On the Energy Efficiency-Spectral Efficiency Trade-Off in MIMO-OFDMA Broadcast Channels

Jie Tang, *Member*, *IEEE*, Daniel K. C. So, *Senior Member*, *IEEE*, Emad Alsusa, *Senior Member*, *IEEE*, Khairi Hamdi, *Senior Member*, *IEEE* and Arman Shojaeifard, *Member*, *IEEE*

***Abstract*— This paper investigates the fundamental energy efficiency-spectral efficiency (EE-SE) relationship in a multiple-input multiple-output (MIMO)- orthogonal frequency division multiple access (OFDMA) broadcast channel (BC) with a practical power model considering the power consumption due to the number of admitted users as well as the number of active transmit antennas. However, with this power model, the EE-SE trade-off optimization problem which jointly optimizes the transmit covariance matrices whilst determining the optimal admitted user set and active transmit antenna set is non-convex, and hence it is extremely difficult to solve directly. As a result, we propose an algorithm that decouples the multi-carrier EE optimization problem to a set of single-carrier EE optimization problems. For the single-carrier EE optimization problem, we first investigate the EE-SE trade-off problem with fixed admitted user set and transmit antenna set. Under this setup, we prove that the EE-SE relationship is a quasiconcave function. Furthermore, EE is proved to be either strictly decreasing with SE or first strictly increasing and then strictly decreasing with SE. Based on these findings, we propose a two-layer resource allocation algorithm in order to tackle the comprehensive EE-SE trade-off problem. Meanwhile, since admitting more users and activating more transmit antennas can achieve higher sum-rate but at the cost of larger transmit-independent power consumption, there exists a trade-off between the sum-rate gain and the power consumption. We therefore study the user and antenna selection approach to further explore the optimal trade-off. Both the optimal exhaustive search and the Frobenius norm based dynamic selection schemes are developed to further improve the achievable EE. To further reduce the computational complexity, a strategy that chooses a fixed admitted user set for all the subcarriers is developed. Simulation results confirm the theoretical findings and demonstrate that the proposed resource allocation algorithm can efficiently approach the optimal EE-SE trade-off.**

***Index Terms*— Green radio (GR), multiple-input multiple-output (MIMO), energy efficiency (EE), spectral efficiency (SE).**

I. Introduction

Over the past decades, significant efforts have been directed towards improving the spectral efficiency (SE) of wireless communication systems in order to support the massive increase in network traffic demand. This trend makes SE to be the main performance indicator for the design and optimization of wireless systems, but at the same time constitutes to ever-rising network power consumption which has severe implications in terms of both economic and ecological costs. Green radio (GR) is a recent research direction dedicated to devising novel solutions for tackling the overwhelming capacity crunch in a sustainable and economically viable way. Energy-efficiency (EE), which is an indication of the delivered bits per-unit energy, is widely considered as a first-order design constraint in GR research, and has attracted much interest recently, e.g., single link optimization [1], multi-antenna system [2]–[7], single cell scenario [8]–[12], multi-cell deployment [13] and cognitive radio network [14], [15].

A prominent transmission technology for the next generation of cellular networks is multiuser (MU) multiple-input multiple-output (MIMO). The information-theoretic capacity limit of MIMO broadcast channels (BC) has been extensively studied in the existing literature. In contrast to existing research on SE of MIMO-BC which only considers transmit power constraints, studying the EE of MIMO-BC requires a comprehensive understanding of the power consumption of downlink MU-MIMO systems. A random opportunistic beamforming algorithm has been proposed in [16] to maximize the EE in a broadcast channel. Authors in [17] provided some general analytical tools and insights into optimizing the EE at the BS through power control with elastic traffic. In [18], the authors tackled the EE maximization problem in downlink MIMO systems and extended their work in [19] where a novel optimization approach with transmit covariance optimization and antenna selection scheme is developed for improving the EE in the context of MIMO-BC. In [20], a framework which relies on dirty paper coding (DPC) for MIMO-BC to find the globally optimal energy-efficient solution is proposed.

Joint optimization of EE and SE is not always practically feasible and may even result in conflicts sometimes [21], [22]. Therefore, finding the optimal trade-off between EE and SE is a problem well worth studying. The concept of EE-SE trade-off has first been introduced in [23]; this work has inspired numerous other research activities where the same analytical approach was used to approximate the EE-SE trade-off of correlated multi-antenna [24], multi-user [25] and cooperative [26] communication systems in the low-SE regime. A framework to integrate the connections between EE and SE trade-offs has been proposed in [27]. In [28], EE-SE trade-off considering circuit power was studied for energy-constrained wireless multihop networks with a single source-destination pair. Authors in [29] and [30] provided a tight approximation of the EE-SE trade-off over single-input single-output (SISO) Rayleigh frequency-flat fading channels and MIMO Rayleigh

The authors are with the School of Electrical and Electronics Engineering, The University of Manchester, Manchester, M13 9PL, UK, Emails: {jie.tang, d.so, e.alsusa, k.hamdi, a.shojaeifard}@manchester.ac.uk.

This work has been supported by the Engineering and Physical Sciences Research Council of UK, under grant EP/J021768/1.

fading channel respectively. The authors in [8] investigated the EE-SE relationship in a single-cell downlink OFDMA network and proved that the EE-SE relationship is a quasi-concave function. Based on the findings, they developed a low-complexity suboptimal resource allocation algorithm for practical applications of the EE-SE trade-off. Furthermore, a tight lower bound and a tight upper bound on the EE-SE curve were introduced by Lagrange dual decomposition (LDD) and continuous relaxation, respectively. In [31], the tradeoff between EE and SE is further exploited by balancing consumption power and occupied bandwidth using resource efficiency for downlink cellular network.

A. Main Contributions

While it is evident that the study of this important joint-metric (EE-SE) has gathered pace in recent years, its impact on the MIMO-OFDMA BC has not been investigated. Furthermore, user selection and antenna selection is a widely discussed technology in spectral efficient MIMO systems, but its impact on energy efficient MIMO systems is still an open question. Unlike the EE optimization problem in existing literature that maximizes EE through power control, the EE-SE trade-off optimization problem requires the balancing of EE and SE and results in efficient use of power as well as bandwidth.

In this paper, we investigate the EE-SE trade-off problem in multiuser MIMO-OFDMA BC. A practical power model considering the power consumption due to the number of admitted users as well as the number of active transmit antennas is employed. In terms of SE maximization, admitting all users is always optimal. However, this conclusion does not hold under the energy efficient scenario. As more admitted users achieve higher sum-rate at the cost of higher signal processing power, there exists a trade-off between the power consumption cost and the sum-rate gain. Moreover, more active transmit antennas achieve higher sum-rate at the cost of higher circuit power, and hence there also exists a trade-off between the power consumption cost and the sum-rate gain. As a result, user and antenna selection is necessary. However, with this power model, the EE-SE trade-off optimization problem which jointly optimize the transmit covariance matrices whilst determining the optimal admitted user set and active transmit antenna set is non-convex, and hence it is extremely difficult to solve directly.

To tackle the multi-carrier EE-SE trade-off problem, we propose an efficient algorithm to tackle the EE optimization problem for MIMO-OFDMA system which is close to that of the optimal value. The idea is to decouple the multi-carrier optimization problem to a set of single-carrier optimization problems, and we just need to solve the decoupled single-carrier EE optimization problem and repeat for all the sub-carriers in order to obtain the maximum system EE. For the latter problem, we first investigate the EE-SE trade-off problem with fixed admitted user set and transmit antennas set. Under this setup, we prove that the EE-SE relationship is a quasiconcave function. Furthermore, EE is proved to be either strictly decreasing with SE or first strictly increasing and then strictly decreasing with SE. Based on the quasiconcave property, a two-layer resource allocation algorithm is then proposed to solve EE-SE trade-off problem with fixed admitted user set and transmit antennas set. In particular, an inner-layer is used to find the maximum EE $\lambda_{EE}^{*}(\lambda_{SE})$ for a given SE, λ_{SE}, while an outer-layer is designed to find the optimal EE, λ_{EE}^{opt}, via a gradient based algorithm. With the proposed two-layer solution for the EE-SE trade-off problem, we then study the user and antenna selection approach to further explore the optimal trade-off. A dynamic user and antenna selection approach based on Frobenius norm method are developed. Moreover, in contrast to the proposed dynamic solution where admitted user set is considered for different subcarriers, a selection strategy that chooses a fixed admitted user set for all the subcarriers is developed to reduce the computational complexity.

B. Organization and Notation

The remainder of this paper is organized as follows. The system model and problem formulation is described in Section II. In Section III, we study the fundamental EE-SE relation and develop a resource allocation scheme based on fixed admitted user set and transmit antennas set. In Section IV, we transform the MIMO-BC problem to the dual MIMO-multiple access channel (MAC) problem based on MAC-BC duality. In Section V, a computationally efficient algorithm is proposed to solve the dual MAC optimization problem. In Section VI, we further study the user and antenna selection approach to explore the optimal trade-off curve, both the optimal exhaustive search and the Frobenius norm based dynamic selection schemes are developed to further improve the EE. In Section VII, a low computational complexity based on fixed selection strategy is investigated. Simulation results are provided in Section VIII and conclusions are drawn in Section IX.

The following notations are used in the paper. Bold upper and lower case letters denote matrices and vectors, respectively; $(\cdot)^{-1}$ denotes the matrix inversion, $(\cdot)^{T}$ denotes the matrix transpose, $(\cdot)^{H}$ denote the matrix conjugate transpose, $\mathbf{I}_{N_t \times N_t}$ denotes an $N_t \times N_t$ identity matrix; $E[\cdot]$ denotes the expectation operator; $\mathrm{Tr}(\cdot)$ denotes the trace of a matrix, $[x]^{+}$ denotes $\max(x, 0)$; $(\cdot)^{b}$ and $(\cdot)^{m}$ denote the quantities associated with a broadcast channel and a multiple access channel, respectively.

II. System model and problem formulation

In this section, we introduce the system model of a MIMO-OFDMA BC and mathematically formulate the EE-SE trade-off problem.

A. System Model

The system consists of a single base station (BS) with N_{tot} transmit antennas and K_{tot} users ($K \in \{1, 2, \ldots, K_{tot}\}$) each with N_r receive antennas. An OFDM transmission scheme with M subcarriers ($m = 1, 2, \cdots, M$) on W bandwidth is employed and thus multiuser MIMO-OFDMA transmission scheme is considered in this work. As the number of admitted users in the network and the number of active transmit antennas at the BS have a significant impact on the EE, selecting the set of admitted users and active transmit antennas is important. Moreover, the number of admitted users could be varied through different subcarriers in order to further increase the system's EE, and hence we denote the selected user set on subcarrier m as $\mathcal{K}_m \in \{1, \ldots, K_{tot}\}$ with the number of admitted user $K_m = |\mathcal{K}_m|$. The superset $\bar{\mathcal{K}}$ is used to denote the combined user sets for all subcarriers. On the other hand, transmit antenna selection cannot be performed on a per-subcarrier basis because different subcarriers might request a different set of active antennas, resulting in more antennas being activated for one OFDMA frame. Therefore, the transmit antenna set, once selected, is fixed for all the subcarriers, and is denoted as $\mathcal{N} \in \{1, \ldots, N_{tot}\}$ with the number of active antennas $N_t = |\mathcal{N}|$. With the selected users on the m^{th} subcarrier and the selected antennas, we denote the channel matrix from the BS to the k^{th} user on the m^{th} subcarrier with N_t active antennas as $\mathbf{H}^m_{k,\mathcal{N}} \in \ C^{N_r \times N_t}$. Channel state information (CSI) is assumed to be perfectly known at the corresponding transmitter and receivers. Note that the CSI at the receivers (CSIR) can be obtained from the channel estimation of the downlink pilots. CSI at transmitter (CSIT) can be acquired through uplink feedback in frequency division duplex (FDD) systems or via uplink channel estimation in time division duplex (TDD) systems. The received signal from the BS to the k^{th} user on the m^{th} subcarrier can be written as

$$\mathbf{y}^m_k = \mathbf{H}^m_{k,\mathcal{N}} \mathbf{x}^m_{\mathcal{N}} + \mathbf{n}^m_k, \tag{1}$$

where $\mathbf{n}^m_k \in \ C^{N_r \times 1}$ is the independent zero mean Gaussian noise with each entry $\mathcal{CN}(0, \sigma^2)$, $\mathbf{x}^m_{\mathcal{N}_m}$ is the transmitted signal on the downlink on the m^{th} subcarrier. In addition, $\mathbf{x}^m_{\mathcal{N}} = \mathbf{x}^m_{1,\mathcal{N}} + \mathbf{x}^m_{2,\mathcal{N}} + \ldots + \mathbf{x}^m_{K,\mathcal{N}}$ where $\mathbf{x}^m_{k,\mathcal{N}} \in \ C^{N_t \times 1}$, is the signal transmitted to the k^{th} user on the m^{th} subcarrier with the selected user set ($\mathcal{K}_m$) and antenna set ($\mathcal{N}$), and $\mathbf{x}^m_{k,\mathcal{N}} = \mathbf{W}^m_{k,\mathcal{N}} \mathbf{s}^m_{k,\mathcal{N}}$ where $\mathbf{s}^m_{k,\mathcal{N}} \in \ C^{N_r \times 1}$ is the transmit signal corresponding to the precoding matrix $\mathbf{W}^m_{k,\mathcal{N}} \in \ C^{N_t \times N_r}$.

