Massive MIMO at Night: On the Operation of Massive MIMO in Low Traffic Scenarios

Hei Victor Cheng, Daniel Persson, Emil Björnson and Erik G Larsson

Tweet

LINKÖPING UNIVERSITY

Massive MIMO at Night: On the Operation of Massive MIMO in Low Traffic Scenarios

Hei Victor Cheng, Daniel Persson, Emil Björnson and Erik G. Larsson
Department of Electrical Engineering (ISY), Linköping University, Sweden
Email: {hei.cheng, daniel.persson, emil.bjornson and erik.g.larsson}@liu.se

Abstract—For both maximum ratio transmission (MRT) and zero forcing (ZF) precoding schemes and given any specific rate requirement the optimal transmit power, number of antennas to be used, number of users to be served and number of pilots spent on channel training are found with the objective to minimize the total consumed power at the base station. The optimization problem is solved by finding closed form expressions of the optimal transmit power and then search over the remaining discrete variables.

The analysis consists of two parts, the first part investigates the situation when only power consumed in the RF amplifiers is considered. The second part includes both the power consumed in the RF amplifiers and in other transceiver circuits. In the former case having all antennas active while reducing the transmit power is optimal. Adaptive scheme to switch off some of the antennas at the base stations is found to be optimal in the latter case.

I. Introduction

The idea of employing a large number of antennas at the base station (BS) to serve multiple terminals using the same time and frequency resources, commonly known as massive multiple-input-multiput-output (MIMO), has received extensive interests as it is a potential solution for the next generation cellular systems to meet the ever-increasing demand for data traffic [1], [2]. Massive MIMO systems was originally proposed in [3], therein the author illustrated the idea of using large number of antennas by analyzing asymptotically the achievable rates. He showed there that massive MIMO provides much higher data rate using the same amount of time and frequency resources. More surprisingly it was later found out that massive MIMO can give the benefit of increasing the spectral efficiency and radiated energy efficiency simultaneously by more than a hundred times compared to the performance of single antenna system [4]. However these benefits do not come for free, as having more antennas will increase the power consumed by circuits and baseband processing at the BS. Recent reports have shown that the total energy consumed by communication technology contributes more than 3 percent of the electrical energy in the world [5]–[7]. Therefore reducing the energy consumption in the BSs is very important, particularly as the data traffic increases rapidly. Unlike the approaches in [8]–[10] which suggest turning off some of the BSs to save energy, we keep the BSs turned on to

maintain the coverage in the cell. Assuming massive MIMO has been deployed in the cell, in some situations, e.g. at late night where the traffic demand is the lowest [11], having all the antennas turned on may cause a huge waste of power. Therefore one important question we are posing here is: what should the massive MIMO base stations do when the data traffic load is low?

In this work we try to answer this question by considering the operation of massive MIMO systems when the rate requirement is low compared to the peak rate that the system can offer. This corresponds to the case "at night" or other situations when there is less demand for data rate. In these cases the bulk of traffic is not likely to be delay sensitive (e.g. video streaming), therefore one can iterate between being fully turned off and sending data to everyone to meet the required data rate. In this case the most important performance metric is the system sum-rate. Each individual terminal can get what they desire and deserve by user scheduling. Therefore one can formulate an optimization problem with the objective to minimize the power consumed in the RF amplifiers (not to be confused with the radiated power), while satisfying a given performance requirement in terms of the sum-rate. An extended model is also investigated where the power consumed by other circuit components is taken into account. Here we consider only the power consumed by the base stations in the downlink, while the extension to include the uplink phase is left for future work. In the following all the rate expressions are in terms of bit per channel use (bpcu).

The general conclusion when we consider only the power consumed in RF amplifiers is that the optimal strategy is to keep all antennas turned on but reduce the transmit power. When the power consumed in other circuit components is considered as well, using a subset of the antennas and turning off the rest is optimal.

The paper is organized as follows: In Section II notations are introduced. In Section III we develop the models and necessary assumptions for this work. In Section IV we formulate the problem that we are going to solve. In Section V we give the solution approach to the problem. In Section VI the solution approach is extended to the case when taking power consumed in other circuit components into account. Finally in Section VII we provide some numerical results and in Section VIII we draw some conclusions.

II. NOTATIONS

We list the notations and definitions here for easy reference.

