Performance of ZF Precoder in Downlink Massive MIMO with Non-Uniform User Distribution

Chuili Kong, Caijun Zhong, and Zhaoyang Zhang

Abstract: **In this paper, we investigate the achievable sum rate and energy efficiency of downlink massive multiple-input multiple-output antenna systems with zero-forcing precoding, by taking into account the randomness of user locations. Specifically, we propose two types of non-uniform user distributions, namely, center-intensive user distribution and edge-intensive user distribution. Based on these user distributions, we derive novel tight lower and upper bounds on the average sum rate. In addition, the impact of user distributions on the optimal number of users maximizing the sum rate is characterized. Moreover, by adopting a realistic power consumption model which accounts for the transmit power, circuit power and signal processing power, the energy efficiency of the system is studied. In particular, closed-form solutions for the key system parameters, such as the number of antennas and the optimal transmit signal-to-noise ratio maximizing the energy efficiency, are obtained. The findings of the paper suggest that user distribution has a significant impact on the system performance: for instance, the highest average sum rate is achieved with the center-intensive user distribution, while the lowest average sum rate is obtained with the edge-intensive user distribution. Also, more users can be served with the center-intensive user distribution.**

Index Terms: **Energy efficiency, massive MIMO systems, non-uniform user distribution, sum rate.**

I. INTRODUCTION

TO meet the explosive increase in the demands for high speed and high quality wireless services in next decade, novel spectral efficient wireless technologies are required. One of the promising candidates is the massive multiple-input multiple-output (MIMO) technology, which deploys a very large number of antennas at the base-station (BS) while serving tens of users. By doing so, it was shown in the pioneering work of Marzetta [1], that the detrimental effects in conventional cellular networks, such as small-scale fading, co-channel interference and white noise, vanish asymptotically even adopting the most simple linear processing technique. As such, unprecedented spectral and energy efficiencies can be obtained.

The huge potential of massive MIMO technology has captured the attention of both industry and academia, and triggered enormous research activities striving to understand the fundamental limits of massive MIMO systems in various propagation environments [2]–[13]. Assuming Rayleigh fading channels, [2] studied the spectral efficiency and energy efficiency of an uplink massive MIMO system. Later in [4], a comprehensive comparison between conjugate and zero-forcing (ZF) beamforming was presented. Later in [5], [14], the analysis was extended to the more general Rician fading channels. Taking into account the circuit power consumption, the energy efficiency of massive MIMO systems was examined in [6], where it was shown that the energy efficiency is strictly quasi-concave with respect to the transmit power.

In the above mentioned papers, the distances between the users and BS are assumed to be deterministic. As such, the randomness of the user location is not taken into consideration during the analysis. On the other hand, tackling the issue of random user distribution is also very important, since it is required in the evaluation of some key performance metrics such as the area spectral efficiency. Thus far, in the context of massive MIMO, there are mainly two approaches in the literature to model the random user distribution, i.e., the stochastic geometry model (see [15] and references therein), where the users are modeled by the poisson point process (PPP) and the probability density function (PDF) model [16], where the distance between the user and BS is modeled by a specific probability distribution. Compared to the PDF model, the PPP model embodies an additional randomness of user number, hence is more difficult to handle analytically.

Thus far, most of existing works addressing the random user distribution consider the uniform user distribution [15], [16]. However, in practice, the user location is determined by the physical topography of the land, i.e., hills and rivers, and the distribution of man made infrastructure, such as malls, buildings and parks. Therefore, the users within the cellular network are more likely to be clustered, which, in general, does not lead to a uniform user distribution [17]. Few works have examined the impact of non-uniform user distribution on the system performance. In [18], the authors applied the thinning procedure to the PPP to obtain a non-uniform distribution. However, the ensuing analysis is purely based on numerical simulations due to analytical intractability. In contrast, [19], [20] adopted the PDF approach, and proposed different probability distributions to model the non-uniform user behavior. However, the proposed distributions lack sufficient flexibility, since they are not capable of controlling how strongly the users cluster towards the cell center or edge.

Manuscript received May 12, 2016; This paper is specially handled by EICs and Division Editors with the help of three anonymous reviewers in a fast manner.

The work was supported in part by the National Key Basic Research Program of China (No. 2012CB316104), the Zhejiang Provincial Natural Science Foundation of China (LR15F010001), the National High-Tech. R&D Program of China under grant 2014AA01A705 and 2014AA01A702, the Zhejiang Science and Technology Department Public Project (2014C31051), and the open research fund of National Mobile Communications Research Laboratory, Southeast University (No.2013D06).

C. Kong, C. Zhong, and Z. Zhang are with the Institute of Information and Communication Engineering, Zhejiang University, China, email: kcl_dut@163.com, {caijunzhong, ning_ming}@zju.edu.cn. C. Zhong is also with the National Mobile Communications Research Laboratory, Southeast University, Nanjing, 210018, China.

Digital object identifier 10.1109/JCN.2016.000099

Motivated by this, in the current paper, we propose two types of non-uniform user distributions, namely, center-intensive user distribution and edge-intensive user distribution, which are quite flexible, and can be applied in a wide range of different propagation scenarios by properly setting the system parameters. Considering the ZF precoding scheme, novel closed-form tight bounds for the ergodic sum rate are presented. Based on which, we characterize the optimal number of users maximizing the sum rate, and prove that, among the center-intensive, uniform, and edge-intensive user distributions, the optimal number of users in the center-intensive user distribution case is the largest, while that in the edge-intensive user distribution case is the smallest. In addition, we examine the energy efficiency of the system by taking into account the transmit power, circuit power and signal processing power consumption. We show that there exist unique operating points where the maximum energy efficiency can be achieved.

The remainder of this paper is organized as follows: Section II introduces the system model. The average sum rate of the system is analyzed in Section III. Section IV studies the energy efficiency of the system. Numerical results and discussions are provided in Section V. Finally, Section VI concludes the paper and summarizes the key findings.

II. SYSTEM MODEL

We consider a single cell downlink system as illustrated in Fig. 1, where the BS is equipped with M antennas, and serves K ($K \leq M$) single antenna users. Since the main focus of the work is to investigate the impact of non-uniform user distribution, we adopt the perfect CSI assumption as in many prior works [2], [21] to facilitate the theoretical analysis.

Taking into account the location of different users, the effective channel from the BS to the kth user is given by $\mathbf{g}_k = \sqrt{\beta_k}\mathbf{h}_k$, where $\beta_k = d_k^{-\alpha}\varsigma_k$, with d_k being the distance between the BS and the kth user, $\alpha \geq 2$ being the path loss exponent, and ς_k denoting the shadowing effect which follows the log-normal distribution, i.e., $10\log_{10}\varsigma_k \sim \mathcal{N}\left(0, \sigma_{\text{sh}}^2\right)$. Moreover, $\mathbf{H} = \left[\mathbf{h}_1^T, \cdots, \mathbf{h}_K^T\right]^T \in \mathcal{C}^{K\times M}$ is the fast fading channel matrix, where $(\cdot)^T$ denotes the transpose operation. As we only consider correlation amongst transmit antenna elements and assume a separable correlation model, the channel matrix can be written as $\mathbf{H} = \tilde{\mathbf{H}}\mathbf{R}_t^{1/2}$, where the elements of $\tilde{\mathbf{H}}$ are independent and identically distributed complex Gaussian variables with zero mean and unit variance, which are constant for T channel uses, and $\mathbf{R}_t \in \mathcal{C}^{M\times M}$ denotes the transmit correlation matrix.

A. Non-Uniform User Distributions

In this subsection, we propose the two types of non-uniform distributions, namely, center-intensive user distribution and edge-intensive user distribution, which are not only analytically simple, but also can model practical scenarios. For instance, the center-intensive user distribution is particularly suitable for urban scenarios with populated buildings (e.g., shopping malls, stations), while the edge-intensive user distribution is suitable for rural mountainous propagation scenarios.

For center-intensive user distribution, the PDF of the distance d_k is given by

$$f_{d_k}(r) = \frac{a_c(R-r)^2 + 2r}{R^2 - {r_0}^2} b_c, \quad r_0 \leq r \leq R, \tag{1}$$

where R is the cell radius, r_0 denotes the guard interval which specifies the nearest distance between the users and BS. The scaling factor a_c controls how strongly the users cluster towards the BS, and b_c is given by $b_c = \frac{3(R+r_0)}{a_c(R-r_0)^2+3(R+r_0)}$.

For the edge-intensive user distribution, the PDF of the distance d_k is given by

$$f_{d_k}(r) = \frac{a_e r^2 + 2r}{R^2 - {r_0}^2} b_e, \quad r_0 \leq r \leq R, \tag{2}$$

where a_e controls how strongly the users cluster towards the cell edge and b_e is $b_e = \frac{3(R+r_0)}{a_e\left(R^2+Rr_0+r_0^2\right)+3(R+r_0)}$.

