Energy Efficiency Scaling Law of Massive MIMO Systems

Wenjia Liu, Shengqian Han, and Chenyang Yang

***Abstract*— Massive multi-input multi-output (MIMO) can support high spectral efficiency with simple linear transceivers, and is expected to provide high energy efficiency (EE). In this paper, we analyze the scaling laws of EE with respect to the number of antennas M at each base station of downlink multi-cell massive MIMO systems under spatially correlated channel, where both transmit and circuit power consumptions, channel estimation errors, and pilot contamination (PC) are taken into account. We obtain the maximal EE for the systems with maximum-ratio transmission and zero-forcing beamforming for given numbers of antennas and users by optimizing the transmit power subject to the minimal data rate requirement and maximal transmit power constraint. The closed-form expressions of approximated EE-maximal transmit power and maximal EE, and their scaling laws with M are derived. Our analysis shows that the maximal EE scales with M in $O(\log_2 M/M)$ for the system without PC, and in $O(1/M)$ for the system with PC. The EE-maximal transmit power scales up with M in $O(\sqrt{M/\ln M})$ until reaching the maximal transmit power for the system without PC, and in $O(1)$ for the system with PC. The analytical results are validated by simulations under a more realistic 3D channel model.**

***Index Terms*— Energy efficiency, scaling law, massive multi-input multi-output (MIMO), pilot contamination.**

I. Introduction

MASSIVE multi-input multi-output (MIMO) is a promising technique for fifth-generation (5G) cellular networks. Simply by aggressively increasing the number of antennas at base station (BS), M, massive MIMO can achieve very high spectral efficiency (SE) with linear transceiver [1]. To demonstrate the potential of using more antennas, the scaling law of SE with M was provided in [2], which is $O\left(\log_2 M\right)$ under favorable propagation conditions, i.e., the channel vectors are asymptotically orthogonal. However, once the pilot contamination (PC) is considered, the signal-to-interference ratio and hence the SE approach to a constant independent from M [1], [2].

Energy efficiency (EE) is one major design goal for 5G networks [3], [4]. To support the same data rate, MIMO systems can achieve higher EE than single antenna systems when only taking into account transmit power [5]. Despite that when the circuit power consumed by the radio frequency (RF) links and complicated signal processing are considered, increasing the number of antennas M will lead to the decrease of the EE of traditional MIMO systems [6], massive MIMO systems are expected to be energy efficient. This comes from the observation that the transmit power can be dramatically reduced thanks to the large array gain and multi-user multiplexing gain [7], and inexpensive components can be used to build the system with low transmit power per antenna when M is very large [8].

To confirm such an intuition, significant research efforts have been made for analyzing the EE of massive MIMO systems in the past several years. The results in [9] show that a massive MIMO system is more energy efficient than traditional MIMO systems only when the average channel gain is small or the circuit power consumption is low. The results in [10]–[13] further show that the EE of massive MIMO system first increases and then decreases with M when the circuit power consumption is considered. Inspired by such result, the number of antennas at each BS was optimized to maximize the EE of massive MIMO systems under various scenarios.

While finding the EE-maximal system parameters is important for designing energy efficient massive MIMO systems, further insights can be obtained from analyzing the asymptotic behavior of the system as M grows. This is because the expression of maximal EE is complicated in general [11], which is unable to reveal its inherent connection with core system parameters analytically. Scaling laws have been widely employed in the literature to show the potential gain of massive MIMO system, e.g, [2], [7], [14]–[16]. Early studies along the line of analyzing energy efficient massive MIMO investigated the uplink transmit power scaling laws with fixed data rate requirement, i.e., how fast the transmit power per user can reduce with the increase of M in order to support a given data rate [7], [14]. Specifically, it was shown in [7] that under independent and identically distributed (i.i.d.) Rayleigh fading channel model, the transmit power per user reduces with the law of $O\left(\frac{1}{M}\right)$ and $O\left(\frac{1}{\sqrt{M}}\right)$ if BSs have perfect and imperfect channel state information (CSI), respectively. In [14], the Ricean channel model with non-zero Ricean K-factor was considered, where the transmit power per user was shown to be scaled down with $O\left(\frac{1}{M}\right)$ even for the case with imperfect CSI at the BS. Different from [7] and [14] that analyze the minimal transmit power required to support a given data rate, the

Manuscript received February 6, 2016; revised July 24, 2016; accepted September 12, 2016. Date of publication September 26, 2016; date of current version January 13, 2017. This work was supported in part by National Basic Research Program of China (No. 2012CB316003), by National High Technology Research and Development Program of China (No. 2014AA01A705), and by National Natural Science Foundation of China (No. 61120106002). The associate editor coordinating the review of this paper and approving it for publication was Z. Zhang.

The authors are with the School of Electronics and Information Engineering, Beihang University, Beijing 100191, China (e-mail: liuwenjia@buaa.edu.cn; sqhan@buaa.edu.cn; cyyang@buaa.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCOMM.2016.2613535

asymptotic behaviors of EE-maximal transmit power at BSs for large M were investigated in [16], where no minimal data rate requirement was considered and both transmit power and circuit power consumption were taken into account. In [16], a single-cell downlink massive MIMO system with perfect CSI and zero-forcing beamforming (ZFBF) was considered, where it was shown that the EE-maximal transmit power at the BS should increase instead of decreasing with the number of antennas at the BS in a scaling law of $O\left(\frac{M}{\ln M}\right)$. However, when the massive MIMO system suffers from interference, e.g., multi-user interference (MUI) caused by imperfect CSI or caused by using maximum ratio transmission (MRT) at BSs, or inter-cell interference (ICI) in multi-cell scenario, the analysis of scaling law of the maximal EE and EE-maximal transmit power becomes rather difficult, due to the intertwined factors impacting the EE, and remains open.

In this paper, we aim to find scaling laws for the EE of downlink multi-cell massive MIMO systems, which can shed light upon the asymptotic behavior of the maximal EE in large number of antennas regime and provide useful guidance for system parameter configuration of energy efficient massive MIMO systems. We strive to answer the following questions: (1) What is the scaling law of the maximal EE of massive MIMO systems with the number of antennas? (2) What is the impact of PC, channel estimation errors, multi-cell setting, and channel correlation on the scaling law of the EE-maximal transmit power with the number of antennas?