- M_0: number of antennas at the base station
    - M: number of antennas in use ($M \leq M_0$);
    - remaining $M_0 - M$ antennas are entirely turned off
- K_0: maximum number of users that can ever be simultaneously served
- K: number of simultaneously served users ($K \leq K_0$)
- N: number of sub-carriers in the system, all sub-carriers use the same antennas to transmit
- T: number of channel uses such that the channels are constant across both time and frequency intervals, also known as the number of symbols per coherence interval
- τ_p: the number of symbols per coherence interval spent on uplink training pilots ($K \leq \tau_p \leq T$)
- p_{d0}: maximum possible downlink transmit (radiated) power per sub-carrier in the entire array divided by the noise power at the terminal, to be shared among all antennas at the base station
- p_d: actual total downlink transmit power per sub-carrier for the entire array divided by the noise power at the terminal ($p_d \leq p_{d0}$)
- p_u: uplink per user transmit power divided by the noise power at the base station

III. BACKGROUND

We consider the downlink of a single cell massive MIMO system operating over a bandwidth of B Hz with N sub-carriers, all sub-carriers are assumed to use the same antennas to transmit. The BS is equipped with an array with M_0 antennas, M of them will be actually used while the remaining $M_0 - M$ antennas are assumed to be entirely turned off. Denote K_0 the maximum number of users that can be simultaneously served. The BS choose K users from that and serve them simultaneously with time-division duplex (TDD) operation. We assume a block flat fading channel where the channel remains constant over T symbols. During each slot of the T symbols, τ_p of them are spent on uplink training pilots where the BS estimate the channel and in the remaining $T - \tau_p$ symbols are used for data transmissions. The channels between the M antennas in used and the K users are assumed to follow independent identically distributed Rayleigh fading such that each element is an independent $CN(0,1)$ random variable. We denote p_{d0} as the maximum possible downlink transmit (radiated) power per sub-carrier in the entire array divided by the noise power at the terminal, to be shared among all antennas at the base station, p_d as the actual total downlink transmit power per sub-carrier for the entire array divided by the noise power at the terminal ($p_d \leq p_{d0}$) and p_u as the uplink per user transmit power divided by the noise power at the base station.

For unit-energy independent symbols and a specific pre-coding matrix $\boldsymbol{A}(j)$ for the jth sub-carrier, the radiated power on the ith antenna is $\sum_{j=1}^{N} ||\boldsymbol{a}_i(j)||^2$ where $\boldsymbol{a}_i(j)$ corresponds to the ith row of $\boldsymbol{A}(j)$. Since the elements of the channel

matrices are independent identically distributed (i.i.d.), the elements of the precoding matrix $\boldsymbol{A}(j)$ that based on the estimates of the channel matrices (e.g. MRT or ZF) are i.i.d. as well. Moreover N is large (in the order of hundreds or thousands), by the law of large numbers we can conclude that the norms of each row will be approximately equal. Therefore we can assume that the total power at the base station is equally distributed over all the antennas without inducing much loss, so each antenna is allocated with the same amount of power Np_d/M. Assume the available bandwidth is B Hz, we can write the actual transmit power at the base station P_d in relation to the normalized transmit power as

$$p_d = P_d \frac{\lambda^{-\kappa}}{BN_0}, \tag{1}$$

where λ is the distance of the users from the base station measured in meters, κ is the path loss exponent and N_0 stands for the noise spectral density. Similarly we can relate the maximum transmit power P_{d0} at the base station to the maximum normalized transmit power as

$$p_{d0} = P_{d0} \frac{\lambda^{-\kappa}}{BN_0}, \tag{2}$$

and the actual transmit power P_u by the terminal in relation to the normalized transmit power by the terminal as

$$p_u = P_u \frac{\lambda^{-\kappa}}{BN_0}. \tag{3}$$

An achievable sum rate of the downlink channel with MRT precoding is given by [12]

$$R_{\text{mrt}} = \left(1 - \frac{\tau_p}{T}\right) K \log_2 \left(1 + \frac{M}{K} \frac{p_d \tau_p p_u}{(1+p_d)(1+\tau_p p_u)}\right), \tag{4}$$

and an achievable sum rate of the uplink channel with MRC reception is given by [4]

$$R_{\text{mrc}} = \left(1 - \frac{\tau_p}{T}\right) K \log_2 \left(1 + \frac{M}{K} \frac{p_u}{(\frac{1}{K} + p_u)(1 + \frac{1}{p_u \tau_p})}\right). \tag{5}$$

These bounds are tight when M is large, which is the case in massive MIMO systems. Although the paper focus on the downlink, the uplink expression is given as well as it will be used to prove a theorem linking the optimal number of pilot symbols to the optimal number of users in terms of both downlink and uplink sum spectral efficiency.