For the special case $a_c = a_e = 0$, both distributions reduce to the conventional uniform distribution. Fig. 2 gives an illustrative example of the proposed non-uniform distributions.

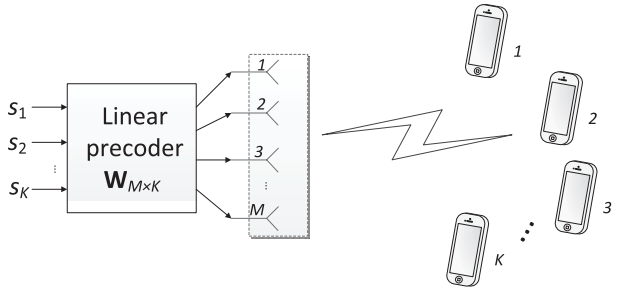


Fig. 1. Illustration of the multi-user MIMO system under consideration.

B. Achievable Sum Rate

The received signal at user k can be expressed as

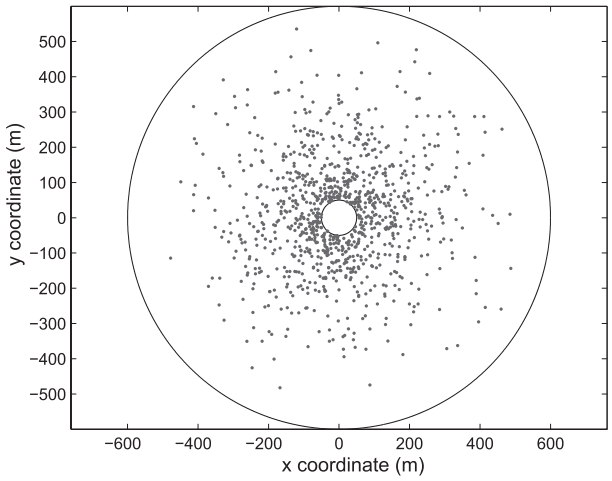
$$y_k = \sqrt{\beta_k}\mathbf{h}_k\mathbf{x} + n_k, \tag{3}$$

where n_k is the additive white Gaussian noise satisfying $n_k \sim \mathcal{CN}\left(0, \sigma^2\right)$, and $\mathbf{x} = \sum_{k=1}^{K}\sqrt{p_k}\mathbf{w}_k s_k$ is the BS transmit vector, where p_k is the transmit power, $\mathbf{w}_k$ is the beamforming vector, and s_k is the modulation symbol taken from Gaussian distribution.

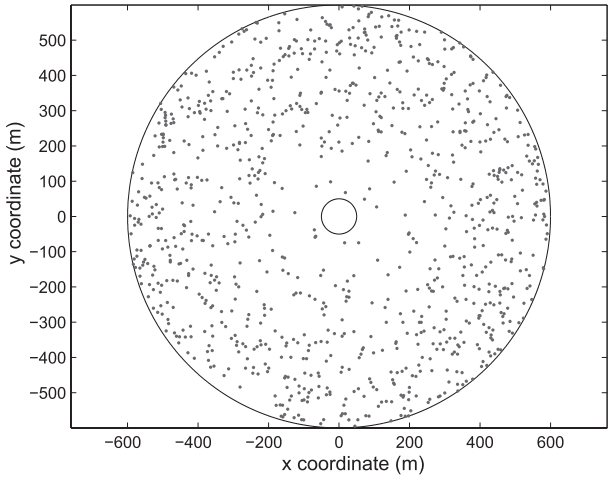
In this paper, we focus on ZF precoding. Hence, we have $\mathbf{w}_k = \mathbf{v}_k/||\mathbf{v}_k||$, where $\mathbf{v}_k$ is the kth column vector of the matrix $\mathbf{V} = \mathbf{H}^\dagger\left(\mathbf{H}\mathbf{H}^\dagger\right)^{-1}$, with $(\cdot)^\dagger$ denoting the conjugate transpose operator and $(\cdot)^{-1}$ denoting the matrix inverse. Moreover, $||\cdot||$ represents the Euclidean 2-norm. As in [16], we adopt the equal power allocation scheme, i.e., $p_k = p = P_{\text{T}}/K$, where P_{T} is the total transmit power. To this end, the instantaneous achievable rate of user k can be computed as

$$R_k^{\text{dl}} = \log_2\left(1 + \frac{P_{\text{T}}\beta_k\gamma_k}{K\sigma^2}\right), \tag{4}$$

where the superscript dl in R_k^{dl} is used to denote the downlink achievable rate, and $\gamma_k = 1/\left[\left(\mathbf{H}\mathbf{H}^\dagger\right)^{-1}\right]_{kk}$. Moreover, $[\mathbf{X}]_{ij}$ gives the (i,j)th entry of $\mathbf{X}$.



(a)



(b)

Fig. 2. Users' locations within a cell. In this example, there are 1000 users (i.e., blue points), $a_c = a_e = 1$, $R = 600$ m, and $r_0 = 50$ m: (a) Center-intensive user distribution and (b) edge-intensive user distribution.

As such, the average sum rate (measurable in bits/channel use) of the system is given by

$$R^{\text{dl}} = \mathtt{E}\left\{\sum_{k=1}^{K} R_k^{\text{dl}}\right\} = \mathtt{E}\left\{\sum_{k=1}^{K} \log_2\left(1 + \frac{P_{\text{T}}\beta_k\gamma_k}{K\sigma^2}\right)\right\}, \quad (5)$$

where the notation $\mathtt{E}\{\cdot\}$ denotes the statistical expectation operator, and the expectation is taken over large-scale fading β_k and small-scale fading $\mathbf{H}$.

III. AVERAGE SUM RATE ANALYSIS

In this section, we analyze the average sum rate of the system, and investigate the impact of key system parameters. Since characterizing the exact average sum rate is intractable, in the following, we derive tight bounds on the average sum rate.

A. Correlated MIMO Channels

Proposition 1: For $\mathbf{R}_t \neq \mathbf{I}_M$, i.e., correlated MIMO channels, the average sum rate of the system is lower bounded by

$$\begin{aligned} R^{\text{dl}} &\geq R^{\text{dl}}_{\text{corr,l}} \quad (6)\\ &= K\log_2\left(1 + \frac{\rho}{K}\exp\left(\psi(K) + \frac{\det(\mathbf{X}_{M-K+1})}{\prod_{i<j}^{M}(\lambda_j - \lambda_i)}\right)\right), \end{aligned}$$

where $\psi(x)$ is Euler's digamma function [22, Eq. (8.360.1)], $\rho = P_{\text{T}}\beta/\sigma^2$ with β being (7) (at the top of the next page) for the center-intensive user distribution, or (8) (at the top of the next page) for the edge-intensive user distribution; moreover, λ_i, $i = 1, 2, \cdots, M$ are the real, positive eigenvalues of $\mathbf{R}_t$, and $\mathbf{X}_k$ is a $M \times M$ matrix with entries

$$\{\mathbf{X}_{M-K+1}\}_{i,j} = \begin{cases} \lambda_i^{j-1}, & j \neq M-K+1; \\ \lambda_i^{j-1}\ln\lambda_i, & j = M-K+1. \end{cases} \quad (9)$$

Proof: See Appendix A. □

Due to the fact that the function $\log_2(1+ae^x)$ is almost linear under very large a, the lower bound given in (11) converges to the exact result in the high SNR regime. However, because of the complex expression in Proposition 1, it is difficult to see the impact of correlation on the average sum rate of the system. As such, we now look into the asymptotic large antenna regime, and we have the following key result.

Proposition 2: In the asymptotic large antenna regime, i.e., $M \to \infty$, the average sum rate of correlated MIMO channels converges to that of uncorrelated MIMO channels.

Proof: From (5), we see that, the difference between the correlated and uncorrelated channels lies in $\gamma_k = \frac{1}{\left[(\mathbf{H}\mathbf{H}^\dagger)^{-1}\right]_{kk}}$. For uncorrelated channels, using the law of large numbers [2], $\lim_{M\to\infty} \frac{\tilde{\mathbf{H}}\tilde{\mathbf{H}}^\dagger}{M} \xrightarrow{a.s.} \mathbf{I}_K$, we have $\lim_{M\to\infty} \gamma_k \xrightarrow{a.s.} M$.