In real-world cellular networks, the traffic load in a cell may vary with time (e.g., day and night) and location (e.g., urban and suburban). This implies that in general the number of users is not a parameter that can be configured. Therefore, we only optimize the transmit power of the BSs to maximize the EE of the system. To answer the proposed questions, we derive and analyze the closed-form expressions of the EE-maximal transmit power and maximal EE and their scaling laws with M. Since PC is a limiting factor for massive MIMO systems, which however can be mitigated by some recently-proposed pilot decontamination methods such as [17], we consider both the systems with and without PC. The analytical results are validated by simulations with a realistic channel model, where different-level power consumption parameters, ranging from those in currently deployed BSs to those that might be possible in the future, are taken into account. The results show that the obtained scaling laws hold for moderate number of antennas at each BS. The main contributions of the paper are summarized as follows.

1) We derive the scaling laws of maximal EE and EE-maximal transmit power for downlink multi-cell massive MIMO systems with MRT and ZFBF under spatially correlated channel, considering minimum mean-square error (MMSE) channel estimation and pilot contamination, subject to minimal data rate requirement and maximal transmit power constraint.
2) We show that the EE-maximal transmit power scales up with the number of antennas as $O\left(\sqrt{\frac{M}{\ln M}}\right)$ until reaching the maximal transmit power for the system without PC, while it converges to a constant independent from M for the system with PC. The maximal EE scales with M as $O\left(\frac{\log_2 M}{M}\right)$ and $O\left(\frac{1}{M}\right)$ for the systems without and with PC, respectively.

The rest of the paper is organized as follows. In Section II, we provide the system and power consumption models. In Section III, we find the maximal EE by optimizing the transmit power and derive the scaling laws. In Section IV, we validate the analytical results by simulations. Finally, we conclude the paper.

II. System and Power Consumption Model

A. System Model

Consider a downlink massive MIMO system consisting of L non-coordinated cells, where each BS equipped with M transmit antennas serves K single-antenna users either with MRT or with ZFBF. For the sake of better readability of the paper, we first consider the following assumptions to simplify the complex cellular network:

- *Power allocation*: the transmit power of each BS P is equally allocated to multiple users.
- *Average channel gain*: the average channel gains α including path loss and shadowing from all BSs to all users are identical.
- *Inter-cell interference intensity*: each user is interfered by the adjacent $L-1$ BSs with the same interference intensity χ, where $\chi \in (0, 1]$.
- *Channel correlation*: the K users in each cell have the same channel covariance matrix $\mathbf{R}$.

These assumptions define a simple scenario, which however captures the basic elements of the scaling laws. As will be analyzed later, the scaling laws derived under these assumptions are the same as those under a more general system model without the assumptions, where the power allocation can be arbitrary, and the average channel gains, interference intensity, and correlation matrices among users are different.

Then, the channel from the ℓth BS (denoted by BS_ℓ) to the kth user in the jth cell (denoted by UE_{jk}) is modeled as

$$\mathbf{h}_{\ell jk} = \sqrt{\alpha \chi_{\ell j}}\, \mathbf{R}^{\frac{1}{2}} \tilde{\mathbf{h}}_{\ell jk} \in \mathbb{C}^{M\times 1}, \tag{1}$$

where $\chi_{\ell j} = 1$ if $\ell = j$ and $\chi_{\ell j} = \chi$ if $\ell \neq j$, and $\mathbf{h}_{\ell jk} \sim \mathcal{CN}(\mathbf{0}, \alpha\chi_{\ell j}\mathbf{R})$ with $\mathcal{CN}(\mathbf{m}, \Sigma)$ denoting complex Gaussian distribution with mean $\mathbf{m}$ and covariance matrix Σ. By taking eigenvalue decomposition, we have $\mathbf{R} = \mathbf{U}\Lambda\mathbf{U}^H$, where $\Lambda = \mathrm{diag}\{[\lambda_1, \cdots, \lambda_N]\}$ is an $N \times N$ diagonal matrix whose elements are the nonzero eigenvalues of $\mathbf{R}$, N determines the spatial degrees of freedom of a massive MIMO system which typically increases with M due to the higher angular resolution of a larger array, and $\mathbf{U} \in \mathbb{C}^{M\times N}$ is composed of the eigenvectors of $\mathbf{R}$ corresponding to the nonzero eigenvalues, i.e., $\mathbf{U}^H\mathbf{U} = \mathbf{I}_N$. Then, the channel vector can be re-expressed as $\mathbf{h}_{\ell jk} = \sqrt{\alpha\chi_{\ell j}}\, \mathbf{U}\Lambda^{\frac{1}{2}}\tilde{\mathbf{h}}_{\ell jk}$, where $\tilde{\mathbf{h}}_{\ell jk} \in \mathbb{C}^{N\times 1} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$ is the i.i.d. channel vector, and $\mathbf{I}_N$ is the N-dimensional identity matrix.

We consider a time-division duplex system with block fading channel, where the channels are constant in a time-frequency coherence block with T channel uses and independent among blocks. The channels can be estimated with the

pilot sequences sent by users. During the uplink training phase, K users in the same cell transmit orthogonal pilot sequences each with $T_{\rm tr}$ channel uses and transmit power $P_{\rm tr}$. The pilots received at BS_j will be contaminated if the same set of pilot sequences used in the jth cell are reused in other cells and the traditional training signal and channel estimation methods are applied. Nonetheless, if some pilot decontamination methods such as that in [17] can be applied, the PC will be largely mitigated without increasing the training overhead. To simplify the analysis, we consider two extreme cases for the systems with and without PC as follows.