Similarly, for zero-forcing precoding an achievable sum spectral rate of the downlink channel is given by

$$R_{\text{zf}} = \left(1 - \frac{\tau_p}{T}\right) K \log_2 \left(1 + \frac{M - K}{K} \frac{p_d \tau_p p_u}{1 + p_d + \tau_p p_u}\right). \tag{6}$$

To get the maximum sum spectral efficiency, all M_0 antennas will be in use and p_d will be at its maximum with $p_d = p_{d0}$, while K and τ_p are variables that need to be optimized.

IV. Power Minimization at Low System Load

When the rate demand is lower than the maximum sum rate, two things can be done to save power. The first option one can think of is to reduce the transmit power. The second option is completely turning off some of the available antennas. The question we are asking is: suppose in some situation the target some rate that we want to provide R_{target} is only a fraction of the maximum rate given above, how many antennas do we need to turn on and what is the corresponding transmit power that should be used to satisfy this target rate? This problem will not involve optimization over M if we only consider the radiated power. However if we take into account the consumed power on each antenna the problem will be different. We use the model from [13]–[15] where the consumed power can be written as

$$
P_{\text{cons}} = M \frac{1}{\eta_{\text{max}}} \left(\frac{P_d}{M}\right)^\epsilon \left(\frac{P_{d0}}{M_0}\right)^{1-\epsilon} = cM \left(\frac{p_d}{M}\right)^\epsilon, \quad (7)
$$

where ϵ is a parameter with $\epsilon \in [0, 0.5]$ varies for different types of power amplifiers, η_{max} is the maximum power efficiency obtained when the amplifiers are running at full power, i.e. when $\frac{P_d}{M} = \frac{P_{d0}}{M_0}$. Since P_d and p_d are essentially the same with just a constant difference, we will work with p_d in the following for convenience. The constant

$$
c = \frac{1}{\eta_{\text{max}}} \left(\frac{p_{d0}}{M_0}\right)^{1-\epsilon} \frac{BN_0}{\lambda^{-\kappa}} \quad (8)
$$

is obtained by grouping other constants that do not depend on p_d. To conclude we would like to solve the following optimization problem:

$$
\begin{aligned}
\underset{M,K,\tau_p,p_d}{\text{minimize}} \quad & cM \left(\frac{p_d}{M}\right)^\epsilon \\
\text{subject to} \quad & M \leq M_0 \\
& K \leq K_0 \\
& K \leq \tau_p \leq T \\
& R_{\text{mrt}} \text{ or } R_{\text{zf}} \geq R_{\text{target}}.
\end{aligned} \quad (9)
$$

V. Solution to the Power Minimization Problem

Problem (9) is certainly non-convex as it involves integer variables and the rate expression depends on the variables in a complicated non-convex manner. A brute force search can be performed by partitioning the space of $[0, p_{d0}]$ and searching for the optimal parameters. However, we observe that for any fixed M, K and τ_p, the optimal p_d can be found in closed form as detailed in the following proposition.

Proposition 1: For any fixed $M > 0, K > 0$ and $\tau_p \geq K$, the optimal p_d for R_{mrt} is given by

$$
p_d = \frac{x}{1-x} \text{ with } x = \frac{K}{M} \frac{1 + \tau_p p_u}{\tau_p p_u} \left(2^{\frac{R_{\text{target}}}{K(1-\tau_p/T)}} - 1\right). \quad (10)
$$

When $x \geq 1$ the desired R_{target} cannot be achieved for any p_d given this choice of $M > 0, K > 0$ and $\tau_p \geq K$. Similarly

the optimal p_d for R_{zf} is given by

$$
p_d = \frac{y(1 + \tau_p p_u)}{\tau_p p_u - y} \text{ with } y = \frac{K}{M-K} \left(2^{\frac{R_{\text{target}}}{K(1-\tau_p/T)}} - 1\right). \quad (11)
$$

When $y \geq \tau_p p_u$ the desired R_{target} cannot be achieved for any p_d given this choice of $M > 0, K > 0$ and $\tau_p \geq K$.