For correlated channels, capitalizing on the result presented in [3, Lemma 4] and using the fact that $\text{tr}\{\mathbf{R}_t\} = M$, we have $\lim_{M\to\infty} \frac{1}{M}\tilde{\mathbf{h}}_i\mathbf{R}_t\tilde{\mathbf{h}}_i^\dagger \xrightarrow{a.s.} \frac{1}{M}\text{tr}\{\mathbf{R}_t\} = 1$, and $\lim_{M\to\infty} \frac{1}{M}\tilde{\mathbf{h}}_i\mathbf{R}_t\tilde{\mathbf{h}}_j^\dagger \xrightarrow{a.s.} 0$, where $\tilde{\mathbf{h}}_i$ denotes the ith row vector of $\tilde{\mathbf{H}}$, and $i \neq j$. Then, we can easily show that $\lim_{M\to\infty} \frac{\tilde{\mathbf{H}}\mathbf{R}_t\tilde{\mathbf{H}}^\dagger}{M} \xrightarrow{a.s.} \mathbf{I}_K$. Therefore, $\lim_{M\to\infty} \gamma_k \xrightarrow{a.s.} M$, which completes the proof. □

In general, increasing the number of BS antennas would result in higher correlation, which reduces the average sum rate. On the other hand, the average sum rate improves with more antennas due to the higher array gain. In this regard, Proposition 2 reveals an interesting result that the sum rate loss due to correlation could be effectively compensated by increasing the number of antennas, and with sufficiently large number of antennas, the effect of spatial correlation vanishes completely.[1]

B. Uncorrelated MIMO Channels

Proposition 1 is quite general and applicable to any number of users and antennas, as well as, arbitrary transmit correlated MIMO channels. However, due to the complex structure

[1]Note that if the correlation channel matrix $\mathbf{R}_t$ is rank deficient, then this observation is no longer valid.

$$\beta = \exp\left(\frac{-\alpha b_c}{R^2 - r_0^2}\left[\frac{a_c}{3}\left(R^3\left(\ln R - \frac{11}{6}\right)\right) + R^2\left(\ln R - \frac{1}{2}\right)\right]\right) \times \exp\left(\frac{-\alpha b_c}{R^2 - r_0^2}\left[\frac{a_c}{3}\left(-3R^2 r_0\left(\ln r_0 - 1\right) + 3Rr_0^2\left(\ln r_0 - \frac{1}{2}\right) - r_0^3\left(\ln r_0 - \frac{1}{3}\right)\right) - r_0^2\left(\ln r_0 - \frac{1}{2}\right)\right]\right) \tag{7}$$

$$\beta = \exp\left(\frac{-\alpha b_e}{R^2 - r_0^2}\left[\frac{a_e}{3}\left(R^3\left(\ln R - \frac{1}{3}\right) - r_0^3\left(\ln r_0 - \frac{1}{3}\right)\right) + R^2\left(\ln R - \frac{1}{2}\right) - r_0^2\left(\ln r_0 - \frac{1}{2}\right)\right]\right) \tag{8}$$

of $\frac{\det(\mathbf{X}_{M-K+1})}{\prod_{i<j}^{M}(\lambda_j - \lambda_i)}$, extracting physical insights appears to be difficult. As such, we now consider the special case of uncorrelated MIMO channels, namely, $\mathbf{R}_t = \mathbf{I}_M$, and we have the following simplified expression.

Proposition 3: For $\mathbf{R}_t = \mathbf{I}_M$, i.e., uncorrelated MIMO channels, the average sum rate of the system is lower bounded by

$$R^{\text{dl}} \geq R^{\text{dl}}_{\text{uncorr,l}} = K\log_2\left(1 + \frac{\rho}{K}\exp\left(\psi(M-K+1)\right)\right), \tag{10}$$

where ρ is given in Proposition 1.

Proof: The sum rate can be lower bounded by

$$R^{\text{dl}} \geq R^{\text{dl}}_{\text{uncorr,l}} = \sum_{k=1}^{K}\log_2\left(1 + \frac{P_{\text{T}}}{K\sigma^2}e^{\mathtt{E}\{\ln\beta_k\}+\mathtt{E}\{\ln\gamma_k\}}\right),$$

where $\gamma_k = 1/\left[\left(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^{\dagger}\right)^{-1}\right]_{kk}$ is a Gamma random variable with parameters $(M-K+1, 1)$. Then, using the integral identity [22, Eq. (4.352.1)], we obtain $\mathtt{E}\{\ln(\gamma_k)\} = \psi(M-K+1)$. To this end, invoking the result of $\mathtt{E}\{\ln\beta_k\}$ given in Appendix A yields the desired result. □

Since $\psi(x)$ is an increasing function with respect to x, it can be easily inferred from Proposition 3 that the average sum rate increases when the number of BS antennas M becomes larger. In addition, we validate the intuitive result that the average sum rate is a decreasing function with respect to the path loss exponent and cell radius R.

Proposition 4: For $\mathbf{R}_t = \mathbf{I}_M$, i.e., uncorrelated MIMO channels, the average sum rate of the system is upper bounded by

$$R^{\text{dl}} \leq R^{\text{dl}}_{\text{uncorr,u}} \tag{11}$$
$$= K\log_2 e\left[\beta_1 + \psi(M-K+1) + \ln\left(\frac{P_{\text{T}}}{K\sigma^2} + \frac{\beta_2}{M-K}\right)\right],$$

where $\beta_1 = \ln\beta$ with β given by Proposition 1, and $\beta_2 = \hat{\beta}\exp\left(\frac{1}{2}\left(\frac{\ln 10}{10}\sigma_{\text{sh}}\right)^2\right)$ with $\hat{\beta}$ being (12) (at the top of the next page) for the center-intensive user distribution, or

$$\hat{\beta} = \frac{b_e}{R^2 - r_0^2}\left[\frac{a_e\left(R^{\alpha+3} - r_0^{\alpha+3}\right)}{\alpha+3} + \frac{2\left(R^{\alpha+2} - r_0^{\alpha+2}\right)}{\alpha+2}\right]$$

for the edge-intensive user distribution.

Proof: See Appendix B. □

The proposed lower and upper bounds are in general quite tight as will be demonstrated later through numerical simulations. In addition, the two bounds converge to each other at high SNRs as shown in the following corollary.

Corollary 1: For asymptotically high SNRs, i.e., P_{T}/σ^2 is large, both the lower and upper bounds, i.e., $R^{\text{dl}}_{\text{uncorr,l}}$ and $R^{\text{dl}}_{\text{uncorr,u}}$ become exact and equal to

$$R^{\infty} = \frac{K}{\ln 2}\left(\ln\left(\frac{P_{\text{T}}}{K\sigma^2}\right) + \beta_1 + \psi(M-K+1)\right). \tag{13}$$

Proof: The result can be obtained with some simple algebraic manipulations based on (10) and (11). □

When uniform distribution is assumed, a different lower bound has been proposed in [16, Eq. (17)]. Since our distribution encompass uniform user distribution, it is of great interest to compare the tightness of the lower bound $R^{\text{dl}}_{\text{uncorr,l}}$ with the lower bound $\hat{R}^{\text{dl}}_{\text{l}}$ proposed in [16].

Corollary 2: For uniform user distribution, the proposed lower bound is strictly tighter than the one proposed in [16], i.e.,

$$R^{\text{dl}}_{\text{uncorr,l}} > \hat{R}^{\text{dl}}_{\text{l}}. \tag{14}$$

Proof: See Appendix C. □

Now, we are interested in finding out the impact of the number of users on the average sum rate. Due to the presence of the digamma function, analytical characterization of the optimal K maximizing the average sum rate appears to be difficult. To this end, recall that in a massive MIMO system, the number of BS antennas M tends to be much larger than the number of users K; hence, $M-K+1$ is in general a large number. As such, exploiting the asymptotic property of digamma function [23, Eq. (6.3.18)], $\psi(x) \approx \ln x$ as $x \to \infty$, the following approximation can be obtained:

$$R^{\text{dl}}_{\text{uncorr,l}} \approx \tilde{R}^{\text{dl}}_{\text{uncorr,l}} = K\log_2\left(1 + \frac{\rho}{K}(M-K+1)\right). \tag{15}$$

Corollary 3: When both M and K become large with a fixed ratio c, i.e., $M = cK$ with $c \geq 1$, the average achievable rate of each user converges to

$$\frac{\tilde{R}^{\text{dl}}_{\text{uncorr,l}}}{K} \to \log_2\big(1 + \rho(c-1)\big). \tag{16}$$

$$\hat{\beta}=\frac{b_c}{R^2-r_0^2}\left[a_c\left(\frac{2R^{\alpha+3}}{(\alpha+1)(\alpha+2)(\alpha+3)}-\frac{R^2r_0^{\alpha+1}}{\alpha+1}-\frac{r_0^{\alpha+3}}{\alpha+3}+\frac{2Rr_0^{\alpha+2}}{\alpha+1}\right)+\frac{2\left(R^{\alpha+2}-r_0^{\alpha+2}\right)}{\alpha+2}\right] \tag{12}$$

It is interesting to note from Corollary 3 that in the asymptotic regime, the average achievable rate of each user depends only on the scaling factor c and the SNR ρ.

