Employing Antenna Selection to Improve Energy-Efficiency in Massive MIMO Systems

Masoud Arash*, Ehsan Yazdian*, Mohammad Sadegh Fazel*, Glauber Brante†, Muhammad Imran‡

* Isfahan University of Technology (IUT), Isfahan, Iran
† Federal University of Paran, Curitiba, Brazil
‡ University of Glasgow, Glasgow, UK

ABSTRACT

Massive MIMO systems promise high data rates by employing large number of antennas, which also increases the power usage of the system as a consequence. This creates an optimization problem which specifies how many antennas the system should employ in order to operate with maximal energy efficiency. Our main goal is to consider a base station with a fixed number of antennas, such that the system can operate with a smaller subset of antennas according to the number of active user terminals, which may vary over time. Thus, in this paper we propose an antenna selection algorithm which selects the best antennas according to the better channel conditions with respect to the users, aiming at improving the overall energy efficiency. Then, due to the complexity of the mathematical formulation, a tight approximation for the consumed power is presented, using the Wishart theorem, and it is used to find a deterministic formulation for the energy efficiency. Simulation results show that the approximation is quite tight and that there is significant improvement in terms of energy efficiency when antenna selection is employed.

*Correspondence

Masoud Arash, Isfahan University of Technology (IUT), Isfahan 8415683111, Iran.
E-mail:M.arash@ec.iut.ac.ir

1. INTRODUCTION

Massive MIMO systems, which employ a very large number of antennas, is one of the promising technologies for fifth-generation (5G) of wireless communication systems [1]. Such transmission techniques are capable of focusing energy into very small regions of space, bringing huge improvements in terms of throughput and energy efficiency (EE). Moreover, the large number of antennas may also help the next generation of wireless systems to manage the growing number of user terminal (UTs), with possible high data rates demands, as is the case for Internet of Things deployments[1, 2]. In addition, massive MIMO may also simplify other parts of a communication system, such as the resource allocation and multiple access layer [3].

On the other hand, worries about global warming and greenhouse gas emissions result in limitations in the total power consumption of the communication systems [4]. This brings an interesting trade-off between the system performance in terms of data rate serving multiple UTs due to the increased number of antennas [1], and the increased radiated power and power consumption [5]. Such a trade-off reflects directly into the EE, which is a parameter that gives a general view of the amount of optimality of a communication system, by encompassing the efficiency of the consumed power relative to the offered data rate [6]. In some works, due to mathematical complexity of EE, an approximation of it is used. For example, in [7] an approximation for EE has been optimized with respect to number of UTs.

In order to reduce the transmit power, some works in the literature proposed power control strategies. For instance, in [8] the authors presented a simplified approach to reduce the complexity of deploying optimal linear precoders, with the goal of ensuring targeted UTs' data rates. In a multi cell massive MIMO scenario, [9] optimized power allocation constrained with a total transmit power budget. In addition, the joint allocation of power and pilot symbols have also been considered, trying to mitigate pilot contamination in massive MIMO in [10], or by using the spectral efficiency as performance metric and setting a total energy budget per coherence interval in [11]. Also, in [12] the authors proposed a joint pilot and data power control scheme to optimize EE subject to each UT's signal-to-interference plus noise ratio (SINR) and power constraints.

Moreover, an important factor that cannot be neglected in the analysis is the power used by the RF chains and signal processing of the transceivers, especially when the number of antennas increases. Neglecting the power

1

consumption of RF chains causes a monotonic increase of the EE relative to the number of antennas. However, as highlighted by [13, 14], EE will not necessarily increase by employing more antennas in the case of massive MIMO systems, which results in an EE optimization problem with respect to the number of antennas and many other system parameters such as the number of UTs, the desired data rate, or other analog devices and residually lossy factors [15]. Then, trying to tackle this problem, the authors in [14] have presented a comprehensive model for the power consumption of a communication system that can be used to accurately analyze the EE of massive MIMO systems.

Furthermore, the complexity of the MIMO transmission schemes may also be an important limitation in some scenarios. Therefore, antenna selection has been adopted in conventional MIMO systems [16], whose idea is to employ only a subset of all available transmit and receive antennas [16] in order to reduce the required number of RF chains, and consequently the energy consumption, as well as the system complexity. Moreover, depending on the optimization goal, the subset of antennas may be formed using different criteria, such as EE, throughput, complexity or *etc*. Recently, the idea of antenna selection has also been analyzed in the massive MIMO context. For instance, the distribution of the mutual information between transmit and/or receive antennas has been derived for a point-to-point massive MIMO communication system employing transmit antenna selection in [17, 5]. However, there is a lack of comprehensive analysis for the effects of antenna selection in the Multi-User (MU) case, which has the greatest potential to be employed in the next generation of wireless systems [18, 2]. As it has been depicted by [5], the big challenge lies in finding an exact relation for the mutual information with antenna selection in the MU case.

In addition, with different approaches to improve the EE, the authors in [19] proposed an algorithm to select a subset of antennas that results in the maximum determinant of the channel matrix. Selecting this submatrix has been shown to be the optimum choice from the capacity point of view in MIMO systems [20], however, the complexity of this method is still very high in massive MIMO systems due to the large number of antennas. In [21], a capacity maximizing algorithm is used to improve EE for two scenarios: fixed or variable power consumption. However, in first scenario complexity of the algorithm is very high in massive MIMO systems and becomes even more when power consumption is assumed to be variable with data rate. Moreover, a limited number of RF chains (much less than the total number of available antennas) has been considered in [20, 5, 22]. Then, the idea is to cycle the RF chains among the whole set of antennas, which is suitable for situations where channel is slow changing enough. The most important limitation of this approach in a massive MIMO framework is that the diversity is limited by the number of RF chains, and not by the number of available antennas. Note that, in this case, number of selected antennas must always be less than the number of available RF chains. Also, another drawback comes in obtaining channel state information (CSI) of all antennas [6], since the cycling of the RF-chains requires more time (or less pilot bits) to achieve full CSI, implying in imperfect or outdated estimation. Therefore, it may be more beneficial to employ the same number of RF-chains as the number of antennas, and then turn off some RF-chains/antennas whenever possible, to operate in a more energy efficient mode.

In this paper, we present an algorithm for antenna selection for MU massive MIMO systems, which is cost-effective related to the system dimensions. In order to reduce the complexity of mathematically analysis of EE, we first propose a selection algorithm based on the norm of the channel coefficients, which is also able to provide a good performance in terms of EE. Then, based on the characteristics of the proposed algorithm, we approximate the EE formulation using ordered statistics, which is shown to be very tight with the optimal case. Our results show a significant improvement in terms of EE when using only a subset of antennas. In general, the optimal number of antennas is always slightly higher than the number of UTs, but the relation depends on the required data rate. If the required data rate is too high (in the order of 100 Mbps in our examples), then the algorithm opts to use the complete set of antennas. Moreover, when both the number of employed antennas and the average data rate to the users are optimized, up to 110% EE improvement is obtained based on our system assumptions.

The rest of the paper is organized as follows: In section 2 system model is presented. Power consumption model and EE is formulated in Section 3. In Section 4 our selection algorithm is proposed, with the calculation of the approximated EE given in Section 4.1. Moreover, in Section 5 numerical results will qualify our analysis and show the improvement of EE, achieved due to antenna selection, while the conclusions of this work are given in Section 6.

Notation: Boldface (lower case) is used for column vectors, $\boldsymbol{x}$, and (upper case) for matrices, $\boldsymbol{X}$. $\boldsymbol{X}^H$ and $\boldsymbol{X}^T$ illustrates conjugate transpose and transpose of $\boldsymbol{X}$, respectively. $\mathbb{E}\{.\}$ denotes expectation, $|\,.\,|$ stands for absolute value of a given scalar variable, $\|\,.\,\|$ is Euclidean norm, tr is trace operator, $(x)^+ = \max(x, 0)$ and $\boldsymbol{X}_{(k,k)}$ represents $(k, k)^{\text{th}}$ entry of $\boldsymbol{X}$. Also, $\boldsymbol{I}_K$ is $K \times K$ identity matrix.