Proof: First we observe both R_{mrt} and R_{zf} are monotonically increasing functions of p_d when $M > 0, K > 0$ and $\tau_p \geq K$ are fixed, and the objective function in (9) is monotonically increasing in p_d as well. Therefore at the optimal point we must have that $R_{\text{mrt}} = R_{\text{target}}$ or $R_{\text{zf}} = R_{\text{target}}$, otherwise one can decrease p_d to achieve equality and decrease the objective value at the same time. Therefore equating $R_{\text{mrt}} = R_{\text{target}}$ or $R_{\text{zf}} = R_{\text{target}}$ gives the desired result. ∎

Having the optimal transmit power given by the closed form solution in Proposition 1, one can solve the optimization problem (9) to global optimality by search through the integer variables M, K and τ_p. However, this approach can be computationally heavy. To reduce the complexity we observe that the search over τ_p is not necessary for MRC/MRT operation. To prove this we need the following lemma.

Lemma 1: For any $x > 0$, we have $\log(x) \geq \frac{x-1}{x}$ with equality if $x = 1$.

Proof: Write $f(x) = \log(x) - \frac{x-1}{x}$, then we have $f'(x) = \frac{1}{x} - \frac{1}{x^2}$. Observing that $f'(x) \leq 0, \forall x \in (0, 1)$ and $f'(x) \geq 0, \forall x > 1$, we can conclude that $x = 1$ is the minimum point of $f(x)$ at which $f(x) = 0$. Thus we have $f(x) \geq 0, \forall x > 0$, which proves the lemma. ∎

Then we can state the following proposition.

Proposition 2: For MRC/MRT operation, we always have $\tau_p = K$ at the optimal solution of problem (9).

Proof: We first prove that for any fixed τ_p, both R_{mrc} and R_{mrt} are monotonically increasing functions of K. This is equivalent to proving that $f(K) = K \log\left(1 + \frac{a}{K}\right)$ and $g(K) = K \log\left(1 + \frac{a}{bK+1}\right)$ are monotonically increasing in K for any strictly positive real constants a, b. We look at their derivatives with respect to K:

$$
\begin{aligned}
f'(K) &= \log\left(1 + \frac{a}{K}\right) + K \cdot \frac{1}{1 + a/K} \cdot (-a/K^2) \\
&= \log\left(1 + \frac{a}{K}\right) - \frac{a}{K+a}.
\end{aligned} \quad (12)
$$

Using Lemma 1 with $x = 1 + a/K$ we have $\log\left(1 + \frac{a}{K}\right) \geq (\frac{K+a}{K} - 1)/\frac{K+a}{K} = \frac{a}{K+a}$. Therefore $f'(K) > 0$ for all $K > 0$. Similarly, we take the derivative of $g(K)$ with respect to K:

$$
\begin{aligned}
g'(K) &= \log\left(1 + \frac{a}{bK+1}\right) \\
&\quad + K \cdot \frac{1}{1 + a/(bK+1)} \cdot (-a/(bK+1)^2 \cdot b) \\
&= \log\left(1 + \frac{a}{bK+1}\right) - \frac{bKa}{(bK+1)(bK+1+a)} \\
&\geq \log\left(1 + \frac{a}{bK+1}\right) - \frac{a}{bK+1+a} \geq 0.
\end{aligned} \quad (13)
$$

The last inequality follows from putting $y = 1 + a/(bK + 1)$ and applying the lemma.

Now we have proved that both R_{mrc} and R_{mrt} are monotonically increasing functions of K for any fixed τ_p. On the other hand, K cannot exceed τ_p, in order for the rate expression to be valid. Thus we conclude that K and τ_p must be equal at the optimum point. ∎

This result has been reported numerically in [4, Fig. 7] for the uplink, and stated without rigorous proof in [12] for the downlink. However we are not aware of any previous formal analytical proof.