Proposition 5: The optimal number of user K maximizing the approximate average sum rate in (15) is $\lfloor K^{\text{opt}} \rfloor$ or $\lceil K^{\text{opt}} \rceil$, where K^{opt} is given by

$$K^{\text{opt}}=\frac{\rho(M+1)W\left(\frac{\rho-1}{e}\right)}{(\rho-1)\left(W\left(\frac{\rho-1}{e}\right)+1\right)}, \tag{17}$$

with $W(x)$ being the Lambert W function [24]. Also, $\lfloor x \rfloor$ denotes the nearest integer that is smaller than x while $\lceil x \rceil$ denotes the nearest integer that is larger than x.

Proof: It is easy to verify that $\tilde{R}^{\text{dl}}_{\text{uncorr,1}}$ is a strictly concave function with respect to K. Hence, there exists a unique K maximizing the average sum rate, which can be obtained by solving $\partial \tilde{R}^{\text{dl}}_{\text{uncorr,1}}/\partial K=0$. To this end, the desired result can be obtained after some algebraic manipulations. □

Proposition 5 provides an important design guideline on how many users should be scheduled simultaneously to maximize the average sum rate with specific cell size and number of BS antennas. In addition, we can observe that, the optimal number of user K depends only on the number of antennas M and the SNR ρ. Moreover, it can be easily seen that when the BS employs more antennas, the cell can serve more users to maximize the average sum rate.

Finally, we look into the impact of user distribution on the optimal number of users to serve, and we have the following key result.

Corollary 4: For different user distribution models, the optimal number of users K^{opt} satisfies $(K^{\text{opt}})^{\text{center-intensive}} \geq (K^{\text{opt}})^{\text{uniform}} \geq (K^{\text{opt}})^{\text{edge-intensive}}$, where the equality is met when $a_e=a_c=0$.

Proof: See Appendix D. □

In general, when the number of user K increases, the individual user rate decreases as shown in Proposition 3 and 4. Corollary 4 suggests that, in terms of maximizing the sum rate, the optimal number of users in the center-intensive user distribution is the largest compared to the other two distributions, indicating that more users can be accommodated in the center-intensive user distribution case. This is an expected result, since the average path loss tends to be significantly reduced for the center-intensive user distribution, such that the sum rate gain of having an additional user is larger than that of the other two distributions. As such, more users can be accommodated before the loss due to inter-user interference dominates the extra rate gain of having more users.

C. Impact of Random Number of Users

In practice, the number of users actively contending for channel access fluctuates across time. Hence, it is also of great interest to study the impact of random number of users on the sum rate performance. When the cell contains a large number of users and each user is active with a small probability, the distribution of the number of users q can be best approximated by the Poisson distribution [25]. As such, we assume that the probability mass function of q is given by $\mathcal{P}(q=k)=e^{-\mu}\mu^k/k!$, where $\mu=\mathtt{E}\{q\}=\lambda\pi\left(R^2-r_0^2\right)$ is the average number of users.

Therefore, the average sum rate taking into account of the randomness induced by the number of users can be expressed as

$$R^{\text{poisson}}=\mathtt{E}_q\left\{\tilde{R}^{\text{dl}}_{\text{uncorr,1}}\right\}=\mathtt{E}_q\left\{q\log_2\left(1+\frac{\rho}{q}(M-q+1)\right)\right\}.$$

To make a fair comparison, it is assume that the average number of active users is the same, i.e., $\mu=\mathtt{E}\{q\}=K$. In general, exact evaluation of the expectation is mathematically intractable. However, in the asymptotic large system regime, simple analytical expressions can be obtained.

Corollary 5: When both M and K grow large while keeping a finite and fixed ratio c, i.e., $M=cK$ with $c\geq 1$, we have

$$R^{\text{poisson}} \to \tilde{R}^{\text{dl}}_{\text{uncorr,1}} \to K\log_2(1+\rho(c-1)). \tag{18}$$

Proof: Substituting $M=cK$ into (15) yields

$$\tilde{R}^{\text{dl}}_{\text{uncorr,1}} \to \bar{R}^{\text{dl}}_{\text{uncorr,1}}=K\log_2(1+\rho(c-1)). \tag{19}$$

Then, we have

$$R^{\text{poisson}} \to \mathtt{E}\left\{\bar{R}^{\text{dl}}_{\text{uncorr,1}}\right\} \tag{20}$$

$$=\sum_{k=0}^{\infty}\frac{e^{-K}K^k}{k!}k\log_2(1+\rho(c-1)) \tag{21}$$

$$\overset{(a)}{=} K\log_2(1+\rho(c-1)), \tag{22}$$

where in (a), we have used the Taylor series expansion for e^{μ}. □

Corollary 5 indicates that, in the asymptotic large system regime, the randomness of user number does not affect the achievable sum rate. However, as shown in the following Proposition, this observation does not hold in the finite regime.

Proposition 6: The achievable average sum rate of the scenario with Poisson distributed number of users is inferior to that of the scenario with fixed number of users, i.e.,

$$R^{\text{poisson}} \leq \tilde{R}^{\text{dl}}_{\text{uncorr,1}}. \tag{23}$$

Proof: Noticing that $\tilde{R}^{\text{dl}}_{\text{uncorr,1}}$ is a concave function with respect to K, applying the Jensen's inequality yields

$$\mathtt{E}\left\{q\log_2\left(1+\frac{\rho}{q}(M-q+1)\right)\right\} \tag{24}$$

$$\leq \mathtt{E}\{q\}\log_2\left(1+\frac{\rho}{\mathtt{E}\{q\}}(M-\mathtt{E}\{q\}+1)\right). \tag{25}$$

Substituting $\mathtt{E}\{q\}=K$ into (24) arrives at the desired result. □

Proposition 6 suggests that, the randomness due to uncertain number of users results in a loss of sum rate compared to the case with deterministic number of users.

IV. ENERGY EFFICIENCY ANALYSIS

The energy efficiency, defined as the ratio of the spectral efficiency (average sum rate in bits/channel use) to the total power expended (in Joule/channel use), has become an important performance metric in the design of wireless networks. In this section, we consider a realistic energy consumption model and investigate how the three key parameters (the number of users K, the number of antennas M and the total transmit SNR ρ) affect the energy efficiency (η_{E}). In addition, the trade-off between η_{E} and average sum rate is characterized.

A. Power Consumption Model

Unlike most prior works, which only consider the transmit power, we adopt a more realistic energy consumption model as in [26] by considering three parts: The power consumption of the power amplifier (PA) P_{PA}, static circuit power P_{C}, and signal processing power P_{SP}. Hence, the total consumption of the system in one channel use is $P_{\text{total}} = P_{\text{PA}} + P_{\text{C}} + P_{\text{SP}}$.

In the following, we give a more detailed discussion on each power component.

1) Power consumption of PA: The power consumption of PA is proportional to the transmit power of the BS, and can be computed as $P_{\text{PA}} = P_{\text{T}}/\eta$, where $0 < \eta \leq 1$ denotes the efficiency of the PA.
2) Circuit power: The circuit power consumption is computed by $P_{\text{C}} = MP_{\text{BS}} + (K+1)P_{\text{syn}} + KP_{\text{UE}} + P_0$, where P_{BS} denotes the power consumption of the collection of components attached to each BS antenna, including the digital to analog converter, the mixer and the filter, and P_{UE} is the power consumption of one user, which includes the low-noise amplifier, the intermediate frequency amplifier, the filter and the analog to digital converter. Finally, P_{syn} is the power consumed by the frequency synthesizer, and P_0 is the fixed power consumption which includes active cooling system.
3) Signal processing power: The total power consumption due to signal processing can be expressed as $P_{\text{SP}} = K\left(P_{\text{cod}} + P_{\text{dec}}\right) + \underbrace{\frac{2K^2M + 2KM}{LT} + \frac{2K^3}{3LT}}_{\text{Part 1}} + \underbrace{\frac{MK}{L}}_{\text{Part 2}}$, where P_{cod} denotes the power consumption due to channel coding and modulation at the BS, where P_{dec} accounts for the power required for decoding symbols at one user. Part 1 comes from the computation of the ZF precoding matrix using the LU-based matrix inversion, where L denotes the computational efficiency, i.e., L operations per Joule. Part 2 is due to the multiplication of the precoding matrix with the vector of information symbols.

B. Maximum Energy Efficiency

Given the above energy consumption model, the η_{E} (bits/Joule) can be approximately expressed as (26), shown on the top of the next page.

The main objective is to obtain the optimal values for the key parameters M, ρ, and K, maximizing the η_{E}. We start with ρ.