Denote $\mathcal{S}_j$ as the index set of the cells using the same set of pilot sequences as the jth cell including the jth cell itself, and $L_P \triangleq |\mathcal{S}_j|$ as the number of these cells, where $\mathrm{UE}_{\ell k}$ transmits the same pilot sequence as UE_{jk} for $\ell \in \mathcal{S}_j$. Then: (1) *for the system with PC*, $\mathcal{S}_j = \{1, 2, \ldots, L\}$ and $L_P = L$, where the same set of orthogonal pilot sequences are reused in all the L cells; (2) *for the system without PC*, we set $\mathcal{S}_j = \{j\}$ and $L_P = 1$.

The MMSE channel estimate can be obtained at BS_j as

$$\hat{\mathbf{h}}_{jjk} = \sqrt{\alpha}\mathbf{U}\Lambda\mathbf{V}\cdot\Big(\Lambda^{\frac{1}{2}}\tilde{\mathbf{h}}_{jjk} + \sum_{\ell\in\mathcal{S}_j\backslash\{j\}}\sqrt{\chi_{\ell j}}\Lambda^{\frac{1}{2}}\tilde{\mathbf{h}}_{j\ell k} + \frac{1}{\sqrt{\alpha K T_{\rm tr} P_{\rm tr}}}\mathbf{U}^H\mathbf{n}_{jk}^{\rm tr}\Big), \quad (2)$$

where $\mathbf{V} = \left(\overline{L}_P\Lambda + \frac{1}{\gamma_{\rm tr}}\mathbf{I}_N\right)^{-1}$, $\overline{L}_P \triangleq \sum_{\ell\in\mathcal{S}_j}\chi_{\ell j} = 1 + \chi(L_P - 1)$, the term $\mathcal{S}_j\backslash\{j\}$ denotes the index set of all interfering cells using the same set of orthogonal pilot sequences as the jth cell, $\mathbf{n}_{jk}^{\rm tr} \sim \mathcal{CN}(\mathbf{0}, \sigma_{\rm tr}^2\mathbf{I}_M)$ is the noise at the BS, and $\gamma_{\rm tr} \triangleq \frac{\alpha K T_{\rm tr} P_{\rm tr}}{\sigma_{\rm tr}^2}$, which is defined as the signal to noise ratio (SNR) of the uplink training.

It is not hard to find that $\hat{\mathbf{h}}_{jjk} \sim \mathcal{CN}(\mathbf{0}, \alpha\Phi)$, where $\Phi \triangleq \mathbf{U}\mathbf{Q}\mathbf{U}^H$, $\mathbf{Q} = \Lambda\mathbf{V}\Lambda = \mathrm{diag}\{[q_1, \ldots, q_N]\}$, and $q_i = \frac{\lambda_i^2}{\overline{L}_P\lambda_i + \frac{1}{\gamma_{\rm tr}}}$ is the nonzero eigenvalues of Φ. Comparing the two matrices Φ and $\mathbf{R}$, we can see that their eigenvalues are different but the corresponding eigenvectors are identical. Since $\hat{\mathbf{h}}_{jjk}$ and $\mathbf{h}_{jjk}$ have different eigenvalues under the same eigenvectors, $\lambda_i - q_i$ reflects the channel estimation error in the direction of the ith eigenvector. When uplink training SNR $\gamma_{\rm tr}$ is high, or PC does not exist (i.e., $L_P = 1$), or the interference is weak (i.e., with smaller χ), the value of $\lambda_i - q_i$ is smaller, indicating a more accurate channel estimate.

The downlink signal received at UE_{jk} is given by

$$y_{jk} = \sqrt{\frac{P}{K}}\sum_{\ell=1}^{L}\sum_{m=1}^{K}\mathbf{h}_{\ell jk}^H\frac{\mathbf{w}_{\ell m}}{\|\mathbf{w}_{\ell m}\|}x_{\ell m} + n_{jk}, \quad (3)$$

where $\mathbf{w}_{\ell m} \in \mathbb{C}^{M\times 1}$ is the beamforming vector of BS_ℓ to $\mathrm{UE}_{\ell m}$, $x_{\ell m}$ is the transmit signal with $\mathbb{E}\{|x_{\ell m}|^2\} = 1$, $n_{jk} \sim \mathcal{CN}(0, \sigma^2)$ is the noise at the user, and $\|\cdot\|$ and $\mathbb{E}\{\cdot\}$ denote Euclidean norm and expectation, respectively. For MRT, $\mathbf{w}_{\ell m} = \hat{\mathbf{h}}_{\ell\ell m}$. For ZFBF, $\mathbf{w}_{\ell m}$ is the mth column of $\hat{\mathbf{H}}_{\ell\ell}(\hat{\mathbf{H}}_{\ell\ell}^H\hat{\mathbf{H}}_{\ell\ell})^{-1}$, and $\hat{\mathbf{H}}_{\ell\ell} = [\hat{\mathbf{h}}_{\ell\ell 1}, \ldots, \hat{\mathbf{h}}_{\ell\ell K}]$ is the estimated channel matrix at BS_ℓ for all users in the ℓth cell.

Every user experiences *MUI* (when using ZFBF, it comes from the channel estimation errors), *coherent ICI* caused by PC, and *non-coherent ICI* generated by the signals from the interfering BSs. From (3), the signal to interference plus noise ratio (SINR) of UE_{jk} can be expressed as (4), shown at the bottom of this page, where $\mathcal{K} = \{1, 2, \cdots, K\}$ is the index set of K users, $\{\ell, m|\ell \in \mathcal{S}_j, m \in \mathcal{K}\backslash\{k\}\} \bigcup \{\ell, m|\ell \notin \mathcal{S}_j, m \in \mathcal{K}\}$ denotes the set of users transmitting orthogonal pilot sequences with UE_{jk}, and $\bigcup$ denotes the union of two sets.