Given the above results, we can solve (9) by searching over different M and K (and also τ_p for ZF precoding). We consider $\epsilon = 0.5$ as in [15] with different values of p_u and R_{target}. Surprisingly we always find that the optimal solution gives $M = M_0$ which means all antennas are turned on. The reason is when we substitute the optimal p_d given by Proposition 1 into the objective of (9) we can see that for a fixed K the objective function is monotonically decreasing in M, which explains why all antennas should be turned on. However this is only true when ϵ is chosen to be 0.5, for $\epsilon < 0.5$ this does not always hold. Note that this conclusion stands in sharp contrast to the results for single user MIMO in [14] [15], where it was shown that the optimal strategy with low traffic demand typically is to turn off a subset of antennas. Simulation results will be given in the following section.

VI. EXTENSIONS TO INCLUDE OTHER CIRCUIT POWER CONSUMPTIONS

The optimization problem (9) only considers the power consumed in the RF amplifiers of the base station, however in practice power consumed by other RF circuits and by baseband processing is non-negligible [16]. In this section we extend the analysis to include also these parts of the total power consumption. In most cases, the power consumed by the transceiver circuits can be viewed as a linear function of M, since the same components are duplicated for each antenna branch. Therefore one can formulate the following optimization problem:

$$\begin{aligned}
\underset{M, K, \tau_p}{\text{minimize}} \quad & M\left(\frac{p_d}{M}\right)^{\epsilon} + bM + f(M, K, \tau_p) \\
\text{subject to} \quad & M \leq M_0 \\
& K \leq K_0 \\
& K \leq \tau_p \leq T \\
& R_{\mathrm{mrt}} \text{ or } R_{\mathrm{zf}} \geq R_{\mathrm{target}}.
\end{aligned} \quad (14)$$

where bM represents the power consumed by the other parts of the circuits that increase linearly with M divided by $c = \frac{1}{\eta_{\max}}\left(\frac{p_{d0}}{M_0}\right)^{1-\epsilon}\frac{BN_0}{\lambda^{-\kappa}}$ and $f(M, K, \tau_p)$ represents the power consumed for baseband processing divided by c.

Fig. 1 shows the number of antennas in use against the target rate with MRT pre-coding. The parameters are chosen given by for moderate SNR with $p_{d0} = 1$, $p_u = 0.1$, $T = 100$, $M_0 = 100$, $K_0 = 100$, $\eta = 0.55$, $\epsilon = 0.5$. Since $f(M, K, \tau_p)$ is usually small compared to bM, it is assumed to be 0 here for
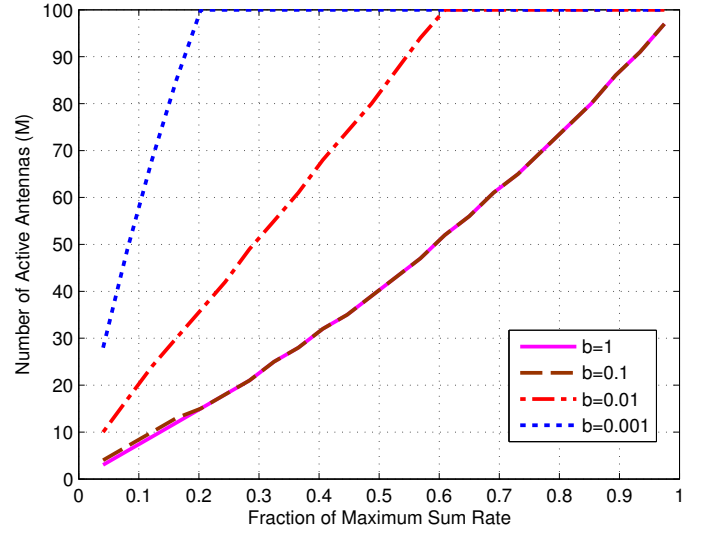

Fig. 1. Number of antennas in use against target rate with MRT precoding, when operating at moderate SNR with $p_{d0} = 1$, $p_u = 0.1$ for different values of b.

simplicity. Similar behaviours are observed in results with ZF precoding, and are not given here for brevity. The reason for the observed behaviour is the term bM added to the objective, this can be viewed as a regularization term a bit reminiscent of the sparse signal recovery literature (e.g. LASSO) as it acts as a sparsity inducing term.