Proposition 7: For fixed K and M, the optimal operating SNR ρ maximizing η_{E} is given by

$$\rho^* = \frac{e^{W\left(\frac{\beta\eta}{\sigma^2}\frac{(M-K+1)}{K}(P_{\text{C}}+P_{\text{SP}})-\frac{1}{e}\right)+1} - 1}{\frac{M-K+1}{K}}. \tag{27}$$

Proof: The result can be obtained by using the same technique presented in [26]. Hence, is omitted due to space limitation. □

Since the Lambert W function $W(x)$ is an increasing function with regard to x, Proposition 7 indicates that the optimal operating SNR is an increasing function of the circuit power P_{C} and the signal processing power P_{SP}. Therefore, reducing the circuit power consumption and enhancing the signal processing efficiency are important directions to pursue in order to make future wireless systems more energy efficient.

Proposition 8: For fixed K and ρ, the optimal number of BS antennas is $\lfloor M^* \rfloor$ or $\lceil M^* \rceil$, where M^* is given by (28), shown on the top of the next page.

Proof: The desired result can be obtained by following similar lines as in the proof of Proposition 7. □

Let $P_x = \rho\sigma^2/(\beta\eta) + (K+1)P_{\text{syn}} + KP_{\text{UE}} + P_0 + K\left(P_{\text{cod}} + P_{\text{dec}}\right) + 2K^3/(3LT)$, which can be interpreted as the portion of total energy consumption that is independent of the antenna number M and $P_y = P_{\text{BS}} + (2K^2 + 2K)/(LT) + K/L$ which can be interpreted as the energy consumption per BS antenna. It is easy to observe that M^* is an increasing function of P_x and decreasing function of P_y. This is intuitive, since if the cost of deploying an additional antenna is too high, i.e., P_y is large, less number of antennas should be used.

C. Energy Efficiency and Average Sum Rate Trade-off

We now present the relationship between energy efficiency and average sum rate. Capitalizing on the simplified sum rate approximation given in (15), we have

$$\rho = \frac{\left(2^{\tilde{R}^{\text{dl}}_{\text{uncorr,l}}/K} - 1\right)K}{M - K + 1}. \tag{29}$$

Then, substituting (29) into (26), we get

$$\eta_{\text{E}} \approx \frac{\tilde{R}^{\text{dl}}_{\text{uncorr,l}}}{\frac{\sigma^2 K}{\beta\eta(M-K+1)}\left(2^{\tilde{R}^{\text{dl}}_{\text{uncorr,l}}/K} - 1\right) + P_{\text{C}} + P_{\text{SP}}}. \tag{30}$$

As expected in (30), increasing the average sum rate is associated with increasing the SNR ρ, which has an unclear impact on the energy efficiency. Therefore, we need to know more about this relationship.

Proposition 9: The optimal average sum rate maximizing the energy efficiency is given by

$$\begin{aligned}&(\tilde{R}^{\text{dl}}_{\text{uncorr,l}})^* \\ &= \frac{K}{\ln 2}\left(W\left(\frac{\beta\eta\left(M-K+1\right)\left(P_{\text{C}}+P_{\text{SP}}\right)}{\sigma^2 Ke} - \frac{1}{e}\right) + 1\right).\end{aligned} \tag{31}$$

$$\eta_{\mathrm{E}} \approx \frac{K\log_2\left(1+\frac{\rho}{K}\left(M-K+1\right)\right)}{\frac{\rho\sigma^2}{\beta\eta}+MP_{\mathrm{BS}}+(K+1)P_{\mathrm{syn}}+KP_{\mathrm{UE}}+P_0+K\left(P_{\mathrm{cod}}+P_{\mathrm{dec}}\right)+\frac{2K^2M+2KM}{LT}+\frac{2K^3}{3LT}+\frac{MK}{L}} \tag{26}$$

$$M^* = \frac{e^{W\left(\frac{\frac{\rho}{eK}\left(\frac{\rho\sigma^2}{\beta\eta}+(K+1)P_{\mathrm{syn}}+KP_{\mathrm{UE}}+P_0+K\left(P_{\mathrm{cod}}+P_{\mathrm{dec}}\right)+\frac{2K^3}{3LT}\right)}{\left(P_{\mathrm{BS}}+\frac{2K^2+2K}{LT}+\frac{K}{L}\right)}-\frac{1}{e}\left(1-\frac{(K-1)\rho}{K}\right)\right)+1}+\frac{(K-1)\rho}{K}-1}{\frac{\rho}{K}} \tag{28}$$

Table 1. Parameters used in the simulations.

Parameter	Value
Cell size: r_0, R, path loss α	50 m, 600 m, 3.7
Symbol time S	$\frac{1}{9\times10^6}$ (s/channel use)
Noise variance σ^2	-90 dB
Shadow fading σ_{sh}	8 dB
Coherence bandwidth B	180 kHz
Coherence time T	10 ms $\cdot$ B $\cdot$ channel use
Operations/Joule L	12.8×10^9
Efficiency of PA η	0.3
P_{BS}, P_{UE}, P_{syn}	1W$\cdot S$, 0.3W$\cdot S$, 2W$\cdot S$
P_0, P_{cod}, P_{dec}	2W$\cdot S$, 4W$\cdot S$, 0.5W$\cdot S$

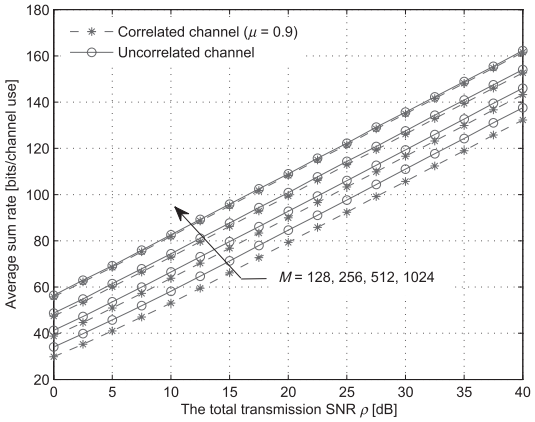


Fig. 3. Average sum rate of correlated and uncorrelated channels: K = 8.

Proof: The desired result can be obtained by following similar lines as in the proof of Proposition 5. □

From Proposition 9, there indeed exists a unique globally optimal $(\tilde{R}^{\mathrm{dl}}_{\mathrm{uncorr,l}})^*$ maximizing η_{E}. We observe that, the optimal average sum rate is an increasing function with respect to the number of antennas M. However, caution should be taken when interpreting the result. With larger M, the optimal average sum rate increases. Nevertheless, the energy efficiency achieved with the optimal average sum rate may decrease. This indicates that employing a very large number of antennas is not always beneficial from an energy efficiency perspective, which conforms with the corresponding results in [26].

V. NUMERICAL RESULTS

In this section, we provide numerical simulation results to validate the analytical expressions derived in the previous sections. Unless otherwise specified, the following set of parameters are used in simulations (see Table 1). These parameters follow the 3GPP propagation specification, and have been adopted in prior works such as [26]. As for the transmit channel correlation matrix $\mathbf{R}_t$, we use the exponential correlation model [27] with a single parameter μ measuring the correlation level.

Fig. 3 compares the simulated average sum rate of correlated and uncorrelated channels. As can be seen when the number of antenna M is in the order of a few hundreds, there exists a constant sum rate gap between the two correlated curves and the uncorrelated curves at all SNRs. Moreover, the sum rate gap shrinks with an increase in the number of antennas. For instance, with $M = 1024$, the sum rate of correlated channels with $\mu = 0.9$ is almost identical to that of the uncorrelated channels. This observation is consistent with Proposition 2.

Fig. 4 depicts the effect of transmit correlation on the sum rate of the system. As readily observed, the proposed lower bound in Proposition 1 is sufficiently tight across the entire SNR range, and becomes almost exact in the high SNR regime. This is due to the fact that the function $\log_2(1+ae^x)$ is almost linear under very large a, which results in the Jensen's inequality used in (32) being almost an equality. In addition, we notice that, the stronger the correlation effect (i.e., larger μ), the lower the average sum rate. However, as the number of transmit antenna becomes larger, the sum rate gap diminishes gradually, which indicates that deploying more transmit antennas is an effective approach to compensate for the sum rate loss due to correlation.

Fig. 5 examines the tightness of the proposed analytical bounds of (10) and (11), and compares the proposed lower bound against the lower bound proposed in [16] for different M. As can be readily observed, when the transmit power increases, the tightness of the proposed lower and upper bounds improves, and eventually both bounds converge to the exact results, which is agreement with Corollary 1. We also observe that the proposed upper bound becomes more accurate when the number of antennas is large. In addition, we see that the proposed lower bound is significantly tighter than the previously reported bound in [16] at all SNRs.