By employing DPC at the transmitter with antenna set $\mathcal{N}$, and without loss of generality, an encoding order $(1, \cdots, K_m)$ for the user set $\mathcal{K}_m$, i.e., the codeword of user 1 is encoded first, the data rate $R^b_{k,m}$ for the k^{th} user on the m^{th} subcarrier can be written as [32]

$$R^b_{k,m} = W \log \frac{|\mathbf{I}_{N_r \times N_r} + \frac{1}{\sigma^2} \mathbf{H}^m_{k,\mathcal{N}} (\sum_{i=k}^{K_m} \mathbf{Q}^{m,b}_{i,\mathcal{N}}) \mathbf{H}^{mH}_{k,\mathcal{N}}|}{|\mathbf{I}_{N_r \times N_r} + \frac{1}{\sigma^2} \mathbf{H}^m_{k,\mathcal{N}} (\sum_{i=k+1}^{K_m} \mathbf{Q}^{m,b}_{i,\mathcal{N}}) \mathbf{H}^{mH}_{k,\mathcal{N}}|}. \tag{2}$$

where $\mathbf{Q}^{m,b}_{k,\mathcal{N}} = \mathbb{E}(\mathbf{x}^m_{k,\mathcal{N}} \mathbf{x}^{mH}_{k,\mathcal{N}})$ is the corresponding transmit covariance matrix, $\mathbf{Q}^{m,b}_{k,\mathcal{N}} \succeq 0$ i.e., $\mathbf{Q}^{m,b}_{k,\mathcal{N}}$ is a positive semidefinite matrix. Hence, the sum rate of the MIMO-OFDMA BC is given by

$$C_{BC} = \sum_{m=1}^{M} \sum_{k=1}^{K_m} R^b_{k,m}.$$

On the other hand, the sum rate of the MIMO-OFDMA MAC is given by

$$C_{MAC} = \sum_{m=1}^{M} \log |\mathbf{I}_{N_t \times N_t} + \frac{1}{\sigma^2} \sum_{k=1}^{K_m} \mathbf{H}^{mH}_{k,\mathcal{N}} \mathbf{Q}^m_{k,\mathcal{N}} \mathbf{H}^m_{k,\mathcal{N}}|.$$

With regards to the power model, since the BSs are the primary power-hungry component in cellular networks, the users' power consumption is not considered here. Advanced circuit technology has made it possible for wireless transceivers to consume different powers in different operational modes such as sleep, idle, transmit and receive modes [22]. In the transmit/active mode, besides the transmit power, the BS power consumption also includes the consumption by signal processing and active circuit blocks, such as analog-to-digital converter, digital-to-analog converter, synthesizer, and mixer [33]. From [33] and [34], the overall power consumption model for downlink transmission at the BS is

$$P = \zeta P_T + P_{sp} + P_c \tag{3}$$

where ζ and P_T represent the reciprocal of drain efficiency of power amplifier and transmit power at the BS, respectively. P_{sp} denotes the power consumption of signal processing at BS, which depends on baseband processing including the computation of the precoding matrix $\mathbf{W}$ [35]

$$P_{sp} = W P_{sp1} [\dim(\mathbf{W})]^{\beta+1} + W P_{sp2} \tag{4}$$

where the first term of (4) is proportional to the number of active RF chains with order of $\beta \geq 0$, and P_{sp1} is the parameter related to the computation of the precoding matrix $\mathbf{W}$. The active RF-chain number is the same as the dimension (number of columns) of the precoding matrix $\mathbf{W}$. The exponent β implies the overhead power consumption of MU processing compared to single user (SU) processing. For example, if $\beta = 0$, there is no overhead for MU-MIMO signal processing computation. If $\beta > 0$, MU-MIMO signal processing computation consumes relatively higher power than SU signal processing. The maximum of exponent β is assumed to be no greater than two as the computational complexity for l-dimensional MU-MIMO precoding, e.g., zero-forcing (ZF) MU-MIMO precoding, is roughly $\mathcal{O}(l^3)$, while that for SU is $\mathcal{O}(l)$; therefore, $0 \leq \beta \leq 2$ is a reasonable assumption. Moreover, it was shown in [36] that GQRD-based DPC scheme has a similar complexity as SVD-based design which is roughly $\mathcal{O}(l^3)$, and thus β is set to 2 in this work. In the second term of (4), P_{sp2} is the signal processing related power consumption per unit frequency at the baseband module, which is independent of the number of active RF chains.

On the other hand, P_c denotes the circuit power consumption which can be divided into static (fixed) and dynamic parts that depend on the parameters of the active links. Here, the transmission associated circuit power consumption is modeled as a linear function of the number of active antennas and the throughput

$$P_c = P_s + P_{ant} N_t + \gamma C_{BC} \tag{5}$$

where P_s is the static circuit power in transmit mode, $P_{ant} N_t$ denotes the dynamic power consumption proportional to the number of active transmit antennas, and γ is a constant denoting the dynamic power consumption per unit data rate. Hence, we model the overall power consumption as

$$\begin{aligned} P = \zeta P_T + P_{ant} N_t + W P_{sp1} [\dim(\mathbf{W})]^3 + W P_{sp2} \\ + \gamma C_{BC} + P_s \end{aligned} \tag{6}$$

which is dependent on the transmission power, number of active antennas, number of admitted users, rate related power consumption and static circuit power.

B. Problem Formulation

Conventional EE for downlink transmission is defined as the total number of delivered bits per unit energy, where energy consumption includes transmission energy consumption and circuit energy consumption in active mode. In this paper we consider the DPC that achieves the sum-rate capacity for MIMO-OFDMA BC. It is important to note that there exists a performance gap between the capacity and the actual rate achieved by the cellular networks in practice due to practical constraints such as acquiring of CSIT/CSIR, overhead of pilots, practical coding and modulation schemes, practical demodulation and decoding algorithms, etc. Nevertheless, the sum rate capacity achieved by DPC under perfect CSIT/CSIR is the information-theoretic upper bound for MIMO-OFDMA BC, which helps to reveal the theoretical achievable limits, and thus is employed in this paper.

It has been shown that the optimal EE is achieved by transmitting with either full bandwidth with minimum transmit power or full transmit power with minimum bandwidth [37]. However in MIMO-OFDMA, transmitting at full power with minimum bandwidth (i.e., one subcarrier) is not feasible as it requires a very high rate modulation scheme to approach the achievable rate. As a result, activating all available subcarriers in MIMO-OFDMA (i.e., full bandwidth are occupied) is the optimal in terms of EE. With the selected users and transmit antennas, EE in MIMO-OFDMA BC is defined as follows

$$\lambda_{EE} \triangleq \frac{C_{BC}}{P} = \frac{\sum_{m=1}^{M}\sum_{k=1}^{K_m} R_{k,m}^{b}}{P}. \tag{7}$$

On the other hand, SE is defined as the total number of delivered bits per unit bandwidth.

$$\lambda_{SE} \triangleq \frac{C_{BC}}{W} = \frac{\sum_{m=1}^{M}\sum_{k=1}^{K_m} R_{k,m}^{b}}{W}. \tag{8}$$

Since our practical power model considers signal processing power and circuit power, the number of admitted users and active transmit antennas plays an important role in system's EE. Hence, we need to select the suitable users and transmit antennas to exploit the trade-off between the power consumption cost and the sum-rate gain. Furthermore, since the objective of this paper is to maximize EE of MIMO-OFDMA BC whilst achieving a desirable SE, it is reasonable to impose a minimum SE requirement for the optimization problem. Based on the sum-rate expression in (3), the total power consumption model in (6) and noting that $P_T = \sum_{m=1}^{M}\sum_{k=1}^{K_m} \mathrm{Tr}(\mathbf{Q}_{k,\mathcal{N}}^{m,b})$, the optimization problem can be formulated as

$$\max_{\bar{\mathcal{K}},\mathcal{N},\{\mathbf{Q}_{k,\mathcal{N}}^{m,b}\succeq 0\}} \frac{C_{BC}(\mathbf{H}_{k,\mathcal{N}}^{m}, \mathbf{Q}_{k,\mathcal{N}}^{m,b})}{P} \tag{9}$$

$$\text{s.t.} \quad \sum_{m=1}^{M}\sum_{k=1}^{K_m} \mathrm{Tr}(\mathbf{Q}_{k,\mathcal{N}}^{m,b}) \leq P_{max}, \tag{10}$$

$$\frac{C_{BC}(\mathbf{H}_{k,\mathcal{N}}^{m}, \mathbf{Q}_{k,\mathcal{N}}^{m,b})}{W} \geq \lambda_{SE(\min)}, \tag{11}$$

where $P = \zeta P_T + P_{ant}N_t + WP_{sp1}[\dim(\mathbf{W})]^3 + WP_{sp2} + \gamma C_{BC} + P_s$, P_{max} and $\lambda_{SE(\min)}$ are the maximum total transmit power constraint at the BS and the minimum SE requirement, respectively.

Since both the admitted user set $\bar{\mathcal{K}}$ and the active transmit antenna set $\mathcal{N}$, affect the EE in a comprehensive manner, i.e., $\bar{\mathcal{K}}$ and $\mathcal{N}$ are related to both channel matrices and the dynamic power consumption, solving $\bar{\mathcal{K}}$ and $\mathcal{N}$ jointly with $\mathbf{Q}_{k,\mathcal{N}}^{m,b}$ for all the subcarriers is not straightforward, and hence the solution of the above problem is very hard due to the non-concavity of the objective function.

Therefore, we propose an efficient suboptimal algorithm to tackle the EE optimization problem for MIMO-OFDMA system in (9)-(11) by decoupling the multi-carrier optimization problem to a set of single-carrier optimization problems as follows

$$\max_{\mathcal{K}_m,\mathcal{N},\{\mathbf{Q}_{k,\mathcal{N}}^{m,b}\succeq 0\}} \frac{C_{BC}^{m}(\mathbf{H}_{k,\mathcal{N}}^{m}, \mathbf{Q}_{k,\mathcal{N}}^{m,b})}{P} \tag{12}$$

$$\text{s.t.} \quad \sum_{k=1}^{K_m} \mathrm{Tr}(\mathbf{Q}_{k,\mathcal{N}}^{m,b}) \leq P_{max}^{m}, \tag{13}$$

$$\frac{C_{BC}^{m}(\mathbf{H}_{k,\mathcal{N}}^{m}, \mathbf{Q}_{k,\mathcal{N}}^{m,b})}{W_c} \geq \lambda_{SE(\min)}, \tag{14}$$

where $P = \zeta P_T + P_{ant}N_t + W_cP_{sp1}N_r^3K^3 + W_cP_{sp2} + \gamma C_{BC} + P_s$, C_{BC}^{m} and W_c denote the sum rate on subcarrier m achieved by DPC and the subcarrier's bandwidth respectively, $P_{max}^{m} = \frac{P_{max}}{M}$ represents the maximum power budget per subcarrier, and $P_{sp} = W_cP_{sp1}N_r^3K^3 + W_cP_{sp2}$ is the signal processing power consumption at the BS. Consequently, we just need to solve the decoupled single-carrier EE optimization problem in (12)-(14) and repeat for all the subcarriers in order to obtain the maximum system EE. Given that the multi-carrier scenario is transformed into a single-carrier scenario, the subcarrier index m is removed for the single carrier EE optimization problem in (12)-(14), i.e., the admitted user set is denoted as $\mathcal{K}$ with $K = |\mathcal{K}|$, and the channel matrix and covariance matrix as $\mathbf{H}_k$ and $\mathbf{Q}_k^{b}$ respectively. Furthermore, for any optimization problems, we can first optimize over some of the variables and then over the remaining ones [38]. Therefore, we will first study the fundamentals of EE-SE relationship per subcarrier under a fixed admitted user set and transmit antenna set, and develop a resource allocation scheme. Based on that, we then introduce the user and antenna selection approach to further explore the optimal trade-off.