2. SYSTEM MODEL

We consider both uplink and downlink of a cellular system, with a single cell consisting of K single antenna UTs and a massive MIMO base station (BS) at the center of the cell. Moreover, we assume that the BS selects F out of M available antennas to serve the UTs. Operational mode is assumed to be Time-Division-Duplex (TDD), so that the

CSI acquired at the BS can be used in both uplink and downlink transmissions, due to channel reciprocity.

Then, at the BS in the uplink, the received signal vector $\boldsymbol{y}^{(ul)} \in \mathbb{C}^{F \times 1}$ can be expressed as

$$\boldsymbol{y}^{(ul)} = \tilde{\boldsymbol{G}} \boldsymbol{Q}^{(ul)} \boldsymbol{s} + \boldsymbol{n}^{(ul)}, \qquad (1)$$

where $\tilde{\boldsymbol{G}}$ is the $F \times K$ channel matrix between F selected antennas in BS and K UTs, $\boldsymbol{Q}^{(ul)}$ is a $K \times K$ diagonal matrix, whose $(k,k)^{\text{th}}$ element is the square root of the kth UT's allocated power, $\boldsymbol{s} \in \mathbb{C}^{K \times 1}$ is the vector of transmitted symbols, with $\mathbb{E}\{\boldsymbol{s}\} = \boldsymbol{0}$ and $\mathbb{E}\{\boldsymbol{s}\boldsymbol{s}^H\} = \boldsymbol{I}_K$, and $\boldsymbol{n}^{(ul)} \sim \mathcal{CN}(0, \sigma_n^2 \boldsymbol{I}_F)$ is additive white Gaussian noise. $\boldsymbol{y}^{(ul)}$ is then multiplied by combining matrix, $\boldsymbol{U}$.

In the downlink, by considering that the BS employs a precoding matrix $\boldsymbol{V}^T$ in order to simplify detection of the transmitted symbols at UTs, the signal received by the UTs can be written as

$$\boldsymbol{y}^{(dl)} = \tilde{\boldsymbol{D}} \boldsymbol{V}^T \boldsymbol{Q}^{(dl)} \boldsymbol{s} + \boldsymbol{n}^{(dl)}, \qquad (2)$$

where $\tilde{\boldsymbol{D}} = \tilde{\boldsymbol{G}}^T$ is the downlink channel matrix, $\boldsymbol{Q}^{(dl)}$ is the downlink power allocation matrix and $\boldsymbol{n}^{(dl)} \sim \mathcal{CN}(0, \sigma_n^2 \boldsymbol{I}_K)$ is the additive white Gaussian noise at the UTs. Then, due to the dimensions of the system, it is more practical to use linear combiners such as Matched-Filter (MF), Minimum Mean-Squared Error (MMSE) or Zero-Forcing (ZF) due to their lower complexity [1] compared to nonlinear methods such as DPC [23]. Throughout this paper we assume a ZF processing at the BS, which allows us to achieve tractable mathematical expressions with performance close to MMSE [14] as it is shown in the Numerical Result section. Moreover, similar to [14], we employ same precoder and combiner operators to reduce the computational complexity of the system, therefore $\boldsymbol{U} = \boldsymbol{V}$. In this case, the matrix $\boldsymbol{V}$ can be expressed as

$$\boldsymbol{V} = (\tilde{\boldsymbol{G}}^H \tilde{\boldsymbol{G}})^{-1} \tilde{\boldsymbol{G}}^H. \qquad (3)$$

Consider $\boldsymbol{G}$ as a $M \times K$ matrix which models independent and identically distributed (i.i.d.) small-scale and large-scale fading channel, whose elements represent the channel between each of the M antennas of the BS and the each of K UTs, so that we express it as

$$\boldsymbol{G} = \boldsymbol{H} \boldsymbol{B}^{\frac{1}{2}}, \qquad (4)$$

where

$$\boldsymbol{H}(m,k) \sim \mathcal{CN}(0, 2\sigma^2), \qquad (5)$$

models the small-scale fading coefficient between the kth UT, $k \in [1, K]$, and the mth antenna at the BS, $m \in [1, M]$. Let us remark that we assume perfect CSI availability in the BS side. Also, the variance of each real and imaginary part of $\boldsymbol{H}(m,k)$ is equal to σ^2. $\boldsymbol{B}$ is a diagonal $K \times K$ matrix, whose $(k,k)^{\text{th}}$ element represents the kth UT's path-loss coefficient, r_k. Since the BS antennas are usually placed in much smaller dimensions relative to the distance between UTs and the BS, it is

sensible to assume that r_k is the same for all BS antennas. Note that the matrix $\tilde{\boldsymbol{G}}$ in Eq. (1) is the channel matrix using only the F selected antennas ($F \leq M$), thus it is obtained using F appropriately selected rows from $\boldsymbol{G}$, as will be detailed in Section 4.

3. ENERGY EFFICIENCY

Assume R_k (in bit/channel use) as the average achievable information rate of the kth UT and P_{tot} as the total power consumption in transceivers to achieve sum rate of all UTs. Then, according to [13], the average EE is defined as

$$\eta_{\text{E}} = \frac{\sum_{k=1}^K R_k}{P_{\text{tot}}}. \qquad (6)$$

In this paper we assume that all UTs achieve an equal gross rate, $R_k = \bar{R}, k = 1, \cdots, K$. To do so, we employ the power allocation scheme presented in [24], which guarantees a constant rate to the UTs, and which provides closed-form results for the consumed energy. Therefore, despite different channel conditions, all UTs will serve with the same rate $\bar{R}$ and therefore the power consumption of each UT will be a function of $\bar{R}$ and its channel condition. Following [24] for instantaneous uplink power allocation vector, $\boldsymbol{p}^{(ul)}$, we have

$$\boldsymbol{p}^{(ul)} = \sigma_n^2 (\boldsymbol{A}^{(ul)})^{-1} \boldsymbol{1}_K, \qquad (7)$$

in which $\boldsymbol{A}^{(ul)}$ is defined as

$$\boldsymbol{A}_{(k,l)}^{(ul)} = \begin{cases} \frac{1}{\|\boldsymbol{v}_k\|^2 (2^{\frac{\bar{R}}{BW}} - 1)} & k = l, \\ 0 & k \neq l, \end{cases} \qquad (8)$$

where $\boldsymbol{v}_k$ is the kth column of matrix $\boldsymbol{V}^T$, BW is the transmission bandwidth and $\boldsymbol{1}_K$ is $K \times 1$ vector whose all elements are equal to one. Note that $\boldsymbol{p}^{(ul)}(k) = (\boldsymbol{Q}_{(k,k)}^{(ul)})^2$ and $\boldsymbol{A}^{(ul)}$ is a diagonal matrix. Moreover, due to using the ZF scheme in both uplink and downlink, we also have that $\boldsymbol{A}^{(dl)} = (\boldsymbol{A}^{(ul)})^T$ and, as a result, the instantaneous power allocation at the downlink is the same as at the uplink, i.e., $\boldsymbol{p}^{(dl)} = \boldsymbol{p}^{(ul)}$ and $\boldsymbol{p}^{(dl)}(k) = (\boldsymbol{Q}_{(k,k)}^{(dl)})^2$ [24].

It is worth noting that by employing this power allocation scheme one can specify an equal rate for all users, which can be constant or a function of other system requirements. For example, the authors in [14] considered a variable $\bar{R}$, which is a logarithmic function of number of BS antennas and number of UTs. Using this assumption the data rate of all UTs will be affected when a new UT is added, or disconnected from the network. Thus, in this paper we assume that all users have an arbitrary and equal rate $\bar{R}$, while we maintain our formulation for the general case.