Fig. 6 first investigates the impact of user distribution on the average sum rate of the system when the number of users is fixed. For the non-uniform user distributions, we choose a_c =

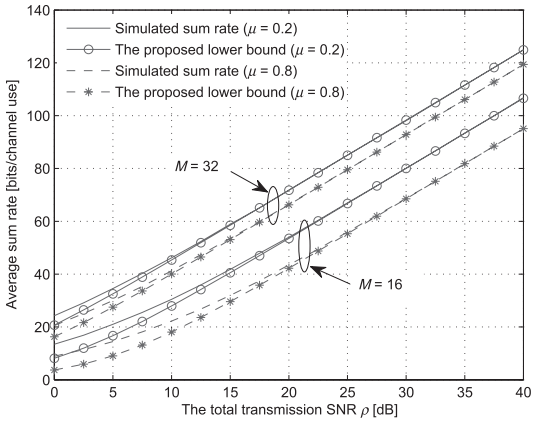


Fig. 4. Average sum rate with transmit antenna correlation: K = 8.

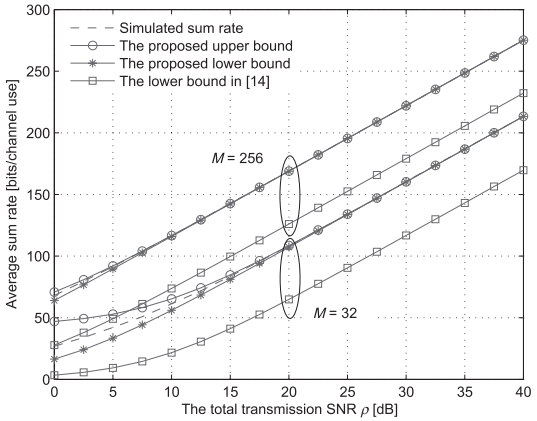


Fig. 5. Average sum rate of the system with uniform user distribution: K = 16.

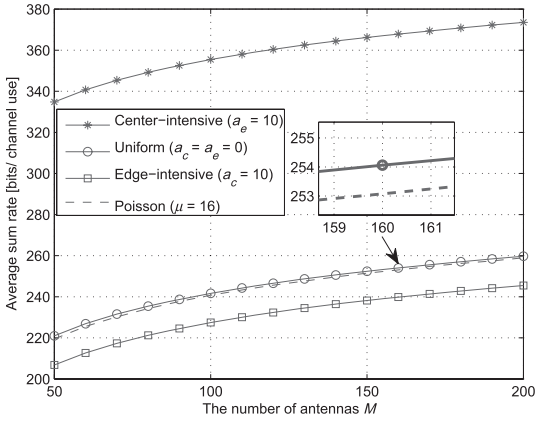


Fig. 6. Average sum rate against the number of antennas M for different user distributions: $P_{\mathrm{T}} = 43$ dBm and $K = 16$.

$a_e = 10$. We see that the center-intensive user distribution scenario achieves the highest sum rate, while the edge-intensive user distribution scenario attains the lowest sum rate. This is rather intuitive, since the communication distance between the user and the BS is in general shorter in the center-intensive distribution scenario. Given the same BS transmit power, shorter distances result in a higher sum rate. Also, intuitively, increasing the number of antennas improves the sum rate of the system in all three scenarios. Finally, we observe that the average sum rate of the scenario with Poisson distributed number of users is slightly lower than that of the scenario with deterministic number of users, conforming the analysis given in Proposition 6.

Fig. 7 shows the impact of the number of users K on the average sum rate for different user distributions. We observe that there exists an optimal K maximizing the average sum rate, which is in line with Proposition 5. Furthermore, the optimal number of users K^{opt} is, in general, different for different user distributions, i.e., the center-intensive user distribution can accommodate the highest number of users as analytically shown in Corollary 4. In addition, the gap in the supported number of users between the center-intensive distribution and the other two distributions appears to be significant, while the gap between uniform user distribution and edge-intensive user distribution is relatively small.

Fig. 8 examines the impact of user distributions on the energy efficiency. As we can clearly observe, the center-intensive user distribution scenario attains the highest energy efficiency. Moreover, there exists an optimal K^* maximizing the energy efficiency, as expected. Interestingly, it is shown that, when K is small, i.e., $K \leq 100$, increasing M actually reduces the energy efficiency. This could be explained by the fact that the sum rate increase due to a large M is not significant to offset the additional power dissipation of the additional antennas since the inter-user interference is not so strong for small K. We also observe that the maximum achievable energy efficiency associated with the case $M = 256$ is higher than the cases with $M = 128$ and $M = 512$, which again indicates that a large M is not always beneficial in terms of energy efficiency, as predicted by Proposition 8. Please note, similar observations have also been reported in prior works such as [26].

Fig. 9 examines the impact of the total transmit SNR on the energy efficiency of the system with different M and K for the uniform user distribution. As expected, we see that there exists an optimal operating SNR indicated by the star marker for each pair of M and K, and the optimal SNRs obtained from simulations are in perfect agreement with the analytical values predicted by Proposition 7. Also, when the number of users K becomes larger, i.e., from 16 to 64, the optimal SNR ρ also increases. This can be explained as follows: When the number of users K increases, the circuit power and signal processing power consumption become larger; hence, it is imperative to increase the transmit power to improve the energy efficiency.

Fig. 10 illustrates the tradeoff between the energy efficiency and average sum rate for the uniform user distribution scenario with different K and M. Interestingly, we see that for fixed M and K, there exists an optimal average sum rate achieving the maximum energy efficiency. Moreover, for a fixed K, the optimal average sum rate increases when the number of antennas changes from 32 to 256 as shown in Proposition 9. Similarly, for a fixed M, the optimal average sum rate increases along with the number of users. Finally, it can be observed that the highest energy efficiency is achieved with $K = 32$ and $M = 128$ among

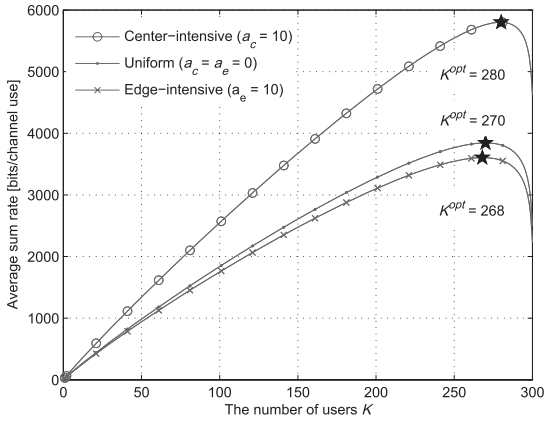


Fig. 7. Average sum rate against the number of users K for different user distributions: $P_{\mathrm{T}} = 43$ dBm and $M = 300$.

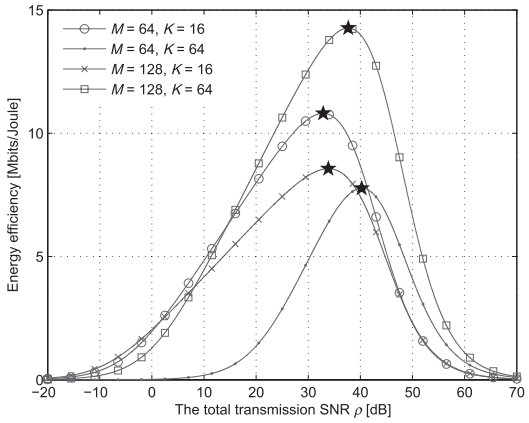


Fig. 9. Energy efficiency against the SNR ρ with different M and K for the uniform user distribution.

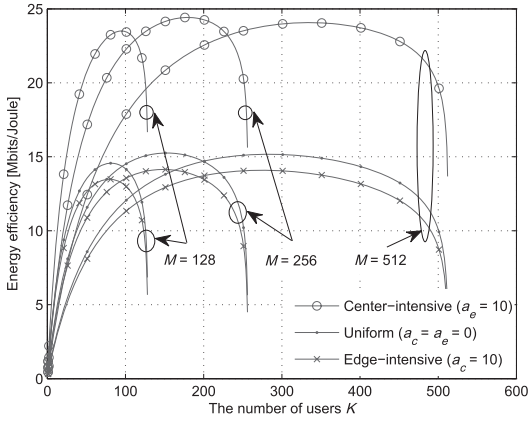


Fig. 8. Energy efficiency for different user distributions: $P_{\mathrm{T}} = 43$ dBm.