III. Fundamentals of EE-SE relationship

In this section, we will study the fundamental EE-SE relationship (on a per subcarrier basis) with fixed admitted user set and transmit antenna set. Hence the MIMO-OFDMA BC has been transformed to MIMO-SC BC scenario in the following sections. The optimization problem in (12)-(14) is thus transformed to

$$\max_{\mathbf{Q}_k^b \succeq 0} \frac{C_{BC}(\mathbf{H}_1, \cdots, \mathbf{H}_K, \mathbf{Q}_1^b, \cdots, \mathbf{Q}_K^b)}{P} \quad (15)$$

$$\text{s.t.} \sum_{k=1}^{K} \text{Tr}(\mathbf{Q}_k^b) \leq P_{max}, \quad (16)$$

$$\frac{C_{BC}(\mathbf{H}_1, \cdots, \mathbf{H}_K, \mathbf{Q}_1^b, \cdots, \mathbf{Q}_K^b)}{W_c} \geq \lambda_{SE(\min)}. \quad (17)$$

To solve the above optimization problem, motivated by the EE-SE relationship in single antenna scenario [8], we first demonstrate the quasiconcavity of EE in SE, and then develop a two-layer resource allocation scheme based on the EE-SE performance.

Theorem I. *For any given SE, $\lambda_{SE} \geq \lambda_{SE(min)}$, achieved with transmit covariance matrix $\mathbf{Q}_k^b, \forall k \in \mathcal{K}$, that satisfies all constraints in (16)-(17), the maximum EE, $\lambda_{EE}^* = \max_{\mathbf{Q}_k^b \succeq 0} \lambda_{EE}(\lambda_{SE})$, is strictly quasiconcave in λ_{SE}.*

Proof: see Appendix A.

Theorem II. *In the SE region $[\lambda_{SE(min)}, \lambda_{SE(max)}]$, the EE, $\lambda_{EE}^*(\lambda_{SE})$*

(i) strictly decreases with λ_{SE} and is maximized at $\lambda_{SE} = \lambda_{SE(min)}$ if

$$\left.\frac{d\lambda_{EE}^*(\lambda_{SE})}{d\lambda_{SE}}\right|_{\lambda_{SE}=\lambda_{SE(min)}} \leq 0,$$

(ii) strictly increases with λ_{SE} and is maximized at $\lambda_{SE} = \lambda_{SE(max)}$ if

$$\left.\frac{d\lambda_{EE}^*(\lambda_{SE})}{d\lambda_{SE}}\right|_{\lambda_{SE}=\lambda_{SE(min)}} > 0$$

$$\text{and} \quad \left.\frac{d\lambda_{EE}^*(\lambda_{SE})}{d\lambda_{SE}}\right|_{\lambda_{SE}=\lambda_{SE(max)}} \geq 0, \quad (18)$$

(iii) first strictly increases and then strictly decreases with λ_{SE} and is maximized at $\lambda_{SE} = \frac{C_{BC}(\lambda_{EE}^{opt})}{W}$ if

$$\left.\frac{d\lambda_{EE}^*(\lambda_{SE})}{d\lambda_{SE}}\right|_{\lambda_{SE}=\lambda_{SE(min)}} > 0$$

$$\text{and} \quad \left.\frac{d\lambda_{EE}^*(\lambda_{SE})}{d\lambda_{SE}}\right|_{\lambda_{SE}=\lambda_{SE(max)}} < 0,$$

(iv) infeasible if

$$\lambda_{SE(min)} > \lambda_{SE(max)},$$

where $\lambda_{SE(max)}$ is the maximum SE under all constraints in (16)-(17) and $C_{BC}(\lambda_{EE}^{opt})$ is the throughput that corresponds to the maximum EE under all constraints in (16)-(17).

Proof: see Appendix B.

Since there exists a unique global maximum for any quasiconcave function, *Theorem I* guarantees the existence and uniqueness of the global maximum solution. Furthermore, $\lambda_{EE}(\lambda_{SE})$ either strictly decreases or first increases and then strictly decreases with λ_{SE} starting from $\lambda_{SE(\min)}$, which is the minimum SE constraint. *Theorem II* further indicates that the maximum point is always achieved at a finite SE. Therefore, problem (15)-(17) can be decomposed into two layers and solved iteratively through the following processes:

(i) Inner-layer: For a given SE, λ_{SE}, finds the maximum EE $\lambda_{EE}^*(\lambda_{SE})$.

(ii) Outer-layer: Finds the optimal EE, λ_{EE}^{opt}, via a gradient-based algorithm in accordance with *Theorem II*.

The key aspect of the proposed scheme lies in the inner-layer algorithm which computes $\lambda_{EE}^*(\lambda_{SE})$ and will be discussed in detail as analysis proceeds. Note that although the optimization procedure is performed based on the affine circuit power model, a more general case in which the circuit power is a convex function of throughput can be similarly proven following the analysis of the linear case.

IV. Equivalence and Duality

Evidently, the MIMO-BC sum rate maximization is a non-convex optimization problem and is difficult to solve directly. Authors in [32] showed that the capacity region of the MIMO MAC with a total power constraint P_{max} for all K transmitters is the same as the dirty paper region of the dual MIMO BC with power constraint P_{max}. In other words, any rate vector that is achievable in the dual MAC with power constraints $(P_1, P_2, \cdots, P_K)$ is in the dirty paper region of the BC with power constraint $\sum_{k=1}^{K} P_k$. Conversely, any rate vector that is in the dirty paper region of the BC is also in the dual MIMO MAC region with the same total power constraint. Furthermore, [39] showed the DPC achievable rate region in the Gaussian MIMO BC is in fact the capacity region. Hence, the dirty paper region of a MIMO BC channel with power constraint P_{max} is equal to the capacity region of the dual MIMO MAC with total power constraint P_{max}:

$$\mathcal{C}_{DPC}(P_{max}, \mathbf{H}) = \mathcal{C}_{BC}(P_{max}, \mathbf{H}) = \mathcal{C}_{MAC}(P_{max}, \mathbf{H}^H). \quad (19)$$

By exploiting the MAC-BC duality theorem, the optimization problem in (15)-(17) is equivalent to the following problem

$$\max_{\mathbf{Q}_k \succeq 0} \frac{C_{MAC}(\mathbf{H}_1^H, \cdots, \mathbf{H}_K^H, \mathbf{Q}_1, \cdots, \mathbf{Q}_K)}{\zeta \sum_{k=1}^{K} \text{Tr}(\mathbf{Q}_k) + P_{\text{fix}} + \gamma C_{MAC}} \quad (20)$$

$$\text{s.t.} \quad \sum_{k=1}^{K} \text{Tr}(\mathbf{Q}_k) \leq P_{max}, \quad (21)$$

$$\frac{C_{MAC}(\mathbf{H}_1^H, \cdots, \mathbf{H}_K^H, \mathbf{Q}_1, \cdots, \mathbf{Q}_K)}{W} \geq \lambda_{SE(\min)}, \quad (22)$$

where C_{MAC} is the rate achieved by all users of the dual MAC, and $\mathbf{Q}_k$ is the transmit signal covariance matrix of the k^{th} user, $P_{\text{fix}} = P_{ant}N_t + W_c P_{sp1} N_r^3 K^3 + W_c P_{sp2} + P_s$. We now need to solve the inner layer to find the maximum EE $\lambda_{EE}^*(\lambda_{SE})$ based on a given SE. Hence, for a given SE, i.e., any λ_{SE} in the SE region $[\lambda_{SE(\min)}, \lambda_{SE(\max)}]$, the optimization problem in (20)-(22) can be expressed as

$$\max_{\mathbf{Q}_k \succeq 0} \frac{W\lambda_{SE}}{\zeta \sum_{k=1}^{K} \text{Tr}(\mathbf{Q}_k) + P_{\text{fix}} + \gamma W \lambda_{SE}} \quad (23)$$

$$\text{s.t.} \quad \sum_{k=1}^{K} \text{Tr}(\mathbf{Q}_k) \leq P_{max}. \quad (24)$$

Given λ_{SE} is fixed in this case, $\mathbf{Q}_k, k = \{1, 2, \ldots, K\}$, are the variables of interest for the optimization problem in

(23)-(24). As a result, we can solve the above maximization problem using the following minimization method

$$\min_{\mathbf{Q}_k \succeq 0} \quad \sum_{k=1}^{K} \mathrm{Tr}(\mathbf{Q}_k) \tag{25}$$

$$\text{s.t.} \quad \frac{C_{MAC}(\mathbf{H}_1^H, \cdots, \mathbf{H}_K^H, \mathbf{Q}_1, \cdots, \mathbf{Q}_K)}{W} = \lambda_{SE} \tag{26}$$

where the capacity region of the dual MIMO-MAC $C_{MAC}(\mathbf{H}_1^H, \cdots, \mathbf{H}_K^H, \mathbf{Q}_1, \cdots, \mathbf{Q}_K) = W \log |\mathbf{I}_{N_t \times N_t} + \frac{1}{\sigma^2} \sum_{k=1}^{K} \mathbf{H}_k^H \mathbf{Q}_k \mathbf{H}_k|$. We hereafter refer to this minimization problem as the dual MAC optimization problem. Since the objective function is convex given the constraint is a convex set, the dual MAC optimization problem is a convex problem and can be solved in an efficient manner. Hence, the inner-layer of the proposed algorithm has been transformed to solve the optimization problem in (25)-(26) based on a given SE λ_{SE}. In the next section, we will introduce a method to solve the dual MAC optimization problem.

V. Dual MAC optimization problem

In this section, we first propose an efficient algorithm to solve the dual MAC optimization problem in (25)-(26), and then we introduce the MAC-to-BC covariance matrix mapping scheme to map the optimal solutions in the dual MAC scenario to the BC scenario. Finally, a complete solution to the EE-SE optimization problem in (15)-(17) is introduced.

Defining $f(\mathbf{Q}_1, \cdots, \mathbf{Q}_K) = \log |\mathbf{I}_{N_t \times N_t} + \frac{1}{\sigma^2} \sum_{k=1}^{K} \mathbf{H}_k^H \mathbf{Q}_k \mathbf{H}_k|$, we rewrite the optimization problem in (25)-(26) as

$$\min_{\mathbf{Q}_k \succeq 0} \quad \sum_{k=1}^{K} \mathrm{Tr}(\mathbf{Q}_k) \quad s.t. \ f(\mathbf{Q}_1, \cdots, \mathbf{Q}_K) = \lambda_{SE}. \tag{27}$$

Recall that the positive semi-definiteness of $\mathbf{Q}_k$ is equivalent to the non-negativeness of the eigenvalues of $\mathbf{Q}_k$ [40], i.e., $q_{k,j} \geq 0$. Correspondingly, the Lagrangian function is written as

$$L(\mathbf{Q}_1, \cdots, \mathbf{Q}_K, \eta, \delta_{k,j}) := \sum_{k=1}^{K} \mathrm{Tr}(\mathbf{Q}_k) \tag{28}$$

$$+\eta[\lambda_{SE} - f(\mathbf{Q}_1, \cdots, \mathbf{Q}_K)] - \sum_{k=1}^{K} \sum_{j=1}^{M} \delta_{k,j} q_{k,j}, \tag{29}$$

where $\eta \geq 0$ and $\delta_{k,j} \geq 0$ are the Lagrangian multipliers associated with the minimum SE constraint and the positive eigenvalues constraints, respectively. According to the KKT conditions of (27), we have

$$\mathbf{I}_{N_r \times N_r} - \eta \frac{\partial f(\mathbf{Q}_1, \cdots, \mathbf{Q}_K)}{\partial \mathbf{Q}_k} - \sum_{j=1}^{M} \delta_{k,j} \frac{\partial q_{k,j}}{\partial \mathbf{Q}_k} = 0, \tag{30}$$

$$\eta(\lambda_{SE} - f[\mathbf{Q}_1, \cdots, \mathbf{Q}_K]) = 0, \tag{31}$$

$$\delta_{k,j} q_{k,j} = 0. \tag{32}$$

Note that it is not necessary to compute the actual value of $\delta_{k,j}$ and $\frac{\partial q_{k,j}}{\partial \mathbf{Q}_k}$, because if $\delta_{k,j} \neq 0$, then $q_{k,j} = 0$. Thus, the semi-definite constraint results in $q_{k,j} = [q_{k,j}]^+$. Without loss of generality, we can assume $\delta_{k,j} = 0$.

