U(\mathcal{P}, \mathcal{F}, \mathcal{S})}{U_{TP}(\mathcal{P}, \mathcal{F}, \mathcal{S})}, \forall \{\mathcal{P}, \mathcal{F}, \mathcal{S}\} \in \mathcal{F},$$
$$\Longrightarrow U(\mathcal{P}, \mathcal{F}, \mathcal{S}) - q^* U_{TP}(\mathcal{P}, \mathcal{F}, \mathcal{S}) \leq 0 \text{ and}$$
$$U(\mathcal{P}^*, \mathcal{F}^*, \mathcal{S}^*) - q^* U_{TP}(\mathcal{P}^*, \mathcal{F}^*, \mathcal{S}^*) = 0. \tag{20}$$

Therefore, we conclude that $\max_{\mathcal{P},\mathcal{F},\mathcal{S}} U(\mathcal{P}, \mathcal{F}, \mathcal{S}) - q^* U(\mathcal{P}, \mathcal{F}, \mathcal{S}) = 0$, which is achievable by resource allocation policy $\{\mathcal{P}^*, \mathcal{F}^*, \mathcal{S}^*\}$. This completes the forward implication.

Next, we prove the converse implication of Theorem 1. Suppose $\{\mathcal{P}_e^*, \mathcal{F}_e^*, \mathcal{S}_e^*\}$ is the optimal resource allocation policy of the *equivalent objective function* such that $U(\mathcal{P}_e^*, \mathcal{F}_e^*, \mathcal{S}_e^*) - q^* U_{TP}(\mathcal{P}_e^*, \mathcal{F}_e^*, \mathcal{S}_e^*) = 0$. Then, for any feasible resource allocation policy $\{\mathcal{P}, \mathcal{F}, \mathcal{S}\} \in \Theta$, we can obtain the following inequality

$$\begin{aligned} & U(\mathcal{P}, \mathcal{F}, \mathcal{S}) - q^* U_{TP}(\mathcal{P}, \mathcal{F}, \mathcal{S}) \\ \leq \quad & U(\mathcal{P}_e^*, \mathcal{F}_e^*, \mathcal{S}_e^*) - q^* U_{TP}(\mathcal{P}_e^*, \mathcal{F}_e^*, \mathcal{S}_e^*) = 0. \end{aligned} \tag{21}$$

The above inequality implies

$$\begin{aligned} \frac{U(\mathcal{P}, \mathcal{F}, \mathcal{S})}{U_{TP}(\mathcal{P}, \mathcal{F}, \mathcal{S})} &\leq q^* \quad \forall \{\mathcal{P}, \mathcal{F}, \mathcal{S}\} \in \mathcal{F} \quad \text{and} \\ \frac{U(\mathcal{P}_e^*, \mathcal{F}_e^*, \mathcal{S}_e^*)}{U_{TP}(\mathcal{P}_e^*, \mathcal{F}_e^*, \mathcal{S}_e^*)} &= q^*. \end{aligned} \tag{22}$$

In other words, the optimal resource allocation policy $\{\mathcal{P}_e^*, \mathcal{F}_e^*, \mathcal{S}_e^*\}$ for the equivalent objective function is also the optimal resource allocation policy for the original objective function.

This completes the proof of the converse implication of Theorem 1. In summary, the optimization of the original objective function and the optimization of the equivalent objective function result in the same resource allocation policy. $\square$

B. Proof of Algorithm Convergence

We follow a similar approach as in [7] for proving the convergence of Algorithm 1. We first introduce the following two propositions. For the sake of notational simplicity, we define the equivalent objective function in (11) as $F(q') = \max_{\mathcal{P},\mathcal{F},\mathcal{S}} \{U(\mathcal{P}, \mathcal{F}, \mathcal{S}) - q' U_{TP}(\mathcal{P}, \mathcal{F}, \mathcal{S})\}$.

Proposition 1: $F(q')$ is a strictly monotonic decreasing function in q', i.e., $F(q'') > F(q')$ if $q' > q''$.

Proof: Let $\{\mathcal{P}', \mathcal{F}', \mathcal{S}'\} \in \mathcal{F}$ and $\{\mathcal{P}'', \mathcal{F}'', \mathcal{S}''\} \in \mathcal{F}$ be the two distinct optimal resource allocation polices for $F(q')$ and $F(q'')$, respectively.

$$\begin{aligned} F(q'') &= \max_{\mathcal{P},\mathcal{F},\mathcal{S}} \{U(\mathcal{P}, \mathcal{F}, \mathcal{S}) - q'' U_{TP}(\mathcal{P}, \mathcal{F}, \mathcal{S})\} \quad (23) \\ &> U(\mathcal{P}', \mathcal{F}', \mathcal{S}') - q'' U_{TP}(\mathcal{P}', \mathcal{F}', \mathcal{S}') \\ &\geq U(\mathcal{P}', \mathcal{F}', \mathcal{S}') - q' U_{TP}(\mathcal{P}', \mathcal{F}', \mathcal{S}') \\ &= F(q'). \end{aligned}$$

$\square$

Proposition 2: Let $\{\mathcal{P}', \mathcal{F}', \mathcal{S}'\} \in \mathcal{F}$ be an arbitrary feasible solution and $q' = \frac{U(\mathcal{P}', \mathcal{F}', \mathcal{S}')}{U_{TP}(\mathcal{P}', \mathcal{F}', \mathcal{S}')}$, then $F(q') \geq 0$.