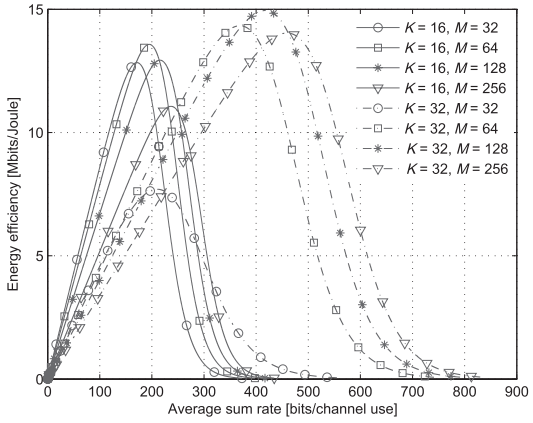


Fig. 10. Energy efficiency versus sum rate tradeoff.

eight different combinations, which agrees with the conclusion drawn in [26] that an energy efficient system should employ a large number of antennas to serve a relatively large number of users.

VI. CONCLUSION

The impact of user distribution on the performance of downlink massive MIMO systems was investigated. Specifically, based on the proposed center-intensive and edge-intensive user distributions, novel tight lower and upper bounds on the achievable sum rate were presented, and the optimal number of users K that maximizes the average sum rate was characterized. In addition, the energy efficiency of the system was analyzed. Specifically, considering a general power consumption model, which includes the transmit power, circuit power and signal processing power, exact closed-form expressions were obtained for the key system parameters such as M and ρ maximizing the energy efficiency. Finally, the trade-off between energy efficiency and average sum rate was established. The main outcome of the paper suggests that user distributions have significant impact on the system performance, as well as, on the choice of the optimal operating point.

APPENDICES

I. Proof of Proposition 1

Starting from (5), and using the fact that the function $f(x) = \ln(1 + \exp(x))$ is convex with respect to x, we obtain the following lower bound on the sum rate

$$R^{\mathrm{dl}} \geq R^{\mathrm{dl}}_{\mathrm{corr,l}} = \sum_{k=1}^{K} \log_2 \left(1 + \frac{P_{\mathrm{T}}}{K\sigma^2} e^{\mathrm{E}\{\ln \beta_k\} + \mathrm{E}\{\ln \gamma_k\}}\right). \quad (32)$$

Let $\mathbf{W} = \mathbf{H}\mathbf{H}^{\dagger}$. Then, we have $\gamma_k = \frac{1}{[\mathbf{W}^{-1}]_{kk}} = \frac{\det(\mathbf{W})}{\det(\mathbf{W}_{kk})}$, where $\mathbf{W}_{kk}$ is the minor of $\left[\mathbf{W}^{-1}\right]_{kk}$. Then, invoking the results of [28], we get $\mathrm{E}\{\ln(\det(\mathbf{W}))\} = \sum_{k=1}^{K} \psi(k) + \frac{\sum_{m=M-K+1}^{M} \det(\mathbf{X}_m)}{\prod_{i<j}^{M}(\lambda_j - \lambda_i)}$, and $\mathrm{E}\{\ln(\det(\mathbf{W}_{kk}))\} = \sum_{k=1}^{K-1} \psi(k) + \frac{\sum_{m=M-K+2}^{M} \det(\mathbf{X}_m)}{\prod_{i<j}^{M}(\lambda_j - \lambda_i)}$.

After some basic algebraic manipulations, we can explicitly compute $\mathtt{E}\{\ln\gamma_k\}$ as

$$\mathtt{E}\{\ln\gamma_k\} = \psi(K) + \frac{\det(\mathbf{X}_{M-K+1})}{\prod_{i<j}^{M}(\lambda_j-\lambda_i)}. \tag{33}$$

The next task is to evaluate $\mathtt{E}\{\ln\beta_k\}$, which is computed via

$$\mathtt{E}\{\ln\beta_k\} = -\alpha\mathtt{E}\{\ln d_k\} + \mathtt{E}\{\ln\varsigma_k\}. \tag{34}$$

Now, since$\mathtt{E}\{\ln d_k\} = \int_{r_0}^{R}\ln(r)\, f_{d_k}(r)dr$, using the identity $\mathtt{E}\{\ln\varsigma_k\} = 0$, the desired result can be obtained after some algebraic manipulations.

II. Proof of Proposition 4

Using the fact that

$$\mathtt{E}\{\log_2(1+ax)\} \le \log_2 e\left(\mathtt{E}\{\ln x\} + \ln\left(a + \mathtt{E}\left\{\frac{1}{x}\right\}\right)\right),$$

(5) can be upper bounded by

$$R^{\text{dl}} \le R^{\text{dl}}_{\text{uncorr,u}} = \sum_{k=1}^{K}\log_2 e \tag{35}$$
$$\left(\mathtt{E}\{\ln\beta_k\} + \mathtt{E}\{\ln\gamma_k\} + \ln\left(\frac{P_{\text{T}}}{K\sigma^2} + \mathtt{E}\left\{\frac{1}{\beta_k}\right\}\mathtt{E}\left\{\frac{1}{\gamma_k}\right\}\right)\right).$$

Since both $\mathtt{E}\{\ln\beta_k\}$ and $\mathtt{E}\{\ln(\gamma_k)\}$ have been evaluated in closed-form, the remaining task is to compute $\mathtt{E}\{1/\beta_k\}$ and $\mathtt{E}\{1/\gamma_k\}$. Exploiting the fact that user locations and shadow fading are mutually independent, we have $\mathtt{E}\{1/\beta_k\} = \mathtt{E}\{d_k^{\alpha}\}\mathtt{E}\{\zeta^{-1}\}$. It is easy to show that

$$\mathtt{E}\{\zeta^{-1}\} = \exp\left(\frac{1}{2}\left(\frac{\ln 10}{10}\sigma_{\text{sh}}\right)^2\right), \tag{36}$$

and

$$\mathtt{E}\{d_k^{\alpha}\} = \int_{r_0}^{R} r^{\alpha} f_{d_k}(r)dr. \tag{37}$$

Substituting (1) or (2) into (37), the desired result can be obtained after some algebraic manipulations.

III. Proof of Corollary 2

To compare $R^{\text{dl}}_{\text{uncorr,l}}$ with $\hat{R}^{\text{dl}}_{\text{l}}$, the following parameters are chosen, $r_0 = 0$ and $a_c = a_e = 0$. As such, the lower bound $R^{\text{dl}}_{\text{uncorr,l}}$ reduces to

$$R^{\text{dl}}_{\text{uncorr,l}} = K\log_2 \tag{38}$$
$$\left(1 + \frac{P_{\text{T}}}{K\sigma^2}\exp\left(-\alpha\left(\ln R - \frac{1}{2}\right)\right)\exp\left(\psi(M-K+1)\right)\right).$$

Using the fact that $\psi(M-K+1) > \ln(M-K)$, we have

$$R^{\text{dl}}_{\text{uncorr,l}} > K\log_2\left(1 + \frac{P_{\text{T}}}{\sigma^2 R^{\alpha}}\exp\left(\frac{\alpha}{2}\right)\left(\frac{M}{K}-1\right)\right). \tag{39}$$

Now consider function $f(x) = \exp(x) - x - 1$. It is easy to show that $f(x)$ is a monotonically increasing function. Since $f(0) = 0$, hence $f(x) > 0$ for $x > 0$, we have $\exp\left(\frac{\alpha}{2}\right) > \frac{\alpha+2}{2}$. Therefore,

$$R^{\text{dl}}_{\text{uncorr,l}} > K\log_2\left(1 + \frac{P_{\text{T}}(\alpha+2)}{2\sigma^2 R^{\alpha}}\left(\frac{M}{K}-1\right)\right). \tag{40}$$

Recall that

$$\hat{R}^{\text{dl}}_{\text{l}} = \tag{41}$$
$$K\log_2\left(1 + \frac{P_{\text{T}}}{\sigma^2}\frac{\alpha+2}{2R^{\alpha}}\exp\left(-\frac{1}{2}\left(\frac{\ln 10}{10}\sigma_{\text{sh}}\right)^2\right)\left(\frac{M}{K}-1\right)\right).$$

Since $\exp\left(-\frac{1}{2}\left(\frac{\ln 10}{10}\sigma_{sh}\right)^2\right) < 1$, we have $R^{\text{dl}}_{\text{uncorr,l}} > \hat{R}^{\text{dl}}_{\text{l}}$.

IV. Proof of Corollary 4

Ignoring the constant $(M+1)$ and let $t = (\rho-1)/e$, then (17) reduces to

$$f(t) = \frac{\left(t+\frac{1}{e}\right)W(t)}{t\left(W(t)+1\right)}. \tag{42}$$

Using the derivative property of the Lambert W function [29], $W'(t) = \frac{W(t)}{t(1+W(t))}$, the first derivative of $f(t)$ can be computed as $f'(t) = \frac{W(t)g(t)}{t^2(1+W(t))^3}$, where $g(t) = t - \frac{2}{e}W(t) - \frac{1}{e}W^2(t)$. It is easy to show that $g(t)$ is a convex function since $g''(t) = \frac{2}{e}\frac{W^2(t)}{t^2(1+W(t))} > 0$. Now, noticing that $g(-1/e) = g(0) = 0$, the t^* achieving the minimum of $g(t)$ is in the range of $(-1/e, 0)$, where $g(t) < 0$. Also, for $t \ge 0$, $g(t) \ge 0$. To this end, it is not difficult to show that $f'(t) > 0$ in the entire range of $(-1/e, \infty)$. Therefore, $f(t)$ is an increasing function with respect to t, which completes the proof of the first part.