The dual objective function of (25) is

$$g(\eta) = \min_{\mathbf{Q}_k \succeq 0} \ L(\mathbf{Q}_1, \cdots \mathbf{Q}_K, \eta), \tag{33}$$

and the dual problem is given by

$$\max_{\eta} g(\eta) \quad s.t. \quad \eta \geq 0. \tag{34}$$

In this work, we use an iterative method to obtain the optimum $\mathbf{Q}_k$ for the dual MAC problem. $\mathbf{Q}_k$ is updated using the gradient of (29) with respect to $\mathbf{Q}_k$ as follows

$$\nabla_{\mathbf{Q}_k} L := \mathbf{I}_{N_r \times N_r} - \eta \frac{\partial f[\mathbf{Q}_1(n), \cdots, \mathbf{Q}_{k-1}(n), \mathbf{Q}_k(n-1), \cdots, \mathbf{Q}_K(n-1)]}{\partial \mathbf{Q}_k(n-1)} \tag{35}$$

$$\mathbf{Q}_k(n) = [\mathbf{Q}_k(n-1) - t \nabla_{\mathbf{Q}_k} L]^+, \tag{36}$$

where t is the step size, and the notation $[\mathbf{A}]^+$ is defined as $[\mathbf{A}]^+ := \sum_i [q_i]^+ \mathbf{v}_i \mathbf{v}_i^H$, q_i and $\mathbf{v}_i$ are the i^{th} eigenvalue and the corresponding eigenvector of $\mathbf{A}$ respectively. The gradient in (35) can be readily computed as

$$\frac{\partial f(\mathbf{Q}_1, \cdots, \mathbf{Q}_K)}{\partial \mathbf{Q}_k} = \mathbf{H}_k (\mathbf{I}_{N_t \times N_t} + \frac{1}{\sigma^2} \sum_{k=1}^{K} \mathbf{H}_k^H \mathbf{Q}_k \mathbf{H}_k)^{-1} \mathbf{H}_k^H. \tag{37}$$

Next, we need to determine the optimal value of η. Since the Lagrangian function $g(\eta)$ is convex over η, the optimal η can be obtained via a one-dimensional search routine. However, because $g(\eta)$ is not necessarily differentiable, the gradient algorithm cannot be applied. Alternatively, the subgradient method can be used to find the optimal solution. In each iterative step, η is updated according to the subgradient direction.

Lemma I. *The subgradient of $g(\eta)$ is $\lambda_{SE} - \log |\boldsymbol{I}_{N_t \times N_t} + \frac{1}{\sigma^2} \sum_{k=1}^{K} \boldsymbol{H}_k^H \boldsymbol{Q}_k \boldsymbol{H}_k|$, where $\eta > 0$ and $\boldsymbol{Q}_k, k = 1, 2, \cdots, K$, are the corresponding optimal covariance matrices for a fixed η in (33).*

Proof: see Appendix C.

Upon convergence of the transmit covariance matrix $\mathbf{Q}_k$, we compare the current SE in dual MAC with λ_{SE}. *Lemma I* indicates that the value of η should increase if $\lambda_{SE} > \log |\mathbf{I}_{N_t \times N_t} + \frac{1}{\sigma^2} \sum_{k=1}^{K} \mathbf{H}_k^H \mathbf{Q}_k \mathbf{H}_k|$, and decrease otherwise. This process will continue until $g(\eta)$ converges. We are now ready to present the following algorithm, namely the bisection based resource allocation algorithm, in order to solve the dual MAC optimization problem in (25)-(26). The algorithm is detailed in Table I.

A. A Complete Solution to the EE-SE trade-off Optimization Problem with fixed admitted users and active transmit antennas

We are now ready to present a complete algorithm to solve the EE-SE optimization problem in (15)-(17). We initialize SE as $\lambda_{SE}(0)$, and find the maximum EE, $\lambda_{EE}^*(\lambda_{SE})$, using the proposed bisection based resource allocation algorithm. We then update λ_{SE} based on *Theorem III* and utilize the

1) Initialize $\eta_{\min}$ and $\eta_{\max}$;
2) **REPEAT**
3) $\eta = (\eta_{\min} + \eta_{\max})/2$;
4) **REPEAT**, Initialize $\mathbf{Q}_1(0), \cdots, \mathbf{Q}_K(0)$, $n = 1$;
5) **FOR** $k = 1, \cdots, K$
6) $\mathbf{Q}_k(n) = [\mathbf{Q}_k(n-1) - t\nabla_{\mathbf{Q}_k} L]^+$,
7) **END FOR**;
8) $n = n + 1$;
9) **UNTIL** $\mathbf{Q}_K$ for $k = 1, \cdots, K$ converge, i.e., $\|\nabla_{\mathbf{Q}_k} L\|^2 \leq \epsilon$ for a small ϵ.
10) if $\lambda_{SE} > \log\|\mathbf{I}_{N_t \times N_t} + \frac{1}{\sigma^2}\sum_{k=1}^{K} \mathbf{H}_k^H \mathbf{Q}_k \mathbf{H}_k\|$, $\eta_{\min} = \eta$, elseif $\lambda_{SE} < \log\|\mathbf{I}_{N_t \times N_t} + \frac{1}{\sigma^2}\sum_{k=1}^{K} \mathbf{H}_k^H \mathbf{Q}_k \mathbf{H}_k\|$, $\eta_{\max} = \eta$;
11) **UNTIL** $\|\eta_{\min} - \eta_{\max}\| \leq \varepsilon$.

TABLE I

BISECTION BASED RESOURCE ALLOCATION ALGORITHM

following searching scheme

$$\lambda_{SE}(n) = \begin{cases} \frac{\lambda_{SE}(n-1)}{\beta} & \left.\frac{d\lambda_{EE}^*(\lambda_{SE})}{d\lambda_{SE}}\right|_{\lambda_{SE}(n-1)} < 0 \\ \beta\lambda_{SE}(n-1) & \text{otherwise} \end{cases}. \tag{38}$$

where $\beta > 1$ is the searching step. Moreover, β needs to be reduced when the gradient $\frac{d\lambda_{EE}^*(\lambda_{SE})}{d\lambda_{SE}}$ changes sign as in

$$\beta(n) = \frac{\beta(n-1)}{2}, \tag{39}$$

and (38) is repeated until convergence, i.e., $|\lambda_{SE}(n+1) - \lambda_{SE}(n)| \leq \rho$ or end with either $\lambda_{SE(\min)}$ or $\lambda_{SE(\max)}$, where $\lambda_{SE(\max)}$ is the maximum SE under maximum power constraint. Hence, we can obtain $\lambda_{SE(\max)}$ by solving the following problem

$$\max_{\mathbf{Q}_k \succeq 0} \frac{C_{MAC}(\mathbf{H}_1^H, \cdots, \mathbf{H}_K^H, \mathbf{Q}_1, \cdots, \mathbf{Q}_K)}{W} \tag{40}$$

$$\text{s.t.} \quad \sum_{k=1}^{K} \text{Tr}(\mathbf{Q}_k) \leq P_{max}. \tag{41}$$

Since both the objective function and constraint are convex, the problem above is convex and can be solved using a method similar to that of the proposed bisection based resource allocation algorithm. Correspondingly, the Lagrangian function is written as

$$\begin{aligned} \bar{L}(\mathbf{Q}_1, \cdots, \mathbf{Q}_K, \bar{\eta}, \bar{\delta}_{k,j}) &:= f(\mathbf{Q}_1, \cdots, \mathbf{Q}_K) \\ -\bar{\eta}(\sum_{k=1}^{K} \text{Tr}(\mathbf{Q}_k) - P_{max}) &+ \sum_{k=1}^{K}\sum_{j=1}^{M} \bar{\delta}_{k,j} q_{k,j}, \end{aligned} \tag{42}$$

where $\bar{\eta} \geq 0$ and $\bar{\delta}_{k,j} \geq 0$ are the Lagrangian multipliers associated with power constraint and the positive eigenvalues constraints, respectively. Thus, the dual objective function of (40) is

$$\bar{g}(\bar{\eta}) = \max_{\mathbf{Q}_k \succeq 0} \bar{L}(\mathbf{Q}_1, \cdots \mathbf{Q}_K, \bar{\eta}), \tag{43}$$

and the dual problem is given by

$$\min \bar{g}(\bar{\eta}) \quad s.t. \quad \bar{\eta} \geq 0. \tag{44}$$

1) Initialize $\lambda_{SE}(0) \in [\lambda_{SE(\min)}, \lambda_{SE(\max)}]$, $n = 1$;
2) **REPEAT**
3) Find the maximum EE $\lambda_{EE}^*(\lambda_{SE})$ using bisection based resource allocation algorithm in Table I;
4) Update $\lambda_{SE}(n)$ using (38); $n = n + 1$;
5) **UNTIL** $\|\lambda_{SE}(n+1) - \lambda_{SE}(n)\| \leq \rho$;
6) Map the MAC covariance matrix to BC covariance matrix using the approach in [32].

TABLE II

THE COMPLETE SOLUTION TO THE EE-SE OPTIMIZATION PROBLEM

Next, $\mathbf{Q}_k$ is updated using the gradient of (42) with respect to $\mathbf{Q}_k$ as follows

$$\nabla_{\mathbf{Q}_k}\bar{L} := -\bar{\eta}\mathbf{I}_{N_r \times N_r} + \frac{\partial f[\mathbf{Q}_1(n), \cdots, \mathbf{Q}_{k-1}(n), \mathbf{Q}_k(n-1), \cdots, \mathbf{Q}_K(n-1)]}{\partial \mathbf{Q}_k(n-1)} \tag{45}$$

$$\mathbf{Q}_k(n) = [\mathbf{Q}_k(n-1) + t\nabla_{\mathbf{Q}_k} L]^+, \tag{46}$$

where the gradient in (45) is (37). We then use the similar subgradient method we propose in *Lemma 1* to determine the optimal value of $\bar{\eta}$. Hence, the upper bound $\lambda_{SE(\max)}$ is found. Note that the EE-SE optimization problem in (15)-(17) is infeasible if $\lambda_{SE(\min)} > \lambda_{SE(\max)}$. Finally, we map the optimal MAC covariance matrix to BC covariance matrix using the approach in [32]. The complete solution to the EE-SE optimization problem is summarized in Table II.

B. Data rate balancing

Suppose the achievable rate region for the problem in (40)-(41) is $\mathbf{r} \in C(\mathbf{H}, P_{max})$. The technique investigated so far aims to maximize the sum rate of MIMO-BC (or minimize the total transmit power of MIMO-BC). To ensure fairness among users, a better criterion is to maximize the data rate while balancing the rate achieved for each user as [41]

$$\max_{\mathbf{r},\nu} \nu \quad s.t. \;\; \mathbf{r} = \nu\boldsymbol{\rho}, \quad \mathbf{r} \in C(\mathbf{H}, P_{max}). \tag{47}$$

Here, maximization is performed over the choice of the scalar ν and the rate vector $\mathbf{r} = [R_1 \cdots R_K]^T$, which is constrained to belong to the capacity region denoted by $C(\mathbf{H}, P_{max})$ and to lie on the straight line defined by $\boldsymbol{\rho}$ and the origin. For example if $\boldsymbol{\rho} = \mathbf{1}_K$, all users attain the same data rate. For other values of $\boldsymbol{\rho}$, the data rate is maximized while dividing the total rate to users according to the ratio defined by the vector $\boldsymbol{\rho}$. This dual problem was shown as a weighted sum rate optimization of the following form [41]

$$\min_{\boldsymbol{\mu}} \max_{\mathbf{r}} \sum_{k=1}^{K} \mu_k \frac{R_k}{\rho_k} \quad s.t. \;\; \sum_{k=1}^{K} \mu_k = 1 \tag{48}$$

where μ_k are the Lagrangian coefficients for the k^{th} constraints in (47), and ρ_k is the k^{th} element of $\boldsymbol{\rho}$. Hence, for an initial setting of μ_k, $\max \sum_{k=1}^{K}(\frac{\mu_k}{\rho_k})R_k$ can be solved using the method described earlier, and μ_k can be updated using a subgradient method as in [41], i.e. $\mu_k^{(d)} = \mu_k^{(d-1)} - t(R_k -$

R_K), where t is a small positive step size. With the updated $\mu_k^{(d)}$, the weighted sum rate problem is solved again and $\mu_k^{(d+1)}$ is computed. This is repeated until convergence. The resulting algorithm will maximize the data rate for MIMO-BC while ensuring data rate balancing across users.