Then, the average rate of UE_{jk} with transmit power P at BS_j and system bandwidth B can be obtained as

$$R_{jk}(P) = B\cdot\mathbb{E}\left\{\log_2\left(1 + \gamma_{jk}\right)\right\}. \quad (6)$$

B. Power Consumption Model

The power consumed for downlink transmission by a BS and the circuit powers for operating the BS in transmitting and receiving phases can be modeled as (5), shown at the bottom of this page [18], [19],[1] where $\eta_{\rm PA}$ is the power amplifier (PA) efficiency, $\sigma_{\rm DC}$, $\sigma_{\rm MS}$ and $\sigma_{\rm cool}$ are respectively the loss factors of direct-current to direct-current power supply, main supply and cooling [18], P_{BB_2} and $P_{\rm CE}$ are respectively the baseband processing power consumed for computing beamforming vectors and for channel estimation, $P_{\mathrm{BB}_{1\rm d}}$ and $P_{\mathrm{BB}_{1\rm u}}$ are respectively other baseband power consumption in the downlink transmission phase and in the uplink training phase, and $P_{\rm RF}$ is the RF power consumption. According to [21], P_{BB_2} can be modeled as $P_{\mathrm{BB}_2} = \frac{M(K+\delta_{ZF}K^2)R_{flops,0}}{\eta_C}$, where $R_{flops,0}$ is the

[1]According to the analysis in [19], the circuit power consumption of a massive MIMO BS can be modeled in the same way as a traditional BS, but the corresponding parameters differ. Note that in (5) the feeder loss is removed considering that massive MIMO systems will employ active antennas where the RF module is integrated into each antenna, which is different from a traditional BS with passive antennas [20].

$$\gamma_{jk} = \frac{\left|\frac{\mathbf{h}_{jjk}^H\mathbf{w}_{jk}}{\|\mathbf{w}_{jk}\|}\right|^2}{\underbrace{\sum_{\ell\in\mathcal{S}_j\backslash\{j\}}\left|\frac{\mathbf{h}_{\ell jk}^H\mathbf{w}_{\ell k}}{\|\mathbf{w}_{\ell k}\|}\right|^2}_{\text{coherent ICI caused by PC}} + \underbrace{\sum_{\{\ell,m|\ell\in\mathcal{S}_j, m\in\mathcal{K}\backslash\{k\}\}\bigcup\{\ell,m|\ell\notin\mathcal{S}_j, m\in\mathcal{K}\}}\left|\frac{\mathbf{h}_{\ell jk}^H\mathbf{w}_{\ell m}}{\|\mathbf{w}_{\ell m}\|}\right|^2}_{\text{MUI and non-coherent ICI}} + \frac{K\sigma^2}{P}}, \quad (4)$$

$$P_{\rm BS} = \frac{(1-\frac{KT_{\rm tr}}{T})\frac{P}{\eta_{\rm PA}} + (1-\frac{KT_{\rm tr}}{T})P_{\mathrm{BB}_2} + \frac{KT_{\rm tr}}{T}P_{\rm CE} + M(P_{\rm RF} + (1-\frac{KT_{\rm tr}}{T})P_{\mathrm{BB}_{1\rm d}} + \frac{KT_{\rm tr}}{T}P_{\mathrm{BB}_{1\rm u}})}{(1-\sigma_{\rm DC})(1-\sigma_{\rm MS})(1-\sigma_{\rm cool})}, \quad (5)$$

floating-point operations per second (flops) per antenna for each user, η_C is the power efficiency of computing measured in flops/W, and we use a binary variable δ_{ZF} to differentiate MRT ($\delta_{ZF}=0$) and ZFBF ($\delta_{ZF}=1$) in computational complexity. P_{CE} can be modeled as $P_{\mathrm{CE}}=\frac{M\log_2(KT_{\mathrm{tr}})R_{flops,0}}{\eta_C}$. To differentiate the impact of transmit and circuit power consumptions on EE, we rewrite (5) as

$$\begin{aligned}P_{\mathrm{BS}}&=\underbrace{(1-\tfrac{KT_{\mathrm{tr}}}{T})\eta P}_{\text{transmit power}}\\&\quad+\underbrace{MP_c+M((1-\tfrac{KT_{\mathrm{tr}}}{T})(K+\delta_{ZF}K^2)+\tfrac{KT_{\mathrm{tr}}}{T}\log_2(KT_{\mathrm{tr}}))P_{sp}}_{\text{circuit power}}\\&\approx(1-\tfrac{KT_{\mathrm{tr}}}{T})\eta P+MP_0,\end{aligned}\tag{7}$$

where $\eta\triangleq\frac{1}{\eta_{\mathrm{PA}}(1-\sigma_{\mathrm{DC}})(1-\sigma_{\mathrm{MS}})(1-\sigma_{\mathrm{cool}})}$, $P_0=\frac{P_{\mathrm{RF}}+P_{\mathrm{BB_{1d}}}}{(1-\sigma_{\mathrm{DC}})(1-\sigma_{\mathrm{MS}})(1-\sigma_{\mathrm{cool}})}+(K+\delta_{ZF}K^2)P_{sp}$ is an equivalent circuit power consumption per antenna, $P_c\triangleq\frac{P_{\mathrm{RF}}+(1-\frac{KT_{\mathrm{tr}}}{T})P_{\mathrm{BB_{1d}}}+\frac{KT_{\mathrm{tr}}}{T}P_{\mathrm{BB_{1u}}}}{(1-\sigma_{\mathrm{DC}})(1-\sigma_{\mathrm{MS}})(1-\sigma_{\mathrm{cool}})}$ is an equivalent circuit power consumed at each antenna except beamforming, $P_{sp}\triangleq\frac{R_{flops,0}}{\eta_C(1-\sigma_{\mathrm{DC}})(1-\sigma_{\mathrm{MS}})(1-\sigma_{\mathrm{cool}})}$ is an equivalent circuit power consumed for beamforming at each antenna for each user, and the approximation comes from the similarity of the signal processing power consumed in uplink training and downlink transmission [19].

C. Downlink EE

The downlink EE is defined as the ratio of the average downlink throughput to the total power consumption of the L BSs, where the throughput is the sum rate of all cells excluding the uplink training overhead. From (4) and (7), we can express the downlink EE of the network as

$$\mathrm{EE}(P)=\frac{(1-\frac{KT_{\mathrm{tr}}}{T})\cdot\sum_{j=1}^{L}\sum_{k=1}^{K}R_{jk}(P)}{L\big((1-\frac{KT_{\mathrm{tr}}}{T})\eta P+MP_0\big)}.\tag{8}$$

III. Scaling Laws Analysis

In this section, we derive the scaling laws of the maximal EE and EE-maximal transmit power with M for massive MIMO systems with MRT and ZFBF in a unified form.