We formulate the total power consumption by dividing it into three distinct parts: *i.)* the emitted power, P_{e}, which depends on the coefficients of transmitted signal

and is computed as the absolute transmitted power; *ii.)* the power that the BS and the UTs consume to generate and process the transmitted signals, P_{process}, which is a function of number of UTs, number of operational antennas and processing algorithms for precoding and combining; and *iii.)* P_{fix}, which is the constant power consumed by cooling systems, backhaul infrastructure, *etc.*

First, the average power consumption due to emitted power in the downlink can be calculated as [14]

$$P_{\text{e}}^{(dl)} = BW\sigma_n^2 \, \mathbb{E}\{\mathbf{1}_K^T (\boldsymbol{A}^{(dl)})^{-1}\mathbf{1}_K\}, \qquad (9)$$

where expectation is over both small-scale and large-scale fadings. Moreover, since the power allocation matrix in the uplink is the same as in the downlink in the case of the ZF scheme, we can conclude that $P_{\text{e}}^{(ul)} = P_{\text{e}}^{(dl)}$. Then, by assuming half of the coherence block for each uplink and downlink phases, the emitted power will be

$$P_{\text{e}} = \frac{1}{2}P_{\text{e}}^{(ul)} + \frac{1}{2}P_{\text{e}}^{(dl)} = P_{\text{e}}^{(dl)}. \qquad (10)$$

Notice that P_{e} can be expressed analytically as a function of K and M by using the Wishart theorem as in [14]. However, as pointed out by [5], it is too hard to find an exact expression for (9) when antenna selection is employed. Nevertheless, in Section 4.1 we derive an approximation for P_{e} to be used in the case of antenna selection.

Next, the power used to process the transmitted signals, P_{process}, can be decomposed into four main components as

$$P_{\text{process}} = P_{\text{BB}} + P_{\text{RF}} + P_{\text{CSI}} + P_{\text{LP}}, \qquad (11)$$

where P_{BB} is the power consumption due to baseband operations for coding and decoding, P_{RF} is the RF circuitry power consumption, P_{CSI} is used for channel estimation, and P_{LP} encompasses linear precoding operations. Then, building up upon the model presented in [14], we have that

$$P_{\text{BB}} = (P_{\text{cod}} + P_{\text{dec}})K, \qquad (12)$$

$$P_{\text{RF}} = \frac{1}{2}\underbrace{(FP_{\text{tx}} + KP_{\text{rx}})}_{\text{Downlink}} + \frac{1}{2}\underbrace{(KP_{\text{tx}} + FP_{\text{rx}})}_{\text{Uplink}}, \qquad (13)$$

$$P_{\text{CSI}} = M\frac{K}{LT}, \qquad (14)$$

$$P_{\text{LP}} = \frac{9K^2F + 6KF + 2K^3}{3LT} + \left(1 - \frac{K}{T}\right)\frac{FK}{L}, \qquad (15)$$

where P_{cod} and P_{dec} are constants, respectively denoting the power required for coding or decoding a symbol, while P_{tx} and P_{rx} are the constant powers used at transmitter and receiver RF-chains, respectively. Notice that we assume equal parts for downlink and uplink in each coherence block, so that a factor of $\frac{1}{2}$ appears for each term in (13). Moreover, L is the computational efficiency of the operations per Joule, and T is the coherence time of channel.

Rewriting (11) in terms of the dependence on F or K we have,

$$P_{\text{process}} = \sum_{i=1}^{3} C_{i,0}K^i + \sum_{i=0}^{2} C_{i,1}K^i F, \qquad (16)$$

in which $C_{1,0} = P_{\text{cod}} + P_{\text{dec}} + P_{\text{rx}} + \frac{M}{LT}$, $C_{2,0} = 0$, $C_{3,0} = \frac{2}{3LT}$, $C_{0,1} = P_{\text{tx}}$, $C_{1,1} = \frac{3}{LT} + \frac{1}{L}$ and $C_{2,1} = \frac{2}{LT}$.

Then, to sum up, the total power consumption can be written as

$$P_{\text{tot}} = P_{\text{e}} + P_{\text{process}} + P_{\text{fix}}. \qquad (17)$$

Finally, assuming that all UTs are served with constant rate, $\bar{R}$, the EE can be written as

$$\eta_{\text{E}} = \frac{K\bar{R}}{P_{\text{e}} + \sum_{i=0}^{3} C_{i,0}K^i + \sum_{i=0}^{2} C_{i,1}K^i F + P_{\text{fix}}}, \qquad (18)$$

where notice that P_{e}, as defined in (9), is the sole term without a closed-form expression, which prevents us to find a deterministic form for the EE. Then, in the following we propose an algorithm for antenna selection and use its characteristics to provide a tight approximated formulation for P_e.

4. ANTENNA SELECTION ALGORITHM

From both diversity order and multiplexing gain point of views, the system always benefits from increasing the number of antennas. However, a few trade-offs are observed in terms of the EE of massive MIMO systems. For instance, it has been shown in [14, 13] that the EE tends to zero when the number of UTs is equal to the number of antennas ($K = M$) using the ZF precoder, or when $M \to \infty$. Thus, it can be concluded that the optimum EE may be achieved for some F in $K < F \leq M < \infty$, so that the system may employ an optimum number $F^\star$ out of the pool of M available antennas. In the literature, some works such as [14] have optimized the EE with respect to M. However, in an operating massive MIMO system the number of antennas, M, is fixed. Therefore, to improve EE we propose to select a subset of antennas, still taking advantage of the massive MIMO transmission scheme, but consuming less energy. Then, since the mathematical analysis in antenna selection case is usually hard to be obtained [5], our main goal is to provide a low-complex antenna selection algorithm, which also allows us to provide a closed-form approximation to the EE formulation.

In order to select the best set of antennas, several criteria have been proposed for conventional MIMO systems. For instance, the channel strength criterion results in maximum SNR; however, maximizing the SNR does not necessarily maximize the EE of the system [16]. Another criterion that can be employed is to maximize the Frobenius norm, which has been shown in [25] to maximize SNR at the

same time that it minimizes the instantaneous probability of error. Therefore, our approach is to select a subset of best antennas in the sense of maximizing the average absolute value of the channel coefficients. Notice that this method it not necessarily optimal in terms of EE, but its computational complexity is linear with number of UTs and BS antennas.

The antenna selection algorithm is presented in Table I. First we employ the absolute value operator to form the vector $\boldsymbol{a}$, whose m^{th} element contains the sum of the absolute values of the channel coefficients between all UTs and the mth BS antenna. Note that we employ $\boldsymbol{H}$, as defined in (5), rather than $\boldsymbol{G}$ to compute the elements of $\boldsymbol{a}$, since we are selecting the antennas based on their fading channel strength only. Next, in each step the maximum value of this vector is determined and the corresponding antenna is added to the subset $\boldsymbol{s}$, representing the selected antennas. Then, after each new inclusion in the subset $\boldsymbol{s}$, the EE is computed using (18) and compared to the previous EE. This process continues while the overall EE increases with the inclusion of new elements and the algorithm stops whenever EE at the step n is lower than that at the step $n-1$. Finally, the algorithm returns the subset $\boldsymbol{s}$ of the selected antennas as well as $F = \text{card}(\boldsymbol{s})$, where card(.) represents the cardinality of the set.

Moreover, let us remark that the algorithm in Table I requires the numerical computation of (18), since the emitted power P_{e} is not in closed-form. However, due to the use of the Frobenius norm, we are able to provide an approximation for P_{e}, and consequently simplify the antenna selection algorithm.