VI. User and Antenna Selection Strategy

In this section, we further study the user and antenna selection approach to explore the optimal EE-SE trade-off. With DPC, admitting all users is always optimal in terms of SE maximization, but not for EE optimization. This is because although admitting more users will achieve a higher sum-rate, it comes at a cost of higher signal processing power consumption. On the other hand, more active transmit antennas also achieve a higher sum-rate but at the cost of higher circuit power consumption. Therefore, there exists a trade-off between the power consumption cost and the sum-rate gain. As a result, user and antenna selection is necessary.

The optimal user and antenna selection approach is by exhaustive search. Based on the decoupled single-carrier problem, we need to perform the two-layer resource allocation algorithm proposed in Section V, and then chooses the optimal admitted user set and transmit antenna set after comparing the EE as follows

$$\{\mathcal{K}^{opt}, \mathcal{N}^{opt}\} = \arg \max_{\mathcal{K}\subseteq\{1,\cdots,K_{tot}\}, \mathcal{N}\subseteq\{1,\cdots,N_{tot}\}} \lambda_{EE}(\mathcal{K}, \mathcal{N}). \tag{49}$$

Therefore, the computational complexity is extremely high for the optimal solution, especially when there exists large number of users or antennas at the BS. In order to reduce the computational complexity, we investigate the property of these two trade-offs and propose a suboptimal low complexity user and antenna selection strategy. The main idea is to decouple the user and antenna selection into a two-layer selection approach.

Let us first investigate the transmit antenna selection strategy in MIMO-OFDMA, which cannot be conducted in a per-subcarrier manner. It is because different antennas would be selected for different subcarriers, resulting in all antennas being activated. Moreover, when a set of antennas is selected, it should be used for all the subcarriers as it will achieve a higher rate at no additional circuit power consumption. Hence, transmit antenna selection will have to be performed for all subcarriers together. User selection can then be conducted for each subcarrier with the selected antenna set.

Although antenna selection should not be done in a per subcarrier manner, the selection criterion obtained from the single-carrier case can be generalized for the multi-carrier scenario. Considering the full set of users and a given number of active transmit antenna N_t, we can approximate the EE as (50) in the next page, and hence (50) is equivalent to optimizing the following

$$\max_{N_t=|\mathcal{N}|} |\sum_{k=1}^{K_{tot}} \mathbf{H}_{k,\mathcal{N}}^H \mathbf{H}_{k,\mathcal{N}}|. \tag{51}$$

However, calculating the matrix determinant requires a large number of computation, especially when large number of antennas is available. Therefore, we further reduce the complexity using the Frobenius norm of the channel instead of the determinant. Although the channel Frobenius norm cannot characterize the capacity accurately, it is related to the capacity as it indicates the overall energy of the channel [42]. Thus, the selection criterion for the multi-carrier case can be generalized to

$$\text{sort}_{1\le n\le N_{tot}} ||\mathbf{h}_n||_F^2 \tag{52}$$

where $\mathbf{h}_n$ is the n^{th} column of the matrix $\tilde{\mathbf{H}}$ which represents the channel quality of the n^{th} transmit antenna at BS, and $\tilde{\mathbf{H}}$ is defined as $\tilde{\mathbf{H}} = [\mathbf{H}_{\mathcal{K},\mathcal{N}}^{1H}\ \mathbf{H}_{\mathcal{K},\mathcal{N}}^{2H}\ \cdots\ \mathbf{H}_{\mathcal{K},\mathcal{N}}^{MH}]^H$, $\mathbf{H}_{\mathcal{K},\mathcal{N}}^m$ is the channel matrix on the m^{th} subcarrier with all the users are admitted $|\mathcal{K}| = K_{tot}$. After sorting the antennas using Frobenius norm, the active transmit antenna set is selected from the first N_t antennas (first N_t columns). Therefore, we choose the transmit antenna set based on the average value through all the subcarriers in the system.

With the transmit antenna set $\mathcal{N}$ selected, the users can now be chosen for each subcarrier. For a given number of admitted users K, we can have the approximation for EE as (53) in the next page. Hence (53) is equivalent to optimizing the following

$$\max_{K=|\mathcal{K}|} |\sum_{k=1}^{K} \mathbf{H}_{k,\mathcal{N}}^H \mathbf{H}_{k,\mathcal{N}}|. \tag{54}$$

Again, we can sort the user using the Frobenius norm approach similar to the transmit antenna selection scheme

$$\text{sort}_{1\le k\le K_{tot}} ||\mathbf{H}_{k,\mathcal{N}}||_F^2 \tag{55}$$

and the admitted users are selected from the first K users of the sorted list. We then only need to perform the proposed resource allocation algorithm in Section V to maximize EE, and repeat the process for all the subcarriers. This process is continued until all the transmit antenna number has been investigated. The complete solution to the EE-SE trade-off problem for MIMO-OFDMA BC system with user and antenna selection is summarized in Table III.

VII. Low complexity fixed selection approach

The previous section presents a dynamic solution where admitted user set is considered for different subcarriers in order to maximize the system's EE. However, it requires higher computational complexity when there exists a large amount of subcarriers. To reduce the computational complexity, a strategy that selects a fixed admitted user set for all subcarriers is proposed. As shown in [41], a MIMO-OFDMA system with frequency selective channels can be viewed as a MIMO non-frequency selective system using a block diagonal matrix form as $\bar{\mathbf{H}}_{k,\mathcal{N}} = \text{diag}[\mathbf{H}_{k,\mathcal{N}}^1\ \mathbf{H}_{k,\mathcal{N}}^2 \cdots \mathbf{H}_{k,\mathcal{N}}^M]$. Using this model, the received signal can be written as

$$\mathbf{y} = \mathbf{HWs} + \mathbf{n} = \mathbf{Hx} + \mathbf{n}, \tag{56}$$

where $\mathbf{H} = [\bar{\mathbf{H}}_{1,\mathcal{N}}^T\ \bar{\mathbf{H}}_{2,\mathcal{N}}^T \cdots \bar{\mathbf{H}}_{K,\mathcal{N}}^T]^T$, $\mathbf{y} = [\mathbf{y}_{1,\mathcal{N}}^T\ \mathbf{y}_{2,\mathcal{N}}^T \cdots \mathbf{y}_{K,\mathcal{N}}^T]^T$ with $\mathbf{y}_{k,\mathcal{N}} = [\mathbf{y}_{k,\mathcal{N}}^{1T}\ \mathbf{y}_{k,\mathcal{N}}^{2T} \cdots \mathbf{y}_{k,\mathcal{N}}^{MT}]^T$, $\mathbf{W}$ is the overall precoding matrix at the BS which is defined as $\mathbf{W} = [\bar{\mathbf{W}}_{1,\mathcal{N}}^T\ \bar{\mathbf{W}}_{2,\mathcal{N}}^T \cdots \bar{\mathbf{W}}_{K,\mathcal{N}}^T]^T$ with $\bar{\mathbf{W}}_{k,\mathcal{N}} = \text{diag}[\mathbf{W}_{k,\mathcal{N}}^1\ \mathbf{W}_{k,\mathcal{N}}^2 \cdots \mathbf{W}_{k,\mathcal{N}}^M]$, and $\mathbf{s}$ is the signal

$$\max_{N_t=|\mathcal{N}|}\max_{\mathbf{Q}^m_{k,\mathcal{N}}\succeq 0}\frac{W\log|\mathbf{I}_{N_t\times N_t}+\frac{1}{\sigma^2}\sum_{k=1}^{K_{tot}}\mathbf{H}^H_{k,\mathcal{N}}\mathbf{Q}^m_{k,\mathcal{N}}\mathbf{H}_{k,\mathcal{N}}|}{\zeta P_T+P_{ant}N_t+W_cP_{sp1}N_r^3K_{tot}^3+W_cP_{sp2}+\gamma C_{BC}+P_s}$$
$$\geq\max_{N_t=|\mathcal{N}|}\max_{P}\frac{W\log|\mathbf{I}_{N_t\times N_t}+\frac{P}{K_{tot}N_t\sigma^2}\sum_{k=1}^{K_{tot}}\mathbf{H}^H_{k,\mathcal{N}}\mathbf{H}_{k,\mathcal{N}}|}{\zeta P_T+P_{ant}N_t+W_cP_{sp1}N_r^3K_{tot}^3+W_cP_{sp2}+\gamma C_{BC}+P_s}\ \text{(uniform power allocation)}$$
$$\geq\max_{N_t=|\mathcal{N}|}\max_{P}\frac{W\log|\frac{P}{K_{tot}N_t\sigma^2}\sum_{k=1}^{K_{tot}}\mathbf{H}^H_{k,\mathcal{N}}\mathbf{H}_{k,\mathcal{N}}|}{\zeta P_T+P_{ant}N_t+W_cP_{sp1}N_r^3K_{tot}^3+W_cP_{sp2}+\gamma C_{BC}+P_s}\ \text{(high SNR regime)} \tag{50}$$

$$\max_{K=|\mathcal{K}|}\max_{\mathbf{Q}^m_{k,\mathcal{N}}\succeq 0}\frac{W\log|\mathbf{I}_{N_t\times N_t}+\frac{1}{\sigma^2}\sum_{k=1}^{K}\mathbf{H}^H_{k,\mathcal{N}}\mathbf{Q}^m_{k,\mathcal{N}}\mathbf{H}_{k,\mathcal{N}}|}{\zeta P_T+P_{ant}N_t+W_cP_{sp1}N_r^3K^3+W_cP_{sp2}+\gamma C_{BC}+P_s}$$
$$\geq\max_{K=|\mathcal{K}|}\max_{P}\frac{W\log|\mathbf{I}_{N_t\times N_t}+\frac{P}{KN_t\sigma^2}\sum_{k=1}^{K}\mathbf{H}^H_{k,\mathcal{N}}\mathbf{H}_{k,\mathcal{N}}|}{\zeta P_T+P_{ant}N_t+W_cP_{sp1}N_r^3K^3+W_cP_{sp2}+\gamma C_{BC}+P_s}\ \text{(uniform power allocation)}$$
$$\geq\max_{K=|\mathcal{K}|}\max_{P}\frac{W\log|\frac{P}{KN_t\sigma^2}\sum_{k=1}^{K}\mathbf{H}^H_{k,\mathcal{N}}\mathbf{H}_{k,\mathcal{N}}|}{\zeta P_T+P_{ant}N_t+W_cP_{sp1}N_r^3K^3+W_cP_{sp2}+\gamma C_{BC}+P_s}\ \text{(high SNR regime).} \tag{53}$$

---

1) Initialization: sort the antennas using (52) based on the case that all users in the network are admitted $|\mathcal{K}|=K_{tot}$;
2) **For** $N_t=1:N_{tot}$
3) Find the best N_t antennas based on Frobenius norm method;
4) **For** subcarrier $m=1:M$
5) **For** $K=1:K_{tot}$
6) Find the best K users based on Frobenius norm method (55);
7) Calculate the optimal EE using the proposed resource allocation scheme in Section V, denoted as $\lambda_{EE}^{opt}(K,N_t,m)$;
8) **End For**
9) Compare all the EE in the buffer and select the set of users on subcarrier m that maximizes the EE.
10) **End For**
11) **End For**
12) Compare all the EE in the buffer and select the set of transmit antenna that maximizes the EE.