A. Average Rate $R_{jk}(P)$

According to the deterministic equivalent approximation in random matrix theory [22], when M and K grow infinitely while $\frac{M}{K}$ is finite, the asymptotic data rate converges in mean square to the average data rate. We use the asymptotic data rate to approximate the average data rate, which is accurate for realistic system dimensions, e.g., $M=10$ and $K=10$ [23]. By using the same approaches as in [23] and [24] but considering different power allocation and channel correlation model, we can derive the approximate average rate of UE_{jk} as

$$R_{jk}(P)\approx B\log_2\left(1+\frac{S}{I_P+I_{nP}+\frac{K\sigma^2}{\overline{q}\alpha P}}\right),\tag{9}$$

where $\overline{q}=\frac{1}{M}\sum_{i=1}^{N}q_i$, and S, I_P, and I_{nP} are respectively the receive powers of the desired signal, coherent ICI, and MUI plus non-coherent ICI, all normalized by $\overline{q}\alpha P$, which are given in (A.1) and (A.4) of Appendix A for the systems using MRT and ZFBF, respectively.

In order to derive the scaling law with M, we need to analyze the asymptotic behaviors of S, I_P, and I_{nP} with M, which can be easily found in the following two special cases. In the first case, the eigenvalue matrix of the channel is $\Lambda=\frac{M}{N}\mathbf{I}_N$, and $\frac{M}{N}\triangleq\beta$ is fixed (when $N=M$, such spatially correlated channel reduces to the i.i.d. channel). In the second case, the systems are with uniform linear array (ULA) and one-ring channel model. In both cases, for the systems using MRT and ZFBF, we can obtain the relationship of S, I_P, and I_{nP} with M (see Appendix A) as[2]

$$\lim_{M\to\infty}\frac{S^{\mathrm{M}}}{M}=\lim_{M\to\infty}\frac{S^{\mathrm{Z}}}{M}=1,\tag{10a}$$

$$\lim_{M\to\infty}\frac{I_P^{\mathrm{M}}}{M}=\lim_{M\to\infty}\frac{I_P^{\mathrm{Z}}}{M}=\chi(\overline{L}_P-1),\tag{10b}$$

$$I_{nP}^{\mathrm{M}}\sim O(1),^{3}\quad I_{nP}^{\mathrm{Z}}\sim O(1).\tag{10c}$$

This indicates that S^{M} and S^{Z} grow with M asymptotically. For the system with PC (i.e., $L_P>1$ and hence $\overline{L}_P=1+\chi(L_P-1)>1$), I_P^{M} and I_P^{Z} grow with M asymptotically. For the system without PC (i.e., $L_P=\overline{L}_P=1$), we have $I_P^{\mathrm{M}}=I_P^{\mathrm{Z}}=0$, and I_{nP}^{M} and I_{nP}^{Z} are independent of M.[4]

B. Maximal EE and EE-maximal Transmit Power

Based on (8) and (9), we next maximize the EE subject to the constraints on maximal transmit power of each BS and minimal rate requirement of each user, where average rather than instantaneous rate constraint is considered since we aim to provide guidance for parameter configuration of energy efficient massive MIMO systems, which should not change frequently with instantaneous channels. The EE maximization problem is formulated as follows.

$$\max_{P}\quad\mathrm{EE}(P)=\frac{(1-\frac{KT_{tr}}{T})BK\log_2\left(1+\frac{SP}{IP+G}\right)}{(1-\frac{KT_{tr}}{T})\eta P+MP_0}\tag{11a}$$

$$s.t.\quad B\log_2\left(1+\frac{SP}{IP+G}\right)\geq R_{\min}\tag{11b}$$

$$0\leq P\leq P_{\max},\tag{11c}$$

where $S=S^{\mathrm{M}}$ and $I=I_P^{\mathrm{M}}+I_{nP}^{\mathrm{M}}$ given in (A.1) for MRT, $S=S^{\mathrm{Z}}$ and $I=I_P^{\mathrm{Z}}+I_{nP}^{\mathrm{Z}}$ given in (A.4) for ZFBF, and $G=\frac{K\sigma^2}{\overline{q}\alpha}$ for both, $R_{\min}$ is the minimal average rate requirement of each user, and $P_{\max}$ is the maximal transmit power at BSs. Constraint (11b) can be rewritten as $P\geq\frac{G}{\frac{S}{2^{R_{\min}/B}-1}-I}\triangleq P_{\min}$,

[2]Throughout the paper, we use the superscripts $()^{\mathrm{M}}$ and $()^{\mathrm{Z}}$ to distinguish the parameters for MRT and ZFBF. For example, herein we use S^{M} to denote the receive signal power for MRT and S^{Z} for ZFBF.

[3]For two functions $f(n)$ and $g(n)$, the notation $f(n)\sim O(g(n))$ means that $\frac{|f(n)|}{|g(n)|}$ remains bounded as $n\to\infty$.

[4]While the forthcoming scaling law analysis depends on such a relation, simulation results show that the analysis is also valid for more general systems and practical channels, as provided later.

where $P_{\min}$ is the minimal transmit power required to satisfy the minimal rate requirement, $P_{\min} \le P_{\max}$, i.e., the minimal rate requirement is assumed achievable.

It is not hard to show that the numerator of $\mathrm{EE}(P)$ in (11a) is concave, the denominator is convex, and both are differential. Hence, $\mathrm{EE}(P)$ is a pseudoconcave function of P. Then, it is easy to find that the optimal solution of problem (11) is $\tilde{P}^* = \min(P_{\max}, \max(P^*, P_{\min}))$ and the maximal EE under the constraint is $\widetilde{\mathrm{EE}}^* = \mathrm{EE}(\tilde{P}^*)$, where P^* is the transmit power maximizing (11a) without the constraints (11b) and (11c). This indicates that the scaling laws essentially depend on how P^* scales with M.