We see that different values of b affect the optimal solution a lot so now the question is what would be the value in practice? To answer this we use the model developed in [17]. Therein a new circuit power consumption model was given, as in the following equation:

$$P_{\mathrm{CP}} = P_{\mathrm{FIX}} + P_{\mathrm{TC}} + P_{\mathrm{CE}} + P_{\mathrm{C/D}} + P_{\mathrm{BH}} + P_{\mathrm{LP}} \quad (15)$$

where the parameters are defined as:

- P_{FIX}: fixed power consumed by site-cooling, control signaling and other load independent processing.
- P_{TC}: $P_{\mathrm{TC}} = MP_{\mathrm{BS}} + P_{\mathrm{LO}}$ accounts for the power consumed in other parts in the transmitter chain. P_{BS} is the power to run different circuit components except the amplifiers (such as converters, mixers and filters), P_{LO} is the power to run the local oscillator which we assume all antennas share the same oscillator at the base station.
- P_{CE}: $P_{\mathrm{CE}} = 2\tau_p M K^2 / L_{\mathrm{BS}}$ accounts for the power consumed in the channel estimation. L_{BS} is the computational efficiency (measured in flops/Watt).
- $P_{\mathrm{C/D}}$: $P_{\mathrm{C/D}} = (P_{\mathrm{COD}} + P_{\mathrm{DEC}})R_{\mathrm{MRC}}$ (or R_{ZF}) stands for the power consumed in coding and decoding channel codes.
- P_{BH}: $P_{\mathrm{BH}} = P_{\mathrm{BT}}R_{\mathrm{MRC}}$ (or R_{ZF}) stands for the power consumed in backhaul transmission.
- P_{LP}: $P_{\mathrm{LP}} = (1 - \tau_p/T)2MK/L_{\mathrm{BS}} + P_{\mathrm{LPC}}$ stands for the consumed power in the linear processing. The first term accounts for performing a matrix vector multiplication (multiply the precoding matrix to the signal vector).

The second term accounts for the power consumed in computing the pre-coding matrix. $P_{\text{LPC}} = 3MK/L_{\text{BS}}$ for MRC precoding for which we just normalize each column of the channel matrix. $P_{\text{LPC}} = K^3/3L_{\text{BS}} + (3MK^2 + MK)/L_{\text{BS}}$ for ZF precoding by performing Cholesky factorization and back-substitution.

The above model seems complicated and some of the power consumption depends on R_{MRT} or R_{ZF} which may seem hard to tackle. However from the proof of Proposition 1 we know that $R_{\text{MRT}} = R_{\text{target}}$ and $R_{\text{ZF}} = R_{\text{target}}$ at the optimal point. Therefore we can regard $P_{\text{C/D}}$ and P_{BH} as constants as they do not include any optimization variables. Finally we can compute a reasonable b by using the model and parameters given in [17].

VII. NUMERICAL RESULTS

In this section we look at the solutions of (14) in different scenarios by using practical parameters from the literature. In each of the numerical results number of active antennas M is plotted against the fraction of maximum sum rate to see how many antennas we actually need to meet different demand of data rates.

Figs. 2, 3 and 4 show the number of antennas in use against the fraction of maximum rate at low SNR, moderate SNR and high SNR respectively. In all cases $P_{d0} = 2\text{W}$, $P_u = 0.2\text{W}$, $\eta = 0.55$, $\epsilon = 0.5$, $B = 20\text{MHz}$, $BN_0 = 96\text{dBm}$, $\kappa = 3.8$, $M_0 = 100$, $K_0 = 100$, $T = 100$, $P_{\text{BS}} = 1W$ and $L_{\text{BS}} = 12.8 \times 10^6$ flops/W [15], [17].

At low SNR where users are 2.5 km away from the base station, we observe that the curves are approximately linear. For moderate and high SNR where users are 1.5 km and 1 km away from the base station respectively, the curve increases slower than linear and behaves like a convex function. One intuitive reason for this is that at low SNR, the use of excess antennas can provide an array gain to boost the sum spectral efficiency and reduce p_d. However at moderate and high SNR we do not need extra antennas to provide the array gain as the received power is already high enough, while using more antennas increases the power consumptions in other part of the circuits. Therefore even less antennas are needed for moderate and high SNR scenarios.