TABLE III

THE COMPLETE SOLUTION TO THE EE-SE TRADE-OFF PROBLEM WITH USER AND ANTENNA SELECTION

corresponding to the precoding matrix $\mathbf{W}$ defined as $\mathbf{s}=[\mathbf{s}^T_{1,\mathcal{N}}\ \mathbf{s}^T_{2,\mathcal{N}}\cdots\mathbf{s}^T_{K,\mathcal{N}}]^T$ with $\mathbf{s}_{k,\mathcal{N}}=[\mathbf{s}^{1T}_{k,\mathcal{N}}\ \mathbf{s}^{2T}_{k,\mathcal{N}}\cdots\mathbf{s}^{MT}_{k,\mathcal{N}}]^T$. The precoded transmit signal is $\mathbf{x}=[\mathbf{x}^T_{1,\mathcal{N}}\ \mathbf{x}^T_{2,\mathcal{N}}\cdots\mathbf{x}^T_{K,\mathcal{N}}]^T$ with $\mathbf{x}_{k,\mathcal{N}}=[\mathbf{x}^{1T}_{k,\mathcal{N}}\ \mathbf{x}^{2T}_{k,\mathcal{N}}\cdots\mathbf{x}^{MT}_{k,\mathcal{N}}]^T$, and the noise is $\mathbf{n}=[\mathbf{n}^T_1\ \mathbf{n}^T_2\cdots\mathbf{n}^T_K]^T$ with $\mathbf{n}_k=[\mathbf{n}^{1T}_k\ \mathbf{n}^{2T}_k\cdots\mathbf{n}^{MT}_k]^T$.

By employing DPC with MIMO-OFDMA at the transmitter with antenna set $\mathcal{N}$, it was shown in [41] that the optimal covariance matrices for the sum rate maximization problem have a block diagonal structure matching the structure of their respective channels $\bar{\mathbf{H}}_{k,\mathcal{N}}$ as $\bar{\mathbf{Q}}^b_{k,\mathcal{N}}=\mathbb{E}(\mathbf{x}_{k,\mathcal{N}}\mathbf{x}^H_{k,\mathcal{N}})=\text{diag}[\mathbf{Q}^{1,b}_k\cdots\mathbf{Q}^{M,b}_k]$, where $\bar{\mathbf{Q}}^b_{k,\mathcal{N}}\succeq 0$. Without loss of generality, an encoding order $(1,\cdots,K)$ for the user set $\mathcal{K}$, the data rate R^b_k for the k^{th} user can be written as [32]

$$R^b_k=W\log\frac{|\mathbf{I}_{N_r\times N_r}+\frac{1}{\sigma^2}\bar{\mathbf{H}}_{k,\mathcal{N}}(\sum_{i=k}^{K}\bar{\mathbf{Q}}^b_{i,\mathcal{N}})\bar{\mathbf{H}}^H_{k,\mathcal{N}}|}{|\mathbf{I}_{N_r\times N_r}+\frac{1}{\sigma^2}\bar{\mathbf{H}}_{k,\mathcal{N}}(\sum_{i=k+1}^{K}\bar{\mathbf{Q}}^b_{i,\mathcal{N}})\bar{\mathbf{H}}^H_{k,\mathcal{N}}|}. \tag{57}$$

Moreover, the signal processing power consumption at the BS is written as

$$P_{sp}=WP_{sp1}[\dim(\mathbf{W})]^{\beta+1}+WP_{sp2}$$
$$=WP_{sp1}N_r^3M^3K^3+WP_{sp2}. \tag{58}$$

Hence, with the selected users ($\mathcal{K}$) and antennas ($\mathcal{N}$), the optimization problem can be formulated as

$$\max_{\mathcal{K},\mathcal{N},\{\bar{\mathbf{Q}}^b_{k,\mathcal{N}}\succeq 0\}}\frac{C_{BC}(\bar{\mathbf{H}}_{k,\mathcal{N}},\bar{\mathbf{Q}}^b_{k,\mathcal{N}})}{P} \tag{59}$$
$$\text{s.t.}\quad\sum_{k=1}^{K}\text{Tr}(\bar{\mathbf{Q}}^b_{k,\mathcal{N}})\leq P_{max}, \tag{60}$$
$$\frac{C_{BC}(\bar{\mathbf{H}}_{k,\mathcal{N}},\bar{\mathbf{Q}}^b_{k,\mathcal{N}})}{W}\geq\lambda_{SE(\min)}. \tag{61}$$

where $P=\zeta P_T+P_{ant}MN_t+WP_{sp1}N_r^3M^3K^3+WP_{sp2}+\gamma C_{BC}+P_s$. This optimization problem has the same structure as the optimization problem in (9)-(11), and hence the resource allocation algorithm proposed in Section V and transmit antenna selection approach in Section VI can be applied here. However, since we employ a fixed user selection strategy through all subcarriers, users selection criterion in (55) should be modified as follows

$$\text{sort}_{1\leq k\leq K_{tot}}||\bar{\mathbf{H}}_{k,\mathcal{N}}||^2_F \tag{62}$$

As we only need to perform the proposed resource allocation algorithm in Section V once (for all the subcarriers), the computational complexity for the fixed user selection scheme is much lower compared to the proposed dynamic strategy.

In Table IV, the computational complexities of the aforementioned decoupled algorithm and fixed selection approach are listed for comparison. The complexity of DPC are based on QR decomposition in [36]. We calculate the computational complexity based on the number of floating points [42]. As can be seen from Table IV, the proposed fixed selection approach has a lower computational complexity compared to the proposed decoupled approach.

Algorithm	Complexity
Decoupled algorithm	$\mathcal{O}(\frac{1}{\beta^2}\frac{1}{\rho^2}M^2K_{tot}^3N_{tot}^4N_r)$
Fixed selection approach	$\mathcal{O}(\frac{1}{\beta^2}\frac{1}{\rho^2}2MK_{tot}^3N_{tot}^4N_r)$

TABLE IV
COMPLEXITY COMPARISON FOR THE PROPOSED ALGORITHMS

VIII. SIMULATION RESULTS

Maximum power, P_{max}	46 dBm
Dynamic power consumption per unit data rate, γ	1 W/Mbps
Drain efficiency of power amplifier, ζ	38%
Circuit power per RF chains, P_{ant}	7.5 W
Power consumed by computing the precoding matrix, P_{sp1}, P_{sp2}	0.94 μW/Hz, 0.54 μW/Hz
Static circuit power, P_s	20 W

TABLE V
LIST OF SIMULATION PARAMETERS

In this section, we present simulation results to verify the theoretical findings and analyze the effectiveness of the proposed approaches. In our simulation, the total number of subcarriers is $M = 1024$ while the bandwidth for each subcarrier W_c is 15 KHz, the noise power is -174 dBm/Hz, the pathloss is $128.1+37.6\log_{10} d$ with distance d (in kilometers) [43], and the radius of the cell is set to 500 m. The power related parameters in the simulation are similar to those in [34] and are listed in Table V. These system parameters are merely chosen to demonstrate the EE-SE relationship as an example and can easily be modified to any other values to assess different scenarios.

In the first simulation, we evaluate the relationship between EE and SE of MIMO-BC (per subcarrier). We fix the number of admitted users to $K = 3$, and the number of antennas at BS and each user to $N_t = 4$ and $N_r = 2$ respectively. To investigate the EE-SE relationship, no maximum transmit power constraint or specific minimum SE requirement is imposed to investigate the performance limit. From Fig. 1, the EE-SE relationship is quasiconcave; this quasiconcavity is the foundation of the proposed two layer approach. It also infers that the proposed bisection based resource allocation algorithm approach can serve as an optimal inner layer algorithm that

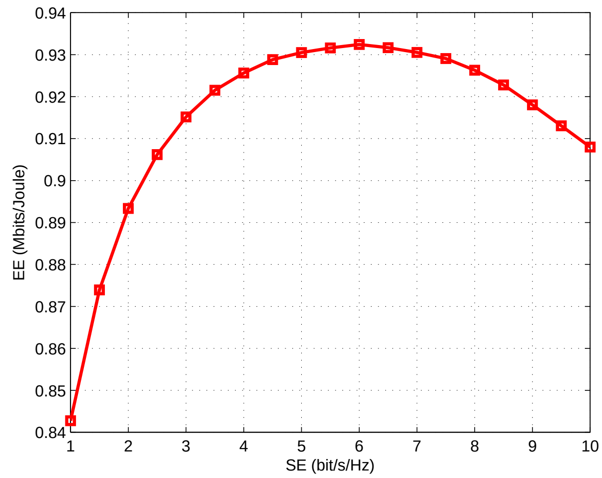

Fig. 1. EE-SE relationship (λ_{EE}^*-versus-λ_{SE}) with fixed admitted users and active transmit antennas, where $K = 3$, $N_t = 4$ and $N_r = 2$.

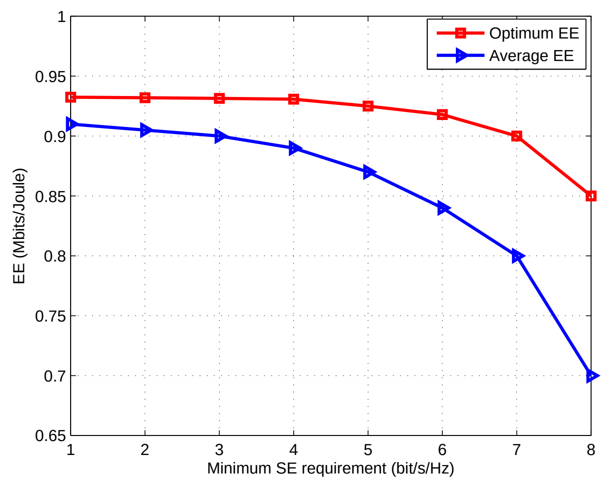

Fig. 2. Comparison of the achievable EE obtained by the proposed two-layer scheme for different minimum SE requirements (λ_{EE}^{opt}-versus-$\lambda_{SE(\min)}$), where $K = 3$, $N_t = 4$ and $N_r = 2$.

leads to the optimal EE-SE performance. We then evaluate the EE-SE relation with different minimum SE requirement and present the optimal and average EE in Fig. 2. It can be observed from Fig. 2 that the optimal EE is the same up to a certain minimum SE requirement, but drops afterwards. This can be explained using the quasiconcavity of EE-SE relationship from Fig. 1. When the minimum SE requirement is low, the required transmit power is also low. Therefore, the most energy efficient design is to operate at a higher transmit power in order to achieve the optimal EE. This is why the optimal EE is constant for low minimum SE requirements. However when the minimum SE requirement is high, the optimal operation is to simply fulfil that SE requirement as in this case the higher the SE, the lower the EE. Finally, the performance of the proposed decoupled approach is compared to the optimal EE in Fig. 3. It can be observed that the proposed decoupled algorithm approaches the optimal EE when the number of subcarriers increases. Hence, although the decoupled approach is suboptimal, it approaches the optimal

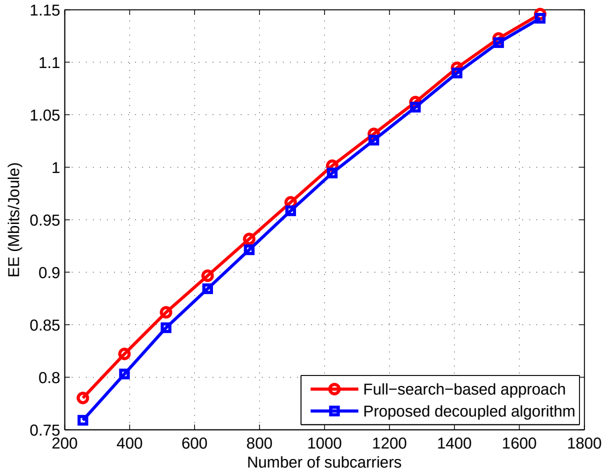

Fig. 3. The performance of the proposed decoupled algorithm in terms of EE.

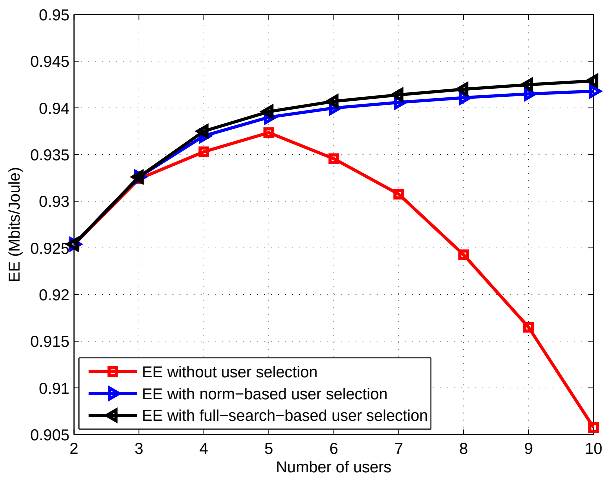

Fig. 4. Comparison of the achievable EE for different number of users in the network, where N_t = 8 and $N_r = 2$.

one with a lower complexity.