For notational simplicity, in the following we first focus on analyzing P^* and the corresponding EE, denoted by EE^*, and then address the impact of the constraints. To differentiate the notations, we call $\widetilde{\mathrm{EE}}^*$ and $\tilde{P}^*$ the *maximal EE* and *EE-maximal transmit power*, and call EE^* and P^* the *relaxed-maximal EE* and *relaxed-EE-maximal transmit power*.

Problem (11) is general and can be used to derive the results of special cases considered in the literature [7] and [16], where the maximal transmit power constraint is not considered in both works. When each user requires to achieve a fixed data rate $R_{\min}$, i.e., constraint (11b) needs to hold with equality, the relaxed-EE-maximal transmit power can be derived as $P^* = \frac{(2^{R_{\min}/B}-1)G}{S-(2^{R_{\min}/B}-1)I}$. Considering $\lim_{M\to\infty} \frac{S}{M} = 1$ in (10), P^* scales with M as $P^* \sim O(\frac{1}{M})$, which is the same as the result in [7], although [7] considered the transmit power of users in uplink while we consider the transmit power of BSs in downlink. When the system suffers no interference, i.e., $I = 0$ in (11), and no constraints are considered, the relaxed-EE-maximal transmit power can be obtained as $P^* = \frac{G}{S} e^{W\left(\frac{S}{G}\frac{MP_0}{(1-\frac{KT_{\mathrm{tr}}}{T})\eta}\right)+1} - \frac{G}{S}$, where $W(x)$ is Lambert W function [25]. Again considering $\lim_{M\to\infty} \frac{S}{M} = 1$ and $e\frac{x}{\ln x} \le e^{W(x)+1} \le (1+e)\frac{x}{\ln x}$ for $x \ge e$, we can show that $P^* \sim O(\frac{M}{\ln M})$, which is the same as the result in [16].

For general cases where the data rate requirement is not a fixed value and $I > 0$, closed-form expressions of the relaxed-EE-maximal power and the corresponding relaxed-maximal EE have not been found in the literature. We next find P^* based on the *Karush-Kuhn-Tucker* (KKT) condition.

Considering that the circuit power consumption $MP_0 > 0$ in practice, after some regular manipulations we can obtain that P^* satisfies

$$\left(P^* + \frac{MP_0}{(1-\frac{KT_{\mathrm{tr}}}{T})\eta}\right)\frac{SG}{((S+I)P^*+G)(IP^*+G)} - \ln\left(1+\frac{SP^*}{IP^*+G}\right) = 0. \tag{12}$$

By expressing the term $\frac{MP_0}{(1-\frac{KT_{\mathrm{tr}}}{T})\eta}$ as a function of P^* from (12), we can show that its first-order derivative over P^* is positive, i.e., P^* *increases monotonically with* $\frac{MP_0}{(1-\frac{KT_{\mathrm{tr}}}{T})\eta}$. That is to say, to maximize the EE of massive MIMO systems, the relaxed-EE-maximal transmit power should increase with the number of antennas M, equivalent circuit power consumption per antenna P_0, training overhead KT_{tr}, and equivalent power amplifier efficiency $\frac{1}{\eta}$.

In order to obtain the scaling laws of EE^* and P^* with respect to M, we need to find the closed-form expression of P^* from the transcendental equation in (12), whose exact solution is however very hard to obtain. To tackle the difficulty, we introduce some approximations, which are accurate for massive MIMO systems. With the derivations in Appendix B, we show that P^* can be approximated as

$$P^* \approx \sqrt{\frac{MP_0K\sigma^2}{(1-\frac{KT_{\mathrm{tr}}}{T})\eta\alpha\bar{q}}}\sqrt{\frac{\frac{1}{I}-\frac{1}{S+I}}{\ln\left(1+\frac{S}{I}\right)}}, \tag{13}$$

which is accurate for large value of M and becomes more accurate for large value of P_0. By substituting (13) into (11a), EE^* can be approximated as

$$\mathrm{EE}^* \approx \frac{(1-\frac{KT_{\mathrm{tr}}}{T})BK}{P_0} \cdot \frac{\log_2\left(1+\frac{S}{I+\frac{c}{f_{\mathrm{EE}}(M)}}\right)}{M + c\cdot f_{\mathrm{EE}}(M)}, \tag{14}$$

where $c = \sqrt{\frac{(1-\frac{KT_{\mathrm{tr}}}{T})\eta K\sigma^2}{P_0\alpha\bar{q}}}$ is a constant, $f_{\mathrm{EE}}(M) \triangleq \sqrt{\frac{\frac{M}{I}-\frac{M}{S+I}}{\ln\left(1+\frac{S}{I}\right)}}$, $S = S^{\mathrm{M}}$, $I = I_{nP}^{\mathrm{M}} + I_P^{\mathrm{M}}$ for MRT, and $S = S^{\mathrm{Z}}$, $I = I_{nP}^{\mathrm{Z}} + I_P^{\mathrm{Z}}$ for ZFBF.

C. Scaling Laws Analysis

In order to obtain the scaling laws of $\tilde{P}^*$ and $\widetilde{\mathrm{EE}}^*$ with M, we first analyze the scaling laws of P^* and EE^*.

1) Relaxed-EE-maximal Transmit Power: The scaling laws of P^* with M are given in the following proposition.

Proposition 1: For both MRT and ZFBF, P^* scales with M as follows

$$\begin{cases} \lim\limits_{M\to\infty} \dfrac{P^*}{\sqrt{\frac{M}{\ln M}}} \approx \sqrt{\dfrac{P_0K\sigma^2}{(1-\frac{KT_{\mathrm{tr}}}{T})\eta\alpha\bar{q}I_{nP}}}, \\ \qquad\qquad \text{i.e., } P^* \sim O\left(\sqrt{\dfrac{M}{\ln M}}\right), \text{ without PC} \\ \lim\limits_{M\to\infty} P^* \approx \sqrt{\dfrac{P_0K\sigma^2(\frac{1}{\chi(\bar{L}_P-1)} - \frac{1}{1+\chi(\bar{L}_P-1)})}{(1-\frac{KT_{\mathrm{tr}}}{T})\eta\alpha\bar{q}\ln(1+\frac{1}{\chi(\bar{L}_P-1)})}}, \\ \qquad\qquad \text{i.e., } P^* \sim O(1), \text{ with PC} \end{cases} \tag{15}$$

where $I_{nP} = I_{nP}^{\mathrm{M}}$ for MRT and $I_{nP} = I_{nP}^{\mathrm{Z}}$ for ZFBF.