4.1. Approximation of P_e When Employing Antenna Selection

According to [14], the Wishart theorem can be used to write P_{e} as a function of K and M when all antennas are used. Let us denote by $\tilde{\boldsymbol{H}}$ the $F \times K$ matrix of selected small-fading coefficients, which is obtained in the same way as $\tilde{\boldsymbol{G}}$, i.e by using F appropriately selected rows from $\boldsymbol{H}$. However, as pointed out by [5], it is quite complex to solve (9) in closed-form when only a subset of the M antennas is used, since the the distribution of the elements of $\tilde{\boldsymbol{H}}$ is no longer Gaussian in the case of antenna selection.

Then, in order to solve this problem, we propose to approximate $\tilde{\boldsymbol{H}}$ using an i.i.d. complex Gaussian matrix $\hat{\boldsymbol{H}}$, so that

$$\hat{\boldsymbol{H}}(f,k) \sim \mathcal{CN}(0, 2\hat{\sigma}^2), \quad f \in [1,F], \ k \in [1,K], \quad (19)$$

where variance of each real and imaginary part of $\hat{\boldsymbol{H}}(f,k)$ is equal to $\hat{\sigma}^2$. Notice that this model has a single parameter to be determined, so that we only need to find $\hat{\sigma}^2$ analytically. Then, in order to achieve a fair approximation for $\tilde{\boldsymbol{H}}$, we first constrain the energies of $\tilde{\boldsymbol{H}}$ and $\hat{\boldsymbol{H}}$ to be

equal, i.e.,

$$\sum_{f=1}^{F} \sum_{k=1}^{K} \mathbb{E}\{| \ \tilde{\boldsymbol{H}}(f,k) \ |^2\} = \sum_{f=1}^{F} \sum_{k=1}^{K} \mathbb{E}\{| \ \hat{\boldsymbol{H}}(f,k) \ |^2\}$$
$$\overset{(\star)}{=} 2FK\hat{\sigma}^2, \quad (20)$$

where $(\star)$ is concluded due to the independence among the elements of $\hat{\boldsymbol{H}}$. Moreover, the left hand side of (20) can be written as

$$\sum_{f=1}^{F} \sum_{k=1}^{K} \mathbb{E}\{| \ \tilde{\boldsymbol{H}}(f,k) \ |^2\} = \sum_{f=1}^{F} \mathbb{E}\{\sum_{k=1}^{K} | \ \tilde{\boldsymbol{H}}(f,k) \ |^2\}$$
$$= \sum_{f=1}^{F} \mathbb{E}\{a_{f:M}\}, \quad (21)$$

where $a_{f:M}$ denotes the fth greatest value in $\boldsymbol{a}$ out of M, so that the sum of the F largest values of $\boldsymbol{a}$ is required in (21) to obtain $\hat{\sigma}^2$. In this case, due to antenna selection algorithm presented in Table I, the mean of entries of $\boldsymbol{a}$ can be achieved using order statistics.

Regarding the distribution of the matrix $\boldsymbol{H}$, we know that $|\boldsymbol{H}(f,k)|^2$ has an exponential distribution for all values of f, k, i.e.,

$$| \ \boldsymbol{H}(m,k) \ |^2 \sim \mathcal{E}(2\sigma^2), \ m = 1, \dots, M, \ k = 1, \dots, K. \quad (22)$$

Thus, according to Table I each element of the $M \times 1$ vector $\boldsymbol{a}$ is the sum of K independent exponentially distributed random variables, which in turn results in a Gamma distributed random variable [26]

$$a_m \sim \mathcal{G}(K, 2\sigma^2), \quad m = 1, \dots, M. \quad (23)$$

In the following step, we must choose the F largest values of $\boldsymbol{a}$. To that end we follow [27], which has shown that for M i.i.d. realizations of a random variable z, with probability density function (pdf) given by $P(z)$, and cumulative distribution function given by $C(z)$, we have the mean of the rth greatest realization, out of M, given by

$$\mu_{r:M} =$$
$$\frac{M!}{(r-1)!(n-r)!} \int_{-\infty}^{\infty} zP(z)C^{r-1}(z)(1-C(z))^{(M-r)}dz, \quad (24)$$

which can be achieved in closed-form for a few specific distributions, such as the Uniform distribution [27], but not for the case of the Gamma distribution.

Nevertheless, by employing (24) we can obtain the mean of the F largest values in $\boldsymbol{a}$, and we resort to the following theorem to find an approximate value for $\sum_{f=1}^{F} \mathbb{E}\{a_{f:M}\}$.

Theorem 1
For jointly distributed random variables $Z_1, Z_2, \dots, Z_M$ with means $\mu_i = \mu$, and variances $\sigma_{iz}^2 = \sigma_z^2$, $i =$

Table I. Antenna selection algorithm

Algorithm steps	Calculation order
$\boldsymbol{a} := [a_1 \; a_2 \; \ldots \; a_M]$	
for $m := 1$ **to** M	
$\quad a_m = \sum_{k=1}^{K} \mid \boldsymbol{H}(m,k) \mid^2$	$O(K)$
end	
$\eta_{\mathrm{E}}(0) := 0$	
for $n := 1$ **to** M	
$\quad u := \arg\max \boldsymbol{a}$	$O(M)$
$\quad \boldsymbol{s}(n) := u$	
$\quad \boldsymbol{a_u} := 0$	
$\quad$ compute $\eta_{\mathrm{E}}(n)$ using (18) using only the subset $\boldsymbol{s}$	
$\quad$ **if** $\eta_{\mathrm{E}}(n) < \eta_{\mathrm{E}}(n-1)$	
$\quad\quad$ delete $\boldsymbol{s}(n)$	
$\quad\quad$ exit	
$\quad$ **end**	
end	
$F = \mathrm{card}\,(\boldsymbol{s})$	
return $F, \boldsymbol{s}$	

$1, \ldots, M$, we have

$$\left| \sum_{i=1}^{M} \lambda_i(\mu_{i:M} - \mu) \right| \le \sigma_z \sqrt{M \sum_{i=1}^{M} (\lambda_i - \bar{\lambda})^2}, \quad (25)$$

for any $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_M] \in \mathbb{R}^M$, where $\bar{\lambda} = \frac{1}{M} \sum_{i=1}^{M} \lambda_i$.

Proof
Please see [29]. □

Assuming that the bound presented in Theorem 1 is tight [29], the following corollary can be employed to derive an upper bound for $\sum_{f=1}^{F} \mathbb{E}\{a_{f:M}\}$.

Corollary 1
According to (25), the mean $\sum_{f=1}^{F} \mathbb{E}\{a_{f:M}\}$ is upper bounded by

$$\sum_{f=1}^{F} \mathbb{E}\{a_{f:M}\} \le 2K\sigma^2 + 2\sigma^2 \sqrt{K \frac{M-F}{F}}. \quad (26)$$

Proof
Please refer to Appendix A. □

Due to the tightness of this upper bound, as the numerical results will show, in the rest of this paper we assume that

$$\sum_{f=1}^{F} \mathbb{E}\{a_{f:M}\} \approx 2K\sigma^2 + 2\sigma^2 \sqrt{K \frac{M-F}{F}}. \quad (27)$$

Then, putting (27) into (20), $\hat{\sigma}^2$ is finally achieved as

$$\hat{\sigma}^2 = \sigma^2 \left(1 + \sqrt{\frac{M-F}{FK}} \right). \quad (28)$$

Moreover, using this Gaussian approximation for the selected antenna matrix, we can also approximate $E\{tr(\tilde{\boldsymbol{H}}^H \tilde{\boldsymbol{H}})^{-1}\}$, which is needed to calculate P_{e}. Based on the Wishart theorem [30] we know that

$$\mathbb{E}\{tr(\hat{\boldsymbol{H}}^H \hat{\boldsymbol{H}})^{-1}\} = \frac{K}{F-K} \frac{1}{2\sigma^2(1 + \sqrt{\frac{M-F}{FK}})}. \quad (29)$$