The second part is to show the impact of user distribution, and consequently $\rho = P_{\text{T}}\beta/\sigma^2$, where β is a user distribution specific parameter, and is given by (7) for center-intensive user distribution and (8) for edge-intensive user distribution. Since r_0 is in general very small compared to the cell radius, which has little impact on the proof, hence, to simplify the proof, we assume $r_0 = 0$. As such, we have $\beta^{\text{center-intensive}} = -\alpha b_c[a_c/3\cdot R(R-11/6) + \ln R - 1/2]$, with $b_c = 3/(a_c R+3)$, $\beta^{\text{uniform}} = -\alpha[\ln R - 1/2]$, and $\beta^{\text{edge-intensive}} = -\alpha b_e[a_e/3\cdot R(R-1/3) + \ln R - 1/2]$, with $b_e = 3/(a_e R+3)$. To this end, it is easy to show $\beta^{\text{center-intensive}} - \beta^{\text{uniform}} = 4\alpha a_c/3(a_c R+3) \ge 0$, and $\beta^{\text{edge-intensive}} - \beta^{\text{uniform}} = -\alpha a_c/6(a_c R+3) \le 0$, which completes the proof.

REFERENCES

[1] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

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

[3] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

[4] 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.

[5] Q. Zhang, S. Jin, K.-K. Wong, H. Zhu, and M. Matthaiou, "Power scaling of uplink massive MIMO systems with arbitrary-rank channel means," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 966–981, Oct. 2014.

[6] G. Miao, "Energy efficient uplink massive multi-user MIMO," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 2302–2313, May 2013.

[7] K. T. Truong and R. W. Heath Jr., "Effect of channel aging in massive MIMO systems,"*IEEE J. Commun. Netw.*, vol. 15, no. 4, pp. 338–351, Aug. 2013.

[8] C. Kong, C. Zhong, A. Papazafeiropoulos, M. Matthaiou, and Z. Zhang, "Sum-rate and power scaling of massive MIMO systems with channel aging," *IEEE Trans. Commun.*, vol. 63, no. 12, pp. 4879–4893, Dec. 2015.

[9] C. Kong, C. Zhong, A. Papazafeiropoulos, M. Matthaiou, and Z. Zhang, "Effect of channel aging on the sum rate of uplink massive MIMO systems," in *Proc. IEEE ISIT*, June 2015, pp. 1222–1226.

[10] S. Park, C.-B. Chae, and S. Bahk, "Before/after precoding massive MIMO systems for cloud radio access networks," *IEEE J. Commun. Netw.*, vol. 15, no. 4, pp. 398–406, Aug. 2013.

[11] M. S. Sim, J. Park, C.-B. Chae, and R. W. Heath Jr. "Compressed channel feedback for correlated massive MIMO systems," *IEEE J. Commun. Netw.*, vol. 18, no. 1, pp. 95–104, Feb. 2016.

[12] R. Kudo, S. M. D. Armour, J. P. McGeehan, and M. Mizoguchi, "A channel state information feedback method for massive MIMO-OFDM," *IEEE J. Commun. Netw.*, vol. 15, no. 4, pp. 352–361, Aug. 2013.

[13] C. Kong, C. Zhong, M. Matthaiou, E. Bjornson, and Z. Zhang, "Multi-pair two-way AF relaying systems with massive arrays and imperfect CSI," in *Proc. IEEE ICASSP*, Mar. 2016, pp. 3651–3655.

[14] C. Kong, C. Zhong, M. Matthaiou, and Z. Zhang, "Performance of downlink massive MIMO in Ricean fading channels with ZF precoder," in *Proc. IEEE ICC*, pp. 1776-1782, June 2015.

[15] C. Li, J. Zhang, and K. B. Letaief, "Throughput and energy efficiency analysis of small cell networks with multi-antenna base stations," *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2505–2517, May 2014.

[16] L. Zhao, K. Zheng, H. Long, and H. Zhao, "Performance analysis for downlink massive MIMO system with ZF precoding," *Trans. Emerging Telecom. Technol.*, 2013.

[17] M. Newton and J. Thompson, "Classification and generation of non-uniform user distributions for cellular multi-hop networks," in *Proc. IEEE ICC*, June 2006, pp. 4549–4553.

[18] H. S. Dhillon, R. K. Ganti, and J. G. Andrews, "Modeling non-uniform UE distributions in downlink cellular networks," *IEEE Wireless Commun. Lett.*, vol. 2, no. 3, pp. 339–342, June 2013.

[19] A. Adhikary, H. S. Dhillon, and G. Caire, "Massive-MIMO meets HetNet: Interference coordination through spatial blanking," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 6, pp. 1171–1186, June 2015.

[20] Z. Gong and M. Haenggi, "Interference and outage in mobile random networks: Expectation, distribution, and correlation," *IEEE Trans. Mobile Comput.*, vol. 13, no. 2, pp. 337–349, Feb. 2014.

[21] L. Liang, W. Xu, and X. Dong, "Low-complexity hybrid precoding in massive multiuser MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 653–656, Dec. 2014.

[22] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Intergrals, Serious and Products,* 7th ed. San Diago: Acadamic Press, 2007.

[23] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, New York: Dover Publication Inc., 1974.

[24] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, "On the Lambert W function," *Advances in Computational Mathematics*, vol. 5, no. 1, pp. 329–359, Nov. 1996.

[25] T. D. Novlan, H. S. Dhillon, and J. G. Andrews, "Analytical modeling of uplink cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2669–2679, June 2013.

[26] E. Björnson, L. Sanguinetti, J. Hoydis, and M. Debbah, "Designing multi-user MIMO for energy efficiency: When is massive MIMO the answer?," in *Proc. IEEE WCNC*, Apr. 2014, pp. 242–247.

[27] C. Martin and B. Ottersten, "Asymptotic eigenvalue distributions and capacity for MIMO channels under correlated fading," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1350–1359, July 2004.

[28] M. Matthaiou, C. Zhong, and T. Ratnarajah, "Novel generic bounds on the sum rate of MIMO ZF receivers," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4341–4353, Sept. 2011.

[29] R. M. Gorless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, "On the Lambert W function," *Adv. Comput. Math.*, vol. 5, no. 4, pp. 329–359, Feb. 1996.

Chuili Kong received the B.S. degree in electronics and information engineering from the Dalian University of Technology, Dalian, in 2013. She is currently working towards the Ph.D. degree in information and communication engineering at the Zhejiang University. Her research interests include massive MIMO systems and cooperative communications.

Caijun Zhong received the B.S. degree in Information Engineering from the Xi'an Jiaotong University, Xi'an, China, in 2004, and the M.S. degree in Information Security in 2006, Ph.D. degree in Telecommunications in 2010, both from University College London, London, United Kingdom. From September 2009 to September 2011, he was a Research Fellow at the Institute for Electronics, Communications and Information Technologies (ECIT), Queen's University Belfast, Belfast, UK. Since September 2011, he has been with Zhejiang University, Hangzhou, China, where he is currently an Associate Professor. His research interests include massive MIMO systems, full-duplex communications, wireless power transfer, and physical layer security.

Dr. Zhong is an Editor of the IEEE Transactions on Wireless Communications, IEEE Communications Letters, EURASIP Journal on Wireless Communications and Networking, and Journal of Communications and Networks. He is the recipient of the 2013 IEEE ComSoc Asia-Pacific Outstanding Young Researcher Award. He and his coauthors has been awarded a Best Paper Award at the WCSP 2013. He was an Exemplary Reviewer for IEEE Transactions on Communications in 2014.

Zhaoyang Zhang received his Ph.D. degree in Communication and Information Systems from Zhejiang University, China, in 1998. He is currently a Full Professor with the Department of Information Science and Electronic Engineering, Zhejiang University. His research interests are mainly focused on information theory and coding theory, signal processing for communications and in networks, and their applications in the next generation wireless mobile communication systems. He has co-authored more than 150 refereed international journal and conference papers as well as two books in the above areas. He was a Co-Recipient of several conference Best Paper Awards / Best Student Paper Award. He is currently serving as Editor for IEEE TRANSACTIONS ON COMMUNICATIONS, IET COMMUNICATIONS and some other international journals. He has served or is serving as TPC Co-Chair or Symposium Co-Chair for many international conferences like WCSP'2013, ICUFN'2011/12/13, and GLOBECOM 2014 Wireless Communications Symposium, etc.