VIII. CONCLUSIONS

In this paper we investigated the downlink operations at the massive MIMO base stations when the demand for data rate is only a fraction of the maximum rate that the system can offer. For both MRT and ZF precoding schemes, and any demanded rate requirement the optimal transmit power, number of antennas to be used, number of users to be served and number of pilots spent on channel training were found with the objective to minimize the total consumed power at the base stations.

Closed form solutions were given for the transmit power at the base station. Results of the optimization suggested that using a fraction of the available antennas proportional to the required rate over maximum supported rate achieved a close
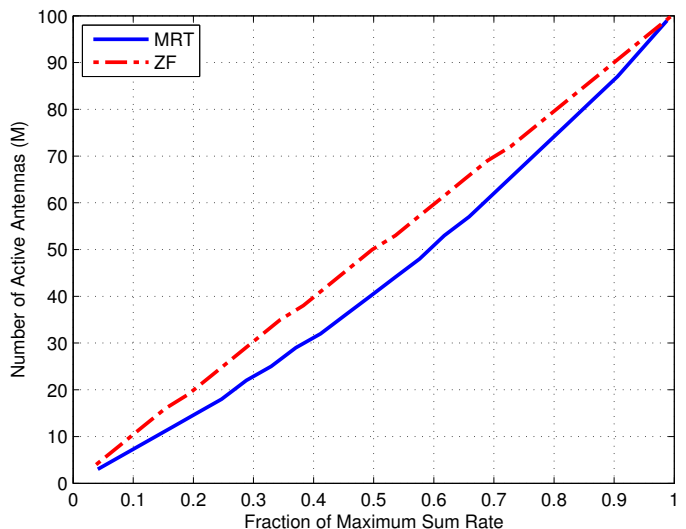

Fig. 2. Number of antennas in use against fraction of maximum rate, when operating at low SNR.
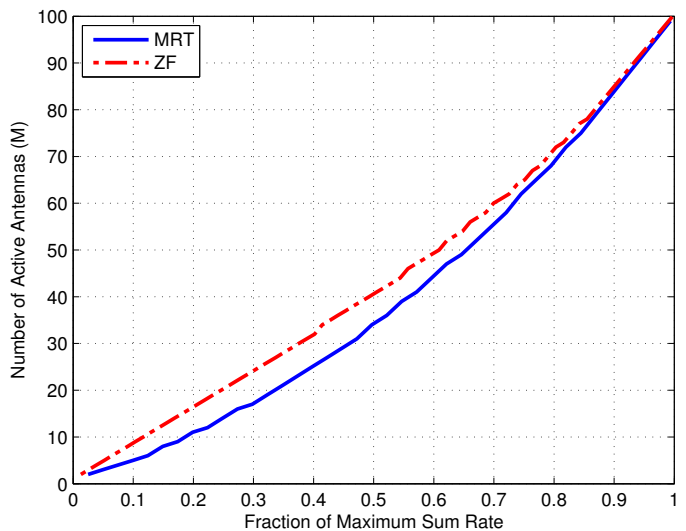

Fig. 3. Number of antennas in use against fraction of maximum rate, when operating at moderate SNR.

to optimal energy consumption. This shed some light on the operations at the massive MIMO base stations.

Future investigation includes taking into account the uplink scenarios and the consideration of multicell operation.

REFERENCES

[1] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Processing Mag.*, vol. 30, no. 1, pp. 40-46, Jan. 2013.

[2] E. G. Larsson, F. Tufvesson, O. Edfors, and T. L. Marzetta "Massive MIMO for next generation wireless systems," *IEEE Communication Mag.*, vol. 52, no. 2, pp. 186-195, Mar. 2014