Then, since the matrix $\tilde{\boldsymbol{H}}$ is approximated by $\hat{\boldsymbol{H}}$,

$$\mathbb{E}\{tr(\tilde{\boldsymbol{H}}^H \tilde{\boldsymbol{H}})^{-1}\} \approx \mathbb{E}\{tr(\hat{\boldsymbol{H}}^H \hat{\boldsymbol{H}})^{-1}\}$$
$$= \frac{K}{F-K} \frac{1}{2\sigma^2(1 + \sqrt{\frac{M-F}{FK}})}. \quad (30)$$

As we can observe, the antenna selection algorithm introduces the term $(1 + \sqrt{\frac{M-F}{FK}})$ to the denominator of (30), comparing to the case when the system employs all M antennas. This term indicates that there is an optimization to be performed in F, since $\mathbb{E}\{tr(\tilde{\boldsymbol{H}}^H \tilde{\boldsymbol{H}})^{-1}\}$, which is directly related to the transmitted power, is not an increasing function of F. This equation shows that $E\{tr(\tilde{\boldsymbol{H}}^H \tilde{\boldsymbol{H}})^{-1}\}$ decreases as the number of selected antennas F increases; however, the speed of this decrease also depends on number of UTs, K. In addition, let us remark that the tightness of the approximation provided by (30) will be validated in Section 5.

Corollary 2
Using (30), an approximation for the emitted power can be written as

$$\hat{P}_{\mathrm{e}} = \frac{BW\sigma_n^2 K \mathbb{E}\{r^{-1}\}(e^{\frac{\bar{R}}{BW}} - 1)}{(F-K)2\sigma^2(1 + \sqrt{\frac{M-F}{FK}})}, \quad (31)$$

where r models the path-loss attenuation[*] and thus $\mathbb{E}\{r^{-1}\}$ is not a function of neither F nor K.

Proof
Please refer to Appendix B. □

4.2. Approximated Antenna Selection Algorithm

Using the approximation of Corollary 2, we can simplify (18) so that it can be approximated in closed-form by

$$\hat{\eta}_{\mathsf{E}} =$$

$$\frac{K\bar{R}}{\frac{BW\sigma_n^2 K \mathbb{E}\{r^{-1}\}(e^{\frac{\bar{R}}{BW}}-1)}{(F-K)2\sigma^2(1+\sqrt{\frac{M-F}{FK}})} + \sum_{i=0}^{3} C_{i,0}K^i + \sum_{i=0}^{2} C_{i,1}K^i F + P_{\text{fix}}},$$
(32)

so that the proposed simplified algorithm can be easily obtained from Table I by replacing η_{E} by $\hat{\eta}_{\mathsf{E}}$, using (32).

Remark 1
Assuming that the UTs are uniformly distributed in a circular cell with radius $d_{\max}$, the path-loss attenuation is given by

$$r(d) = \frac{\bar{d}}{\parallel d \parallel^{\kappa}},$$
(33)

where d is the distance between the user and the BS, κ is the path-loss exponent and the constant $\bar{d}$ regulates the channel attenuation at a reference distance $d_{\min}$. In this case, the $\mathbb{E}\{r^{-1}\}$ can be calculated as [14]

$$\mathbb{E}\{r^{-1}\} = \frac{d_{\max}^{\kappa+2} - d_{\min}^{\kappa+2}}{\bar{d}\left(1+\frac{\kappa}{2}\right)(d_{\max}^2 - d_{\min}^2)}.$$
(34)

Throughout this paper we assume that the UTs are uniformly distributed in a circular cell, so that $\mathbb{E}\{r^{-1}\}$ is given by (34). This assumption allows us to provide a simple closed-form expression for $\hat{\eta}_{\mathsf{E}}$, whose optimum point with respect to F can be calculated using the Newton's method or with a simple one dimensional global search, which we employ in next section to provide a few numerical examples.

Remark 2
In addition to F, (32) also suggests that the system could benefit from the optimization of $\bar{R}$, *i.e.*, the UTs data rate. Since $\bar{R}$ and F are independent variables, one dimensional global search can also be employed to verify its effect on the maximization of the EE.
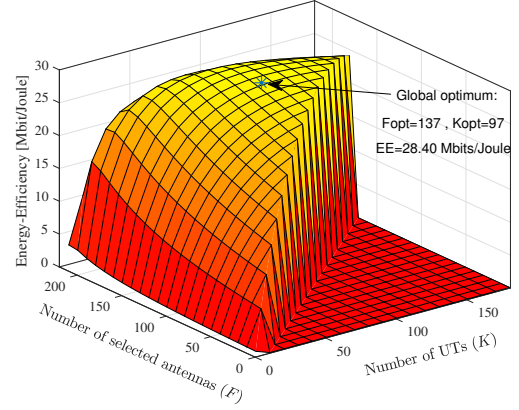
_____
[*]Assuming that the large-scale fading is dominated by the path-loss.

Figure 1. EE as a function of K and F using the proposed approximated algorithm, obtained employing (32). The global optimum point is also shown, for which $K = 97$, $F = 137$ and $\hat{\eta}_{\mathsf{E}} = 28.40$ Mbits/J.

5. NUMERICAL RESULTS

In this section we numerically evaluate the proposed algorithm, using the approximated approach depicted in Section 4.2, as well as the Monte Carlo simulations for the algorithm in Table I. The results are obtained employing 2000 Monte Carlo iterations. Moreover, we use the same parameters for power consumption as in [14], which are summarized in Table II. Also, the 3GPP distance-dependent path-loss model is used and the total number of available antennas is considered to be $M = 220$ in all scenarios. All simulation results are provided for ZF case, unless otherwise stated.

Fig. 1 shows the general behavior of the approximated proposed algorithm, using (32) to obtain the optimum $\bar{R}$ using a global search. Therefore, the maximal $\hat{\eta}_{\mathsf{E}}$ is illustrated for each (F, K) pair. As we can observe, $\hat{\eta}_{\mathsf{E}} = 0$ in the scenarios when $F < K$, so that the UTs cannot be served in this case, and the EE is maximized when the number of selected antennas is in the range $K \leq F \leq M$. The global optimum point, which is marked in Fig. 1, is achieved when $F = 137$ and $K = 97$, yielding $\eta_{\mathsf{E}} = 28.40$ Mbits/J.

Fig. 2 investigates the accuracy of Eq. (31) by comparing the Monte Carlo simulated P_e with approximated one in (31). In this figure, $\bar{R}/BW = 5(bits/s/Hz)$ and simulation is done for $K = 30$, $K = 90$ and $K = 150$. It is seen that the approximation for P_e is quite tight and thus is a reasonable alternative for it. Therefore, since all other parameters are deterministic in Eq. (32), the proposed approximated EE also leads to very tight results with respect to the Monte Carlo simulations. Also, it is observed through simulations that as F grows toward M, the approximation becomes more accurate.