Proof: See Appendix C. ∎

Remark 1: For single-cell massive MIMO system with ZFBF and perfect CSI at the BS and without maximal transmit power constraint, the scaling law $P^* \sim O(\frac{M}{\ln M})$ is obtained in [16]. By setting $L = 1$ in the expression without PC in (15), the scaling law becomes $\sqrt{\frac{M}{\ln M}}$. This indicates that the imperfect CSI will change the scaling law. For the multi-cell massive MIMO system, *Proposition 1* indicates that the relaxed-EE-maximal transmit power increases with $\sqrt{\frac{M}{\ln M}}$ instead of $\frac{M}{\ln M}$ once the system suffers from the interference

TABLE I
SCALING LAWS OF EE-MAXIMAL TRANSMIT POWER AND MAXIMAL EE WITH NUMBER OF ANTENNAS

	$\tilde{P}^*$ for MRT and ZFBF, $M \to \infty$, given K	$\widetilde{\mathrm{EE}}^*$ for MRT and ZFBF, $M \to \infty$, given K
Without PC	$\min\left(\sqrt{\frac{P_0 K\sigma^2}{\left(1-\frac{KT_{\mathrm{tr}}}{T}\right)\eta\alpha\overline{q}I_{nP}}}\sqrt{\frac{M}{\ln M}}, P_{\max}\right) \sim$ first $\mathcal{O}\left(\sqrt{\frac{M}{\ln M}}\right)$, then $\mathcal{O}(1)$	$\frac{\left(1-\frac{KT_{\mathrm{tr}}}{T}\right)BK}{P_0}\frac{\log_2 M}{M} \sim \mathcal{O}\left(\frac{\log_2 M}{M}\right)$
With PC	$\min\left(\sqrt{\frac{P_0 K\sigma^2}{\left(1-\frac{KT_{\mathrm{tr}}}{T}\right)\eta\alpha\overline{q}}\frac{\frac{1}{\chi(\overline{L}_P-1)}-\frac{1}{1+\chi(\overline{L}_P-1)}}{\ln(1+\frac{1}{\chi(\overline{L}_P-1)})}}, P_{\max}\right) \sim \mathcal{O}(1)$	$\frac{\left(1-\frac{KT_{\mathrm{tr}}}{T}\right)BK\log_2\left(1+\frac{1}{\chi(\overline{L}_P-1)}\right)}{P_0}\frac{1}{M} \sim \mathcal{O}\left(\frac{1}{M}\right)$

other than PC, and approaches to a constant independent of M once the system suffers from PC. For massive MIMO system with MRT, MUI always exists, no matter in single-cell or multi-cell system and no matter with perfect or imperfect CSI. Therefore, multi-cell scenario and imperfect CSI do not change the scaling laws of P^* in (15) for systems using MRT.

2) Relaxed-Maximal EE: The scaling laws of EE^* with M are given in the following proposition.

Proposition 2: For both MRT and ZFBF, EE^* scales with M as follows

$$\begin{cases} \lim\limits_{M\to\infty}\dfrac{\mathrm{EE}^*\cdot M}{\log_2 M} \approx \dfrac{(1-\frac{KT_{\mathrm{tr}}}{T})BK}{P_0}, \\ \qquad \text{i.e., } \mathrm{EE}^* \sim \mathcal{O}\left(\dfrac{\log_2 M}{M}\right), \text{without PC} \\ \lim\limits_{M\to\infty}\mathrm{EE}^*\cdot M \approx \dfrac{(1-\frac{KT_{\mathrm{tr}}}{T})BK\log_2(1+\frac{1}{\chi(\overline{L}_P-1)})}{P_0}, \\ \qquad \text{i.e., } \mathrm{EE}^* \sim \mathcal{O}\left(\dfrac{1}{M}\right), \text{with PC} \end{cases} \tag{16}$$

Proof: See Appendix D. ∎

By setting EE^* in (16) as a constant, we can see that if the equivalent circuit power consumption per antenna P_0 can be reduced according to $\mathcal{O}\left(\frac{\log_2 M}{M}\right)$ for the system without PC and $\mathcal{O}\left(\frac{1}{M}\right)$ for the system with PC, then the relaxed-maximal EE will not change with M. Otherwise, the relaxed-maximal EE reduces with M asymptotically, and the relaxed-maximal EE of the system with PC decreases with M with a faster speed than the system without PC.

3) Scaling Laws of EE-maximal Transmit Power $\tilde{P}^*$ and Maximal EE $\widetilde{\mathrm{EE}}^*$: Recall that the optimal solution of problem (11) is $\tilde{P}^* = \min(P_{\max}, \max(P_{\min}, P^*))$ and the corresponding maximal EE is $\widetilde{\mathrm{EE}}^* = \mathrm{EE}(\tilde{P}^*)$, where $P_{\min} = \frac{G}{\frac{S}{2^{R_{\min}/B}-1}-I}$.

Because both S and I_P scale up with M as shown in (10), we know that the denominator of $P_{\min}$ increases with M so that $P_{\min}$ decreases with M. Moreover, (12) and *Proposition 1* suggest that for the system without PC, P^* increases with M, and for the system with PC, P^* first increases with M and then converges to a constant. Therefore, P^* will finally exceed $P_{\min}$ with the increase of M. In other words, the minimal rate requirement in (11b) (and hence $P_{\min}$) does not affect the scaling laws of $\tilde{P}^*$ and $\widetilde{\mathrm{EE}}^*$.