[3] T. L. Marzetta "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Communications*, vol. 9, no. 11, pp. 3590-3600, Nov. 2010.
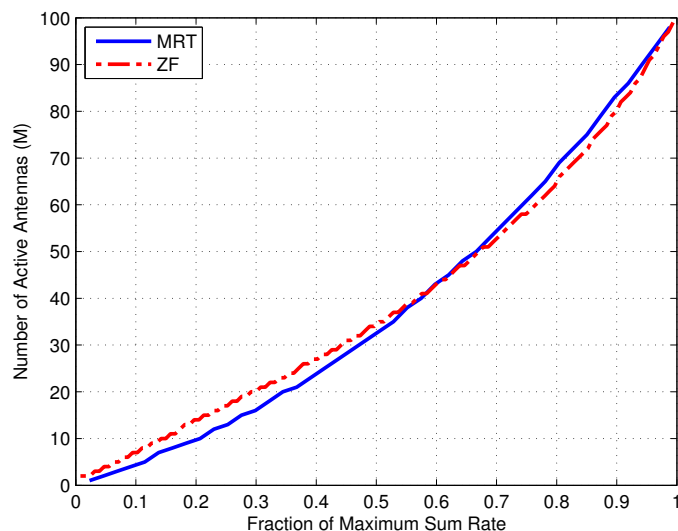
Fig. 4. Number of antennas in use against fraction of maximum rate, when operating at high SNR.

[17] E. Björnson, L. Sanguinetti, J. Hoydis, and M. Debbah, "Optimal design of energy-efficient multi-user MIMO systems: is massive MIMO the answer?," *IEEE Transactions on Wireless Communications, Submitted.* [online] Availble http://arxiv.org/pdf/1403.6150.pdf

[4] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Communications*, vol.61, no.4, pp.1436-1449, April 2013

[5] G. P. Fettweis and E. Zimmermann "ICT Energy Consumption - Trends and Challenges," *Proc. 11th Int. Symp. Wireless Personal Multimedia Commun. (WPMC08)*, Lapland, Finland, Sept. 2008

[6] Y. Chen, S. Zhang, S. Xu, and G. Y. Li, "Fundamental trade-offs on green wireless networks," *IEEE Communication Mag.*, vol. 49, no. 6, pp. 30-37, June 2011.

[7] Z. Hasan, H. Boostanimehr, and V. K. Bhargava, "Green cellular networks: A survey, some research issues and challenges," *IEEE Communication Surveys Tut.*, vol. 13, no. 4, pp. 524-540, Nov. 2011.

[8] M. A. Marsan, L. Chiaraviglio, D. Ciullo, and M. Meo "Optimal energy savings in cellular access networks," *Proc. IEEE ICC09 Workshop, GreenComm*, vol. 1, no. 5, pp. 14-18, 2009.

[9] W.-C. Liao, M. Hong, Y.-F. Liu, and Z.-Q. Luo, "Base station activation and linear transceiver design for optimal resource management in heterogeneous networks," 2014. [online] Avaliable: http://arxiv.org/abs/1309.4138

[10] S. Singh and J. Andrews, "Joint resource partitioning and offloading in heterogeneous cellular networks," *IEEE Trans. Wireless Communications*, vol. 13, no. 2, pp. 888-901, Feb. 2014.

[11] G. Auer, V. Giannini, I. Godor, P. Skillermark, M. Olsson, M. A. Imran, D. Sabella, M. J. Gonzalez, C. Desset, and O. Blume "Cellular Energy Efficiency Evaluation Framework," *Proc. IEEE 73rd Vehicular Technology Conference (VTC Spring)* , vol. 1, no. 6, pp. 15-18, May 2011

[12] H. Yang and T. L. Marzetta "Performance of conjugate and zero-forcing beamforming in large-scale antenna systems," *IEEE J. Sel. Areas Commun.*, vol.31, no.2, pp.172-179, Feb. 2013

[13] S. Mikami, T. Takeuchi, H. Kawaguchi, C. Ohta, and M. Yoshimoto "An efficiency degradation model of power amplifier and the impact against transmission power control for wireless sensor networks," *Proc. IEEE Radio and Wireless Symposium*, Jan. 2007, pp. 447-450.

[14] D. Persson, T. Eriksson, and E. G. Larsson, "Amplifier-aware multiple-input multiple-output power allocation," *IEEE Communications Letters*, vol. 17, no. 6, pp. 1112-1115, Jun. 2013.

[15] H. V. Cheng, D. Persson, and E. G. Larsson, "MIMO capacity under power amplifiers consumed power and per-antenna radiated power constraints," *IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2014.

[16] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M.A. Imran, D. Sabella, M.J. Gonzalez, O. Blume, and A. Fehske, "How much energy is needed to run a wireless network?," *IEEE Wireless Communication Mag.*, vol. 18, no. 5, pp. 40-49, 2012.