Next, Fig. 3 shows the empirical eigenvalue distributions of the selected matrix of small-scale coefficients

Parameter	Value	Parameter	Value
Minimum UT distance: $d_{\min}$	35 m	Maximum UT distance: $d_{\max}$	250 m
Path-loss at distance d: $r(d)$	$\frac{10^{-3.53}}{r^{3.76}}$	Power for coding: P_{cod}	4 W
Coherence bandwidth: BW	180 kHz	Power for decoding: P_{dec}	0.5 W
Coherence time: T	32 ms	Power of the RF-chain at TX: P_{tx}	1 W
Noise variance: σ_n^2	10^{-20} J/c.u.	Power of the RF-chain at RX: P_{rx}	0.3 W
Channel coefficients variance: $2\sigma^2$	1	Constant power consumption: P_{fix}	18 W
Operations efficiency: L	10^9/Joule		
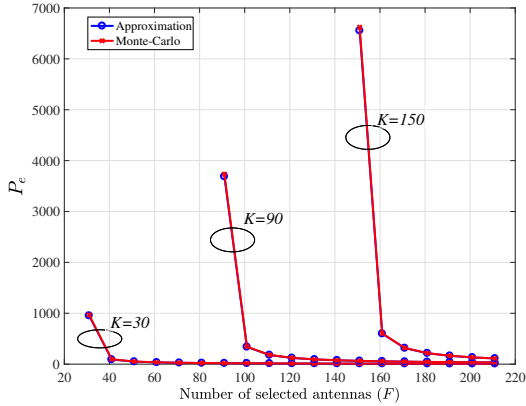
Figure 2. Comparison between the Monte Carlo simulation of P_e with the approximated one in Eq. (31) for $K \in \{30, 90, 150\}$ when $\bar{R}/BW = 5(bits/s/Hz)$.
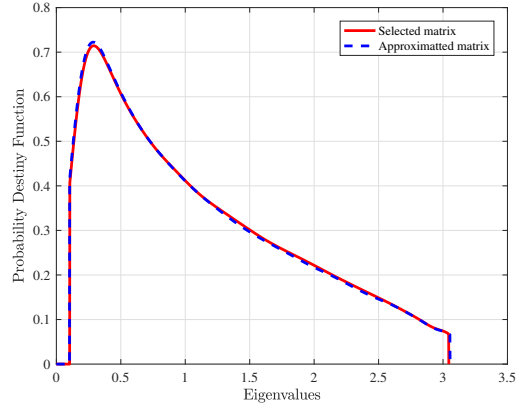

Figure 3. Empirical eigenvalue distribution of the selected matrix ($\tilde{\boldsymbol{H}}^*\tilde{\boldsymbol{H}}$) and its proposed approximation ($\hat{\boldsymbol{H}}^*\hat{\boldsymbol{H}}$), for $F = 140$ and $K = 70$.

($\tilde{\boldsymbol{H}}^*\tilde{\boldsymbol{H}}$) and its proposed approximation ($\hat{\boldsymbol{H}}^*\hat{\boldsymbol{H}}$) for $F = 140$ and $K = 70$. The choice for the empirical eigenvalue distribution is due to its good reliability, and we have used a Kernel density estimator [30] to plot Fig. 3. Moreover, it is known that assuming $\hat{\boldsymbol{H}}$ as an i.i.d. Gaussian matrix produces eigenvalues following a Marčenko-Pastur distribution [30]

$$f(x) = (1 - \frac{F}{K})^+ \delta(x) + \frac{K\sqrt{(x-\alpha)^+(\beta-x)^+}}{2\pi F x \hat{\sigma}^2},$$
(35)

where $\delta(x)$ is Dirac delta function, $\alpha = \hat{\sigma}^2(1 - \sqrt{F/K})^2$ and $\beta = \hat{\sigma}^2(1 + \sqrt{F/K})^2$. As it can be observed, the distribution of the eigenvalues of $\tilde{\boldsymbol{H}}^*\tilde{\boldsymbol{H}}$ is very similar to that of the eigenvalues of $\hat{\boldsymbol{H}}^*\hat{\boldsymbol{H}}$, showing that $\hat{\boldsymbol{H}}$ is a good approximation to $\tilde{\boldsymbol{H}}$.

Fig. 4 plots the maximal EE as a function of the number of UTs. For each K, the optimization of both $\bar{R}$ and F is carried out and we compare the Monte Carlo simulation of the antenna selection algorithm, the approximated antenna selection algorithm of Section 4.2, as well as the case when $F = M$ (without antenna selection) with the optimization of $\bar{R}$. The first observation is with respect to the tightness of our approximations in deriving (32), which is very close to the optimal curve obtained through Monte Carlo simulations. Moreover, this figure also shows a significant

improvement in terms of EE when using only a subset of antennas, especially in the cases when the number of users is much lower than the number of antennas. For example, with $K = 20$ UTs, the EE is about 110% better than when the whole set of antennas is employed. In addition, when the number of UTs increases the optimal F also increases, so that for $K > 160$ the optimum number of selected antennas is the same as the total number of available antennas; thus, both scenarios with and without antenna selection have the same performance in terms of EE.

Fig. 5 represents the EE for MF and MMSE, both of them obtained through Monte Carlo simulation. As it can be seen comparing both results, EE is improved up to 60% using the proposed antenna selection with MF (compared to the case without antenna selection), and up to 88% using the proposed antenna selection with MMSE. However, a big difference is observed when comparing the magnitude of the EE for MF and MMSE, which is mainly caused by a larger amount of interference in the case of MF. From our simulations we also notice that, due to the higher interference, MF tends to select the number of antennas equal to K, even for high number of UTs, jeopardizing the EE. On the other hand, MMSE shows very similar performance compared to ZF, as suggested in Section 2, with MMSE achieving slightly better EE, while
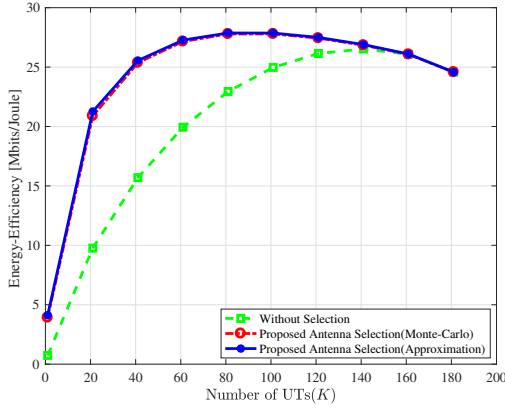
Figure 4. EE of the Monte Carlo simulation for the antenna selection algorithm with optimal F, the approximated antenna selection algorithm with optimal F, as well as the case when $F = M$ (without antenna selection). Moreover, $\bar{R}$ is also optimized for each scenario.
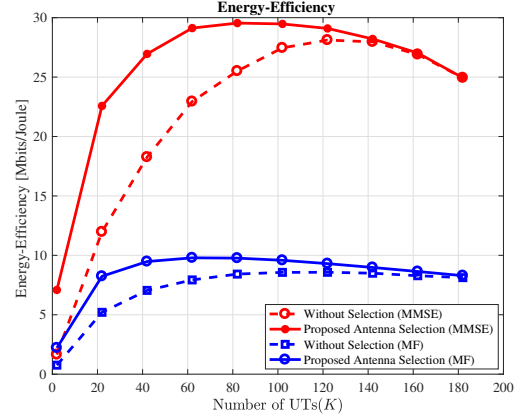

Figure 5. EE of Monte Carlo simulation for the proposed antenna selection algorithm and without selection case for MMSE and MF. Optimization is done for both F and $\bar{R}$.

ZF exhibits better improvement when using the proposed antenna selection scheme. Nevertheless, when EE achieves it optimal point, the corresponding data rates and number of selected antennas for ZF and MMSE are very close, while MF has at least 3 times lower data rates compared to the other schemes.

Fig. 6 shows EE for different values of $\bar{R}$, fixed at 40 Mbits/s, 70 Mbits/s and 100 Mbits/s, respectively, so that the optimization is done with respect to the number of selected antennas, F, only in ZF case. This scenario represents many practical cases when the operators are interested in setting a constant data rate for the users. As it can be observed, there is a larger gap between the EE of the proposed antenna selection scheme and employing all M antennas for lower data rates, which is because the optimum number of selected antennas is small for lower data rates, consequently increasing the EE. On the other hand, by increasing the required data rate, the optimal number of antennas $F^{\star}$ also increases until it reaches the saturation point of being equal to M.