Proof:
$$\begin{aligned} F(q') &= \max_{\mathcal{P},\mathcal{F},\mathcal{S}} \{U(\mathcal{P}, \mathcal{F}, \mathcal{S}) - q' U_{TP}(\mathcal{P}, \mathcal{F}, \mathcal{S})\} \\ &\geq U(\mathcal{P}', \mathcal{F}', \mathcal{S}') - q' U_{TP}(\mathcal{P}', \mathcal{F}', \mathcal{S}') = 0. \end{aligned}$$

$\square$

We are now ready to prove the convergence of Algorithm 1.

Proof of Convergence: We first prove that the energy efficiency q increases in each iteration. Then, we prove that if the number of iterations is large enough, the energy efficiency q converges to the optimal q^* such that it satisfies the optimality condition in Theorem 1, i.e., $F(q^*) = 0$.

Let $\{\mathcal{P}_n, \mathcal{F}_n, \mathcal{S}_n\}$ be the optimal resource allocation policy in the n-th iteration. Suppose $q_n \neq q^*$ and $q_{n+1} \neq q^*$ represent the energy efficiency of the considered system in iterations n and $n+1$, respectively. By Theorem 1 and Proposition 2, $F(q_n) > 0$ and $F(q_{n+1}) > 0$ must be true. On the other hand, in the proposed algorithm, we calculate q_{n+1} as $q_{n+1} = \frac{U(\mathcal{P}_n, \mathcal{F}_n, \mathcal{S}_n)}{U_{TP}(\mathcal{P}_n, \mathcal{F}_n, \mathcal{S}_n)}$. Thus, we can express $F(q_n)$ as

$$\begin{aligned} F(q_n) &= U(\mathcal{P}_n, \mathcal{F}_n, \mathcal{S}_n) - q_n U_{TP}(\mathcal{P}_n, \mathcal{F}_n, \mathcal{S}_n) \\ &= U_{TP}(\mathcal{P}_n, \mathcal{F}_n, \mathcal{S}_n)(q_{n+1} - q_n) > 0 \quad (24) \\ \Longrightarrow q_{n+1} &> q_n, \qquad \because U_{TP}(\mathcal{P}_n, \mathcal{F}_n, \mathcal{S}_n) > 0. \quad (25) \end{aligned}$$

By combining $q_{n+1} > q_n$, Proposition 1, and Proposition 2, we can show that as long as the number of iterations is large enough, $F(q_n)$ will eventually approach zero and satisfy the optimality condition as stated in Theorem 1. $\square$

REFERENCES

[1] D. Aktas, M. Bacha, J. Evans, and S. Hanly, "Scaling Results on the Sum Capacity of Cellular Networks with MIMO Links," *IEEE Trans. Inform. Theory*, vol. 52, pp. 3264–3274, 2006.

[2] T. Marzetta, "Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas," *IEEE Trans. Wireless Commun.*, vol. 9, pp. 3590–3600, Nov. 2010.

[3] K. Kumar, G. Caire, and A. Moustakas, "Asymptotic Performance of Linear Receivers in MIMO Fading Channels," *IEEE Trans. Inform. Theory*, vol. 55, pp. 4398–4418, Oct. 2009.

[4] R. Prabhu and B. Daneshrad, "Energy-Efficient Power Loading for a MIMO-SVD System and its Performance in Flat Fading," in *Proc. IEEE Global Telecommun. Conf.*, Dec. 2010, pp. 1–5.

[5] G. Miao, N. Himayat, and Y. Li, "Energy-Efficient Transmission in Frequency-Selective Channels," in *Proc. IEEE Global Telecommun. Conf.*, Dec. 2008, pp. 1–6.

[6] Z. Hasan, G. Bansal, E. Hossain, and V. Bhargava, "Energy-Efficient Power Allocation in OFDM-Based Cognitive Radio Systems: A Risk-Return Model," *IEEE Trans. Wireless Commun.*, vol. 8, pp. 6078–6088, Dec. 2009.

[7] W. Dinkelbach, "On Nonlinear Fractional Programming," *Management Science*, vol. 13, pp. 492–498, Mar. 1967. [Online]. Available: http://www.jstor.org/stable/2627691

[8] T. Yoo and A. Goldsmith, "On the Optimality of Multiantenna Broadcast Scheduling Using Zero-Forcing Beamforming," *IEEE J. Select. Areas Commun.*, vol. 24, pp. 528–541, Mar. 2006.

[9] O. Arnold, F. Richter, G. Fettweis, and O. Blume, "Power Consumption Modeling of Different Base Station Types in Heterogeneous Cellular Networks," in *Proc. Future Network and Mobile Summit*, 2010, pp. 1–8.

[10] R. Kumar and J. Gurugubelli, "How Green the LTE Technology Can Be?" in *Intern. Conf. on Wireless Commun., Veh. Techn., Inform. Theory and Aerosp. Electron. Syst. Techn.*, Mar. 2011.