In the next simulation, we evaluate the achievable EE under different number of users in the proposed MIMO-OFDMA system. To show the impact of user selection, we fix the number of active antennas at BS and each user to $N_t = 8$ and $N_r = 2$ respectively, and compare the EE achieved by the proposed norm-based user selection scheme with the optimal full-search-based approach and the scheme without user selection. As shown in Fig. 4, the EE achieved by the proposed norm-based user selection and the optimal full-search-based schemes are both monotonically increasing as a function of the total number of users K_{tot} in the system. The performance gain comes from the user selection diversity of DPC as the probability of choosing the admitted users with better channel conditions increases when the total number of users increases. This is in stark contrast to the case without user selection strategy, where the EE first increases then decreases with increasing number of users in the system. The best EE performance in this simulation scenario is achieved

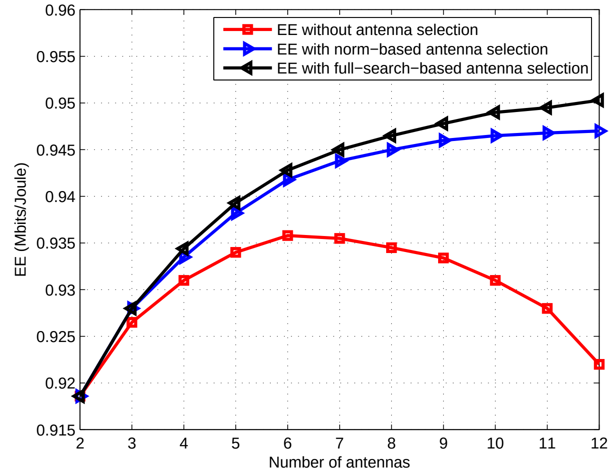

Fig. 5. Comparison of the achievable EE for different number of antennas at the BS, where $K = 3$ and $N_r = 2$.

when there exists $K_{tot} = 5$ users. Since the spatial dimensions for DPC is $\min(N_r \times K_{tot}, N_t)$ [44], the maximum multiplexing gain under these simulation parameters is eight, which also indicates the maximum number of users achieving the maximum spatial dimensions is $K_{tot} = 4$. Nonetheless, there exists multiuser diversity in BC scenario; in other words, the sum rate capacity will still increase marginally if the number of admitted users is increased beyond 4. On the other hand, as shown in (7), the power consumption increases with K^3. Hence when the number of users is further increased, the increase in power consumption will outgrow the gain in sum-rate from multiuser diversity. This explains why the optimal number of user is 5 in this scenario. Also the advantage of having user selection is clearly demonstrated. It must also be noted that the proposed norm-based selection suffers only minor degradation from the high complexity optimal approach.

The achievable EE under different number of transmit antennas at the BS has also been evaluated. Similarly, to show the impact of antenna selection in terms of EE, we fix the number of admitted users in the network to $K = 3$ and each employ $N_r = 2$ receive antennas. Based on these parameter, we compare the EE achieved by the proposed norm-based scheme with the optimal full-search-based approach and the scheme without antenna selection. Fig. 5 shows that the EE achieved by the proposed norm-based and the optimal full-search-based schemes are both monotonically increasing as a function of the total number of antenna N_{tot} at the BS. In this case, the EE gain comes from the spatial diversity through antenna selection. Again, the practicability of the proposed norm-based approach is demonstrated as its EE performance is close to the high complexity optimal approach. However, the scheme without antenna selection has a significantly worse performance where the EE first increases then decreases with increasing number of antennas at the BS. The best EE performance is achieved when there are $N_{tot} = 6$ antennas. This can also be explained using the spatial dimensions of DPC. Since there are three users with two receive antennas each, the maximum spatial dimensions is six. Although further

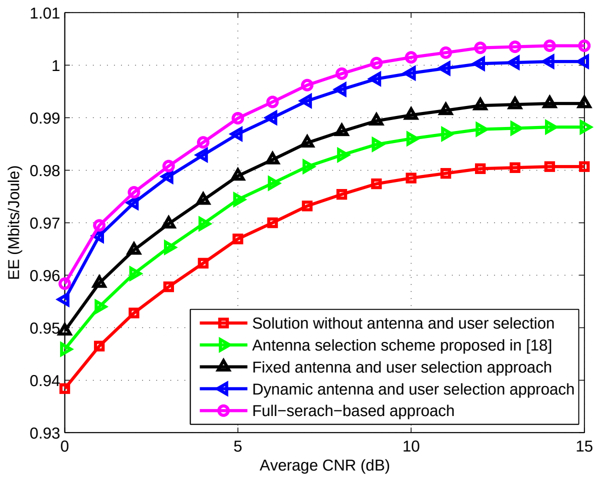

Fig. 6. Performance evaluation of the achievable EE obtained by the proposed norm-based user and antenna selection scheme, where $K_{tot} = 8$, $N_{tot} = 10$ and $N_r = 2$.

increasing the number of antennas can achieve more spatial diversity gain, the power consumption is increased more and hence EE decreases after $N_{tot} = 6$.

Finally, we evaluate the performance of the proposed combined norm-based user and antenna selection algorithm for the MIMO-OFDMA system on different channel to noise ratios (CNRs). To show the EE gain of the proposed scheme, we compare with the scheme in [19] which maximizes the EE in MIMO-BC scenario. To ensure fair comparison, we set the same parameters for the scheme in [19]. We also compare with the scheme without any user and antenna selection. Hence, this scheme is always employing all transmit antennas and admitting all users in the network, and thus achieves the highest SE. Moreover, to show the EE degradation of the proposed combined norm-based user and antenna selection approach, we compare the result with a full-search-based solution, which needs to check all the possible combination of admitted users and active transmit antennas. In other words, this full-search-based solution is the upper bound of the achievable EE in the system. As shown in Fig. 6, the EE achieved by both the proposed dynamic and fixed norm-based selection approach outperform the EE achieved in [19], and the dynamic approach is very close to the optimal full-search-based solution, but with a lower complexity. On the other hand, the EE achieved by the proposed fixed allocation scheme is lower than the proposed dynamic allocation approach but with even lower complexity. Finally, as expected, the scheme without user and antenna selection has the worst performance in terms of EE even though it maximizes the SE.

IX. Conclusions

In this paper, we investigated the fundamental EE-SE relationship in a MIMO-OFDMA BC scenario. A practical power model which is related to the number of admitted users as well as the number of active transmit antennas is considered. By separating the multi-carrier EE optimization problem to a set of single carrier EE optimization problems, a two-layer resource allocation algorithm has been developed based on the quasiconcavity of EE-SE relationship. Accordingly, a novel inner-layer algorithm was proposed and solved by applying the principle of MAC-BC duality. The algorithm in its dual form is implemented using the sub-gradient method and bisection searching scheme. To further explore the EE-SE trade-off, we then study the user and antenna selection approach where a dynamic user and antenna selection approach based on Frobenius norm method is developed. To further reduce the computational complexity, in contrast to the proposed dynamic solution where the admitted user set is selected for each subcarrier, a user selection strategy that is fixed for all subcarriers is developed. Simulation results show that the EE achieved by both the proposed dynamic and fixed norm-based selection approach outperform an existing work [19]. Moreover, the dynamic approach performs very close to the optimal full-search-based solution at a lower complexity. With even lower complexity, the EE achieved by the proposed fixed allocation scheme is lower than the proposed dynamic allocation approach but better than the one without selection. The proposed approaches can therefore improve the EE performance of MIMO-OFDMA BC.

Appendix A
Proof of Theorem I

Proof: In order to prove the quasiconcavity of $\lambda_{EE}(\lambda_{SE})$, we first introduce the definition of a quasiconcave function. A function $f : \mathbf{R}^n - \mathbf{R}$ is called quasiconvex if its domain and all its sublevel sets

$$\mathcal{S}_\theta = \{x \in \mathbf{dom} f | f(x) \le \theta\}, \theta \in \mathcal{R} \tag{63}$$

for $\theta \in \mathbf{R}$, are convex [40]. A function is quasiconcave if $-f$ is quasiconvex, i.e., every superlevel set $\{x | f(x) \ge \theta\}$ is convex. Hence, we denote the superlevel sets of $\lambda^*_{EE}(\lambda_{SE})$ as

$$\mathcal{S}_\beta = \{\lambda_{SE} \ge \lambda_{SE(\min)} | \lambda^*_{EE}(\lambda_{SE}) \ge \beta\}. \tag{64}$$

According to [40], $\lambda^*_{EE}(\lambda_{SE})$ is strictly quasiconcave in λ_{SE} if $\mathcal{S}_\beta$ is strictly convex for any real number β. When $\beta < 0$, no points exist on the counter $\lambda^*_{EE}(\lambda_{SE}) = \theta$. When $\beta \ge 0$, we rewrite λ_{EE} as $\lambda_{EE} = \frac{W\lambda_{SE}}{\zeta P_T + P_s + \gamma C_{BC}}$, and hence $\mathcal{S}_\beta$ is equivalent to $\beta\zeta P_T(\lambda_{SE}) + \beta P_s + (\beta\gamma W - 1)\lambda_{SE} \le 0$. Recall the optimization problem in (15)-(17), we can rewrite this optimization problem as follows

$$\begin{aligned}\lambda_{SE}(P_T) &= \min_{\chi \ge 0} \max_{\mathbf{Q}_k^m} f(\mathbf{Q}_1^m, \cdots, \mathbf{Q}_K^m) \\ &\quad - \chi[\sum_{k=1}^{K} \mathrm{Tr}(\mathbf{Q}_k^m) - P_T] \\ &\le \max_{\mathbf{Q}_k^m} f(\mathbf{Q}_1^m, \cdots, \mathbf{Q}_K^m) - \bar{\chi}[\sum_{k=1}^{K} \mathrm{Tr}(\mathbf{Q}_k^m) - P_T] \\ &= \max_{\mathbf{Q}_k^m} f(\mathbf{Q}_1^m, \cdots, \mathbf{Q}_K^m) - \bar{\chi}[\sum_{k=1}^{K} \mathrm{Tr}(\mathbf{Q}_k^m) - \bar{P}_T] \\ &\quad + \bar{\chi}[P_T - \bar{P}_T]\end{aligned}$$

$$= \lambda_{SE}(\bar{P}_T) + \bar{\chi}[P_T - \bar{P}_T]. \quad (65)$$

$\lambda_{SE}(\bar{P}_T)+\bar{\chi}[P_T-\bar{P}_T]$ is an upper bound to $\lambda_{SE}(P_T)$ depends on the subgradient $\bar{\chi}$ at the point $\bar{P}_T$, and hence $\lambda_{SE}(P_T)$ is a concave function and monotonically increases in P_T, we here denote $\lambda_{SE}(P_T)$ as a concave function h. Let $X = h(x) \in \mathcal{S}$, $Y = h(y) \in \mathcal{S}$, thus we have

$$h(\delta x + (1-\delta)y) \geq \delta h(x) + (1-\delta)h(y) \quad (66)$$

where $\delta \in \{0, 1\}$. h^{-1} is also strictly increasing on $\mathcal{S}$ from inverse of strictly monotone function. Thus, we have

$$\delta h^{-1}(X) + (1-\delta)h^{-1}(Y) = \delta x + (1-\delta)y \geq h^{-1}(\delta X + (1-\delta)Y). \quad (67)$$

Hence, h^{-1} is a convex function and $P_T(\lambda_{SE})$ is a convex function in λ_{SE}. Consequently, $\mathcal{S}_\beta$ is convex and $\lambda^*_{EE}(\lambda_{SE})$ is strictly quasiconcave in λ_{SE}. This completes the proof of *Theorem I*. ■

APPENDIX B
PROOF OF THEOREM II

Proof: In order to prove *Theorem II*, we analyze the limit of $\lambda^*_{EE}(\lambda_{SE})$ as follows

$$\begin{aligned} \lim_{\lambda_{SE}\to\infty} \lambda^*_{EE}(\lambda_{SE}) &= \lim_{\lambda_{SE}\to\infty} \max_{\mathbf{Q}_k^b \succeq 0} \frac{W\lambda_{SE}}{\zeta P^*_T(\lambda_{SE}) + P_s + \gamma W \lambda_{SE}} \\ &= \lim_{\lambda_{SE}\to\infty} \frac{o(P^*_T(\lambda_{SE}))}{P^*_T(\lambda_{SE})} \\ &= 0 \end{aligned} \quad (68)$$