For the same scenarios as Fig. 6, the optimal number of selected antennas is shown in Fig. 7 as a function of the number of UTs. As we can notice, $F^{\star}$ is always slightly higher than K, but the relation depends on the requirement in terms of $\bar{R}$. Moreover, $F^{\star}$ saturates at the maximum value of 220 (equal to M) for higher data rates, when $\bar{R} = 100$ Mbps in this case, for which the EE is the same as in the non-selection scheme. Let us remark that we assume that K is not a parameter to be optimized in this work, since it depends on the particular network scenario. Thus, Fig. 7 provides $F^{\star}$ as a function K.

Finally, Fig. 8 investigates the EE as a function of the system spectral efficiency (defined as $\bar{R}/BW$ in bits/s/Hz [6]), in the case when the number of UTs is $K = 90$ and F is optimized in order to maximize the EE. A
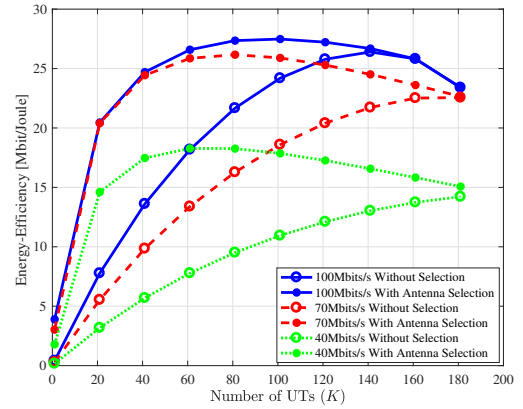

Figure 6. EE of the Monte Carlo simulation for the antenna selection algorithm with optimal F, the approximated antenna selection algorithm with optimal F, as well as the case when $F = M$ (without antenna selection). The data rate is assumed to be fixed at $\bar{R} = 40, 70$ and 100 $Mbits/s$.

remarkable improvement in terms of EE can be observed with the proposed antenna selection scheme. Moreover, the gap between using antenna selection or not using it, decreases for high spectral efficiency, which is due to the requirement for more antennas to serve higher data rates while there exists a limited number of available antennas. It is worth noting that more than 30% of improvement in terms of EE can be achieved when the system operates either with 3.7 bits/s/Hz or 2 bits/s/Hz, which are the spectral efficiency targets for LTE-advanced systems at the downlink and uplink, respectively [31].
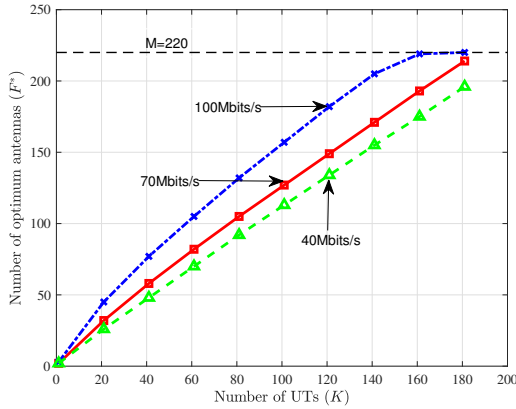
Figure 7. Optimum number of selected antennas ($F^\star$) as a function of the number of UTs (K) for different values of $\bar{R}$
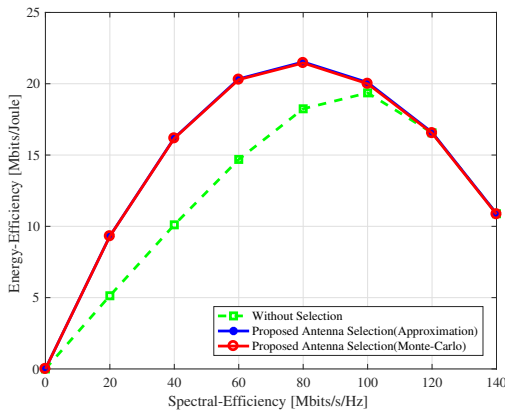

Figure 8. EE as a function of the spectral efficiency. For each point of the curve, the optimization is performed in terms of F.

6. CONCLUSION

Massive MIMO systems yield high data rates to the user terminal by employing a very large number of antennas. However, as a consequence, the power consumption also increases with the number of antennas, while EE is a vital concern for new generation wireless systems. In this paper we considered a multi-user massive MIMO scenario and we proposed an antenna selection scheme so that the system may operate with a subset of $F \leq M$ antennas in order to maximize the EE. By employing a realistic power consumption model for massive MIMO, we derived a closed-form approximation for the EE formulation ir order to reduce the mathematical complexity, which is shown to be very tight with the optimal results. Simulation results showed that using antenna selection significantly improves the EE, especially when the number of UTs is relatively small with respect to the number of available antennas. Moreover, the optimal number of antennas was shown

to be slightly higher than the number of UTs, while the relation depends on the required data rate. Finally, when both the number of employed antennas and the average data rate to the users are optimized, up to 110% for $K = 20$ improvement in EE is observed.

REFERENCES

1. Hoydis, J., Ten Brink, S., & Debbah, M. (2013). Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?. *IEEE Journal on Selected Areas in Communications*, 31(2), 160-171.

2. Andrews, J. G., Buzzi, S., Choi, W., Hanly, S. V., Lozano, A., Soong, A. C., & Zhang, J. C. (2014). What will 5G be?. *IEEE Journal on Selected Areas in Communications*, 32(6), 1065-1082.

3. Larsson, E., Edfors, O., Tufvesson, F., & Marzetta, T. (2014). Massive MIMO for next generation wireless systems. *IEEE Communications Magazine*, 52(2), 186-195.

4. Imran, MA and Katranaras, E and Auer, G and Blume, O and Giannini, V and Godor, I and Jading, Y and Olsson, M and Sabella, D and Skillermark, P and others.(2011).Energy efficiency analysis of the reference systems, areas of improvements and target breakdown.Tech. Rep. ICT-EARTH deliverable.

5. Li, H., Song, L., & Debbah, M. (2014). Energy efficiency of large-scale multiple antenna systems with transmit antenna selection. *IEEE Transactions on Communications*, 62(2), 638-647.

6. Ngo, H. Q., Larsson, E. G., & Marzetta, T. L. (2013). Energy and spectral efficiency of very large multiuser MIMO systems. *IEEE Transactions on Communications*, 61(4), 1436-1449.

7. Zhao, L., Zheng, K., Long, H. & Zhao, H.(2014). Performance analysis for downlink massive MIMO system with ZF precoding. *Transactions on Emerging Telecommunications Technologies*, 25(12), 1219-1230.

8. Sifaou, H., Kammoun, A., Sanguinetti, L. & Debbah, M.(2016). Polynomial expansion of the precoder for power minimization in large-scale MIMO systems. Communications (ICC), 2016 IEEE International Conference on. IEEE.

9. Chien, V., Björnson, E. & Larsson, E. G.(2016). Joint power allocation and user association optimization for Massive MIMO systems. *IEEE Transactions on Wireless Communications*, 15(9), 6384-6399.

10. Chien, V., Björnson, E. & Larsson, E. G.(2017). Joint Pilot Sequence Design and Power Control for Max-Min Fairness in Uplink Massive MIMO. In International Conference on Communications (ICC), 2017. IEEE.

11. Cheng, H. V., Björnson, E. & Larsson, E. G.(2017). Optimal Pilot and Payload Power Control in Single-Cell Massive MIMO Systems. *IEEE Transactions on*

Signal Processing, 65(9), 2363-2378.

12. Guo, K., Guo, Y. & Ascheid, G.(2015). Energy-efficient uplink power allocation in multi-cell MU-Massive-MIMO systems. European Wireless 2015; 21th European Wireless Conference; Proceedings of. VDE.

13. Mukherjee, S., & Mohammed, S. K. (2014). Energy-Spectral efficiency trade-off for a massive SU-MIMO system with transceiver power consumption. *arXiv preprint arXiv:1410.5240*.