Hence, with strict concavity of $\lambda^*_{EE}(\lambda_{SE})$ which is proved in Appendix B, starting from $\lambda_{SE} = \lambda_{SE(\min)}$, $\lambda^*_{EE}(\lambda_{SE})$ either strictly decreases with λ_{SE} if $\left.\frac{d\lambda^*_{EE}(\lambda_{SE})}{d\lambda_{SE}}\right|_{\lambda_{SE}=\lambda_{SE(\min)}} \leq 0$, or first strictly increases and then strictly decreases with λ_{SE} if $\left.\frac{d\lambda^*_{EE}(\lambda_{SE})}{d\lambda_{SE}}\right|_{\lambda_{SE}=\lambda_{SE(\min)}} > 0$. And the maximum EE in the SE region $[\lambda_{SE(\min)}, \lambda_{SE(\max)}]$ is straightforward as indicated in *Theorem II*. This completes the proof. ■

APPENDIX C
PROOF OF LEMMA I

Proof: Let ν be the subgradient of $g(\check{\eta})$. For a given $\check{\eta} > 0$, the subgradient ν of $g(\check{\eta})$ satisfies $g(\hat{\eta}) \leq g(\check{\eta}) + \nu(\hat{\eta} - \check{\eta})$, where $\hat{\eta}$ is any feasible value. Let $\hat{\mathbf{Q}}_k, \{k = 1, \cdots, K\}$, denote the optimal covariance matrices in (33) for $\eta = \hat{\eta}$, and $\check{\mathbf{Q}}_k, \{k = 1, \cdots, K\}$, denote the optimal covariance matrices in (33) for $\eta = \check{\eta}$. We express $g(\hat{\eta})$ as

$$\begin{aligned} g(\hat{\eta}) &= \min_{\mathbf{Q}_k \succeq 0} \sum_{k=1}^K tr(\mathbf{Q}_k) + \hat{\eta}[\lambda_{SE} - f(\mathbf{Q}_1, \cdots, \mathbf{Q}_K)] \\ &= \sum_{k=1}^K tr(\hat{\mathbf{Q}}_k) + \hat{\eta}[\lambda_{SE} - f(\hat{\mathbf{Q}}_1, \cdots, \hat{\mathbf{Q}}_K)] \\ &\leq \sum_{k=1}^K tr(\check{\mathbf{Q}}_k) + \hat{\eta}[\lambda_{SE} - f(\check{\mathbf{Q}}_1, \cdots, \check{\mathbf{Q}}_K)] \\ &= \sum_{k=1}^K tr(\check{\mathbf{Q}}_k) + \check{\eta}[\lambda_{SE} - f(\check{\mathbf{Q}}_1, \cdots, \check{\mathbf{Q}}_K)] \\ &\quad -\check{\eta}[\lambda_{SE} - f(\check{\mathbf{Q}}_1, \cdots, \check{\mathbf{Q}}_K)] \\ &\quad +\hat{\eta}[\lambda_{SE} - f(\check{\mathbf{Q}}_1, \cdots, \check{\mathbf{Q}}_K)] \\ &= g(\check{\eta}) + (\hat{\eta} - \check{\eta})[\lambda_{SE} - f(\check{\mathbf{Q}}_1, \cdots, \check{\mathbf{Q}}_K)]. \end{aligned} \quad (69)$$

Hence the subgradient of $g(\check{\eta})$ is $\nu := \lambda_{SE} - f(\check{\mathbf{Q}}_1, \cdots, \check{\mathbf{Q}}_K)$. This concludes the proof. ■

REFERENCES

[1] R. S. Prabhu and B. Daneshrad, “Performance analysis of energy-efficient power allocation for MIMO-MRC systems,” *IEEE Trans. Commun.*, vol. 60, no. 8, pp. 2048 – 2053, Aug. 2012.

[2] L. Yan, B. Bai, and W. Chen, “On energy efficiency maximization in downlink mimo systems exploiting multiuser diversity,” *IEEE Commun. Lett.*, vol. 18, no. 12, pp. 2161–2164, Dec. 2014.

[3] O. K. Rayel, G. Brante, J. L. Rebelatto, R. D. Souza, and M. A. Imran, “Energy efficiency-spectral efficiency trade-off of transmit antenna selection,” *IEEE Trans. Commun.*, vol. 62, no. 12, pp. 4293–4303, Dec. 2014.

[4] O. Onireti, F. Heliot, and M. A. Imran, “On the energy efficiency-spectral efficiency trade-off of distributed MIMO systems,” *IEEE Trans. Commun.*, vol. 61, no. 9, pp. 3741–3753, Sept. 2013.

[5] H. Zhang, Y. Huang, S. Li, and L. Yang, “Energy-efficient precoder design for MIMO wiretap channels,” *IEEE Commun. Lett.*, vol. 18, no. 9, pp. 1559–1562, Sept. 2014.

[6] E. Bjornson, J. Hoydis, M. Kountouris, and M. Debbah, “Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits,” *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 7112–7139, Nov. 2014.

[7] J. Tang, D. So, E. Alsusa, K. A. Hamdi, and A. Shojaeifard, “Resource allocation for energy efficiency optimization in heterogeneous networks,” *IEEE Journal on Sel. Areas in Commun.*, to appear in 2015.

[8] C. Xiong, G. Li, S. Zhang, Y. Chen, and S. Xu, “Energy- and spectral-efficiency tradeoff in downlink OFDMA networks,” *IEEE Trans. Wireless Commun.*, vol. 10, no. 11, pp. 3874 – 3886, Nov. 2011.

[9] L. Venturino and S. Buzzi, “Energy-aware and rate-aware heuristic beamforming in downlink MIMO OFDMA networks with base station coordination,” *IEEE Trans. Veh. Tech.*, to appear.

[10] X. Xiao, X. Tao, and J. Lu, “Energy-efficient resource allocation in lte-based MIMO-OFDMA systems with user rate constraints,” *IEEE Trans. Veh. Tech.*, vol. 64, no. 1, pp. 185–197, Jan. 2015.

[11] M. Jang, H. Park, Y. Kwon, and T. Hwang, “Energy-efficient adaptation of pilot power, data power, and transmission rate for downlink multiuser MIMO systems,” *IEEE Trans. Veh. Tech.*, to appear.

[12] X. Ge, X. Huang, Y. Wang, M. Chen, Q. Li, T. Han, and C. X. Wang, “Energy-efficiency optimization for MIMO-OFDM mobile multimedia communication systems with QoS constraints,” *IEEE Trans. Veh. Tech.*, vol. 63, no. 5, pp. 2127–2138, June 2014.

[13] J. Tang, D. So, E. Alsusa, K. A. Hamdi, and A. Shojaeifard, “Energy efficiency optimization with interference alignment in multi-cell MIMO interfering broadcast channels,” *IEEE Trans. Commun.*, vol. 63, no. 7, pp. 2486–2499, July 2015.

[14] C. Ghosh and S. Roy, “Coexistence challenges for heterogeneous cognitive wireless networks in TV white spaces,” *IEEE Wireless Communications*, vol. 18, no. 4, pp. 22 – 31, Aug. 2011.

[15] L. Fu, Y. J. A. Zhang, and J. Huang, “Energy efficient transmissions in MIMO cognitive radio networks,” *IEEE Journal on Sel. Areas in Commun.*, vol. 31, no. 11, pp. 2420–2431, Nov. 2013.

[16] Z. Chong and E. Jorswieck, “Energy efficiency in random opportunistic beamforming,” in *Proc. Veh. Technol. Conf.*, 2011.

[17] Z. Chong and E. Jorswieck, “Analytical foundation for energy efficiency optimisation in cellular networks with elastic traffic,” in *Int. ICST Conf. Mobile Lightweight Wireless Syst.*, Bilbao, Spain, May, 2011.

[18] J. Xu, L. Qiu, and C. Yu, "Improving energy efficiency through multimode transmission in the downlink MIMO systems," *EURASIP J. Wireless Commun. and Net.*, 2011.

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

[20] C. Hellings and W. Utschick, "Energy effciency optimization in MIMO broadcast channels with fairness constraints," in *International Workshop on Signal Processing Advances in Wireless Communications*, Darmstadt, Germany, June 2013.

[21] F. Meshkati, H. V. Poor, and S. C. Schwartz, "Energy-efficient resource allocation in wireless networks," *IEEE Signal Process. Mag.*, vol. 24, no. 3, pp. 58 – 68, May 2007.

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

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

[24] A. Lozano, A. M. Tulino, and S. Verdu, "Multiple-antenna capacity in the low-power regime," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2527 – 2544, Oct. 2003.

[25] G. Caire, G. Taricco, and E. Biglieri, "Suboptimality of TDMA in the low-power regime," *IEEE Trans. Inform. Theory*, vol. 50, no. 4, pp. 608 – 620, Apr. 2004.

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

[27] Y. Chen, S. Zhang, S. Xu, and G. Y. Li, "Fundamental tradeoffs on green wireless networks," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 30 – 37, June 2011.

[28] C. Bae and W. E. Stark, "End-to-end energy-bandwidth tradeoff in multihop wireless networks," *IEEE Trans. Inf. Theory*, vol. 55, no. 9, pp. 4051 – 4066, Sep. 2009.

[29] F. Heliot, M. A. Imran, and R. Tafazolli, "A very tight approximation of the SISO energy efficiency-spectral efficiency trade-off," *IEEE Commun. Lett.*, vol. 16, no. 6, pp. 850 – 853, June 2012.

[30] F. Heliot, M. A. Imran, and R. Tafazolli, "On the energy efficiency-spectral efficiency trade-off over the MIMO rayleigh fading channel," *IEEE Trans. Commun.*, vol. 60, no. 5, pp. 1345 – 1356, May 2012.

[31] J. Tang, D. So, E. Alsusa, and K. A. Hamdi, "Resource efficiency: A new paradigm on energy efficiency and spectral efficiency tradeoff," *IEEE Trans. Wireless Commun.*, vol. 13, no. 8, pp. 4656 – 4669, Aug. 2014.

[32] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2658 – 2668, Oct. 2003.

[33] C. Isheden and G. P. Fettweis, "Energy-efficient multi-carrier link adaptation with sum rate-dependent circuit power," in *Proc. IEEE Global Telecommun. Conf.*, Dec. 2010.

[34] A. J. Fehske, P. Marsch, and G. P. Fettweis, "Bit per Joule efficiency of cooperating base stations in cellular networks," in *GLOBECOM Workshops (GC Wkshps), 2010*, pp. 1406 – 1411, Miami, FL, Dec. 2010.

[35] J. Joung, Y. K. Chia, and S. Sun, "Energy-efficient, large-scale distributed-antenna system (L-DAS) for multiple users," *IEEE Journal of Selected Topics in Signal Process.*, vol. 8, no. 5, p. 954 965, Oct. 2014.

[36] L.-N. Tran, M. Juntti, M. Bengtsson, and B. Ottersten, "Weighted sum rate maximization for MIMO broadcast channels using dirty paper coding and zero-forcing methods," *IEEE Transactions on Communications*, vol. 61, no. 6, pp. 2362–2373, June 2013.

[37] J. Xu, S. Li, L. Qiu, B. S. Slimane, and C. Yu, "Energy efficient downlink mimo transmission with linear precoding," *Science China-Information Sciences*, vol. 56, no. 2, pp. 79–81, Feb. 2013.

[38] S. Boyd, *Branch and Bound Methods.* Stanford University.

[39] H. Weingarten, Y. Steinberg, and S. S. (Shitz), "The capacity region of the GAUSSIAN multiple-input multiple-output broadcast channel," *IEEE Trans. Inform. Theory*, vol. 52, no. 9, Sept. 2006.

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

[41] P. Tejera, W. Utschick, J. A. Nossek, and G. Bauch, "Rate balancing in multiuser MIMO OFDM systems," *IEEE Trans. Commun.*, vol. 57, no. 5, pp. 1370 – 1380, May 2009.

[42] M. P. Holmes, A. G. Gray, and C. L. Isbell, "Fast svd for large-scale matrices," College of Computing, Georgia Institute of Technology, Atlanta, GA.

[43] 3GPP, "Coordinated multi-point operation for LTE physical layer aspects (Release 11)," Feb. 2011 [Online]. Available: ftp.3gpp.org, TR 36.819 v0.0.1.

[44] N. Jindal and A. Goldsmith, "Dirty-paper coding versus TDMA for MIMO broadcast channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 5, p. 1783 1794, May 2005.