14. Björnson, E., Sanguinetti, L., Hoydis, J., & Debbah, M. (2015). Optimal design of energy-efficient multi-user MIMO systems: Is massive MIMO the answer? *IEEE Transactions on Wireless Communications*, 14(6), 3059-3075.

15. Ha, D., Lee, K. & Kang, J.(2013). Energy Efficiency Analysis with Circuit Power Consumption in Massive MIMO Systems. In International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), 2013. IEEE.

16. Ghrayeb, A. (2006, March). A survey on antenna selection for MIMO communication systems. In Information and Communication Technologies, 2006. ICTTA'06. 2nd (Vol. 2, pp. 2104-2109). IEEE.

17. Hesami, P., & Laneman, J. N. (2011, March). Limiting behavior of receive antennae selection. In Information Sciences and Systems (CISS), 2011 45th Annual Conference on (pp. 1-6). IEEE.

18. Larsson, E., Edfors, O., Tufvesson, F., & Marzetta, T. (2014). Massive MIMO for next generation wireless systems. *IEEE Communications Magazine*, 52(2), 186-195.

19. Fang, B., Qian, Z., Shao, W., & Zhong, W. (2015). RAISE: A new fast transmit antenna selection algorithm for massive MIMO systems. *Wireless Personal Communications*, 80(3), 1147-1157.

20. Molisch, A. F., Win, M. Z., Choi, Y. S., & Winters, J. H. (2005). Capacity of MIMO systems with antenna selection. *IEEE Transactions on Wireless Communications*, 4(4), 1759-1772.

21. Bai, Q. & Nossek, J.A. (2015). Energy efficiency maximization for 5G multiantenna receivers. *Transactions on Emerging Telecommunications Technologies*, 26(1), 3-14.

22. Dong, K., Prasad, N., Wang, X., & Zhu, S. (2011). Adaptive antenna selection and Tx/Rx beamforming for large-scale MIMO systems in 60 GHz channels. *EURASIP Journal on Wireless Communications and Networking*, 2011(1), 1-14.

23. Jindal, N., & Goldsmith, A. (2005). Dirty-paper coding versus TDMA for MIMO broadcast channels. *IEEE Transactions on Information Theory*, 51(5), 1783-1794.

24. Wiesel, A., Eldar, Y. C., & Shamai, S. (2006). Linear precoding via conic optimization for fixed MIMO receivers. *IEEE Transactions on Signal Processing*, 54(1), 161-176.

25. Gore, D., & Paulraj, A. J. (2002). MIMO antenna subset selection with space-time coding. *IEEE Transactions on Signal Processing*, 50(10), 2580-2588.

26. Papoulis, A., & Pillai, S. U. (2002). *Probability, random variables, and stochastic processes*. Tata McGraw-Hill Education.

27. David, H. A., & Nagaraja, H. N. (2003). *Order statistics*. John Wiley & Sons, Inc.

28. Maiwald, D & Kraus, D. (2000). Calculation of moments of complex Wishart and complex inverse Wishart distributed matrices. *IEE Proceedings on Radar, Sonar and Navigation*, 147(4), 162-168.

29. Arnold, B., & Balakrishnan, N. (2012). *Relations, bounds and approximations for order statistics (Vol. 53)*. Springer Science & Business Media.

30. Tulino, A. M., & Verdú, S. (2004). *Random matrix theory and wireless communications (Vol. 1)*. Now Publishers Inc.

31. Rumney, Moray and others. (2013). *LTE and the evolution to 4G wireless: Design and measurement challenges*. John Wiley & Sons.

A. PROOF OF COROLLARY 1

Using a triangle inequality in (25) we have that

$$\sum_{i=1}^{M} \lambda_i \mu_{i:M} - \left| \sum_{i=1}^{M} \lambda_i \mu \right| \leq \left| \sum_{i=1}^{M} \lambda_i \mu_{i:M} \right| - \left| \sum_{i=1}^{M} \lambda_i \mu \right|$$

$$\leq \left| \sum_{i=1}^{M} \lambda_i (\mu_{i:M} - \mu) \right| \leq \sigma \sqrt{M \sum_{i=1}^{M} (\lambda_i - \bar{\lambda})^2}. \quad (36)$$

Then, noting that

$$\mu = 2K\sigma^2 > 0 \quad , \quad \sigma_z^2 = 4K\sigma^4,$$

we would have

$$\sum_{i=1}^{M} \lambda_i \mu_{i:M} \leq \sum_{i=1}^{M} \lambda_i 2K\sigma^2 + 2\sigma^2 \sqrt{KM \sum_{i=1}^{M} (\lambda_i - \bar{\lambda})^2}. \quad (37)$$

Now let us choose λ_i as the following

$$\lambda_i = \begin{cases} 1 & \text{if the } i^{\text{th}} \text{ antenna is selected} \\ 0 & \text{otherwise,} \end{cases} \quad (38)$$

so that $\bar{\lambda} = \frac{F}{M}$.

By replacing (38) into the right hand side of (37) we have

$$\sum_{i=1}^{M} \lambda_i 2K\sigma^2 + 2\sigma^2 \sqrt{KM \sum_{i=1}^{M} (\lambda_i - \bar{\lambda})^2}$$

$$= 2FK\sigma^2 + 2\sigma^2 \sqrt{KM \left[F \left(1 - \frac{F}{M} \right)^2 + (M - F) \left(-\frac{F}{M} \right)^2 \right]}$$

$$= 2FK\sigma^2 + 2\sigma^2 \sqrt{K[F(M - F)]}. \quad (39)$$

Finally, inserting (39) into (37), replacing λ_i by $E\{a_{f:M}\}$ and dividing both sides of the inequality by F yields (26), concluding the proof. $\qquad\square$

B. PROOF OF COROLLARY 2

Taking into account that $\boldsymbol{A}^{(dl)}$ is a diagonal matrix we can write

$$\mathbb{E}\{\mathbf{1}_K^T(\boldsymbol{A}^{(dl)})^{-1}\mathbf{1}_K\} = \mathbb{E}\{tr(\boldsymbol{A}^{(dl)})^{-1}\}. \qquad (40)$$

Then, using (8) and noting that $\| \boldsymbol{v}_k \|^2 = (\tilde{\boldsymbol{G}}^H\tilde{\boldsymbol{G}})_{(k,k)}^{-1}$ we have

$$\mathbb{E}\{tr(\boldsymbol{A}^{(dl)})^{-1}\} = (e^{\frac{\bar{R}}{BW}} - 1)\mathbb{E}\{tr(\tilde{\boldsymbol{G}}^H\tilde{\boldsymbol{G}})^{-1}\}. \quad (41)$$

Moreover, using (29) in [28, eq.(50)],

$$\mathbb{E}\{tr(\boldsymbol{A}^{(dl)})^{-1}\} \approx \frac{(e^{\frac{\bar{R}}{BW}} - 1)}{(F - K)(1 + \sqrt{\frac{M-F}{FK}})}\mathbb{E}\{tr(\boldsymbol{B}^{-1})\}. \tag{42}$$

Since the expectation is the same for all UTs due to the assumed path-loss model [14], then $\mathbb{E}\{r_k^{-1}\} = \mathbb{E}\{r^{-1}\}$, so that

$$\mathbb{E}\{tr(\boldsymbol{B}^{-1})\} = K\mathbb{E}\{r^{-1}\} \tag{43}$$

and therefore

$$\mathbb{E}\{tr(\boldsymbol{A}^{(dl)})^{-1}\} \approx \frac{K(e^{\frac{\bar{R}}{BW}} - 1)}{(F - K)(1 + \sqrt{\frac{M-F}{FK}})}\mathbb{E}\{r^{-1}\}. \tag{44}$$

Finally, putting (44) into (9) results in (31), which concludes the proof. $\qquad\square$