For the system without PC, P^* increases with M and will finally exceed $P_{\max}$. Hence, in this case $\tilde{P}^*$ first scales up with $\mathcal{O}\left(\sqrt{\frac{M}{\ln M}}\right)$ until reaching the maximal transmit power $P_{\max}$. When $\tilde{P}^* = P_{\max}$, we can derive that

$$\begin{aligned}\lim_{M\to\infty}\frac{\widetilde{\mathrm{EE}}^*\cdot M}{\log_2(M)} &= \frac{(1-\frac{KT_{\mathrm{tr}}}{T})BK}{P_0}\lim_{M\to\infty}\frac{\log_2(1+\frac{S}{I_{nP}+\frac{G}{P_{\max}}})/\log_2(M)}{(M+\frac{(1-\frac{KT_{\mathrm{tr}}}{T})\eta P_{\max}}{P_0})/M} \\ &= \frac{(1-\frac{KT_{\mathrm{tr}}}{T})BK}{P_0},\end{aligned}$$

where S increases with M, and $I_{nP}+\frac{G}{P_{\max}}$ and $\frac{(1-\frac{KT_{\mathrm{tr}}}{T})\eta P_{\max}}{P_0}$ are constants independent from M. Further considering (16), $\widetilde{\mathrm{EE}}^*$ scales with $\mathcal{O}(\frac{\log_2 M}{M})$.

For the system with PC, P^* converges to a constant, which makes $\tilde{P}^*$ also be a constant and thus $\widetilde{\mathrm{EE}}^*$ scale with $\frac{1}{M}$.

D. Summary of the Results and Discussion

The scaling laws are summarized in Table I. The results can be explained as follows. When the number of users K is given, for the system without PC, the sum rate achieved by the EE-maximal transmit power increases with $\log_2 M$ thanks to the increased array gain, while for the system with PC, the achieved sum rate approaches to a constant independent of M owing to the coherent ICI. On the other hand, the overall circuit power consumption increases with M. As a consequent, the maximal EE of massive MIMO system with MRT or ZFBF finally decreases with M. Since the training overhead depends on K, the maximal EE does not simply increase with the multiplexing gain of K. In fact, as have been analyzed in [16] and [21], there exists an optimal number of users for fixed values of M, P and M/K.

Remark 2: The obtained EE-maximal transmit power scaling laws are very different from the power scaling laws in [7] and [14]. This is because [7] and [14] investigated the minimal transmit power of users to guarantee the fixed data rate requirement, while we study the optimal transmit power of BSs to maximize the system EE under the minimal data rate requirement (not fixed). If we let the rate requirement hold with equality, then maximizing EE is equivalent to minimizing transmit power, otherwise, they are different leading to different power scaling laws. In addition, we consider downlink transmission while [7] and [14] considered uplink transmission.

Remark 3: We can observe from Table I that the maximal EE, $\widetilde{\mathrm{EE}}^*$, decreases with M, and the EE-maximal transmit power, $\tilde{P}^*$, scales up with M until reaching the maximal power of BSs in the case without PC and is a constant in the case

with PC. However, this does not mean that $\widetilde{\mathrm{EE}}^*$ decreases more when the BSs transmit with $\tilde{P}^*$ than with other powers. Instead, if the transmit power of a system is not configured according to the derived scaling laws, the EE of the system will not be maximized and will reduce much faster with M than the derived EE scaling laws.

Based on these analysis, we can answer the questions in the introduction as follows.

(1) Given K and P_0, the maximal EE decreases with the increase of M, no matter if the massive MIMO system is with or without PC. In order to maximize the EE with minimal data rate requirement and maximal transmit power constraint, for the system without PC, the EE-maximal transmit power should increase with M until reaching the maximal transmit power according to the laws in Table I. For the system with PC, the EE-maximal transmit power should first increase with M and then converge to $\sqrt{\frac{P_0 K \sigma^2}{(1-\frac{KT_{\mathrm{tr}}}{T})\eta\alpha\overline{q}} \frac{\frac{1}{\chi(\overline{L}_P-1)}-\frac{1}{1+\chi(\overline{L}_P-1)}}{\ln(1+\frac{1}{\chi(\overline{L}_P-1)})}}$ or $P_{\max}$ according to the laws in Table I.

(2) In single-cell scenario, the channel estimation errors change the scaling law of $\tilde{P}^*$ for massive MIMO system with ZFBF and do not change the scaling law of $\tilde{P}^*$ for system with MRT. In multi-cell scenario, PC changes the scaling laws of $\tilde{P}^*$ and $\widetilde{\mathrm{EE}}^*$ for both massive MIMO systems with MRT and ZFBF. Channel correlation only affects the power scaling laws with a multiplicative factor via $\overline{q}$ and I_{nP}.

E. Impacts of Assumptions and Approximations

In what follows, we discuss the impact of the underlying assumptions and approximations on the scaling law analysis.

1) Impacts of Assumptions: In the analysis, we considered a simplified system model with the four assumptions stated in Section II-A. In order to show the impact of these assumptions on the scaling laws, in the following we consider a general channel model $\mathbf{h}_{\ell jk} = \sqrt{\alpha_{\ell jk}}\mathbf{R}_{\ell jk}^{\frac{1}{2}}\tilde{\mathbf{h}}_{\ell jk}$ with different average channel gain α_{jjk}, different inter-cell interference intensity $\frac{\alpha_{\ell jk}}{\alpha_{jjk}}$, and different covariance matrix $\mathbf{R}_{\ell jk}$. We also consider a general power allocation policy $p_{jk} = \mu_{jk}P$ with arbitrary μ_{jk} satisfying $0 \le \mu_{jk} \le 1$ and $\sum_{k=1}^{K}\mu_{jk} = 1$. Then, by using the same approach to obtain (9), we can derive the approximate average rate of UE_{jk} as

$$R_{jk}(P) \approx B\log_2\left(1+\frac{S_{jk}P}{(I_{jk,P}+I_{jk,nP})P+\sigma^2}\right), \quad (17)$$

where $I_{jk,P} \triangleq \sum_{\ell\in\mathcal{S}_j\setminus\{j\}}\alpha_{\ell jk}\overline{q}_{\ell jk}\mu_{\ell k}M$, $I_{jk,nP} \triangleq \sum_{\{\ell,m|\ell\in\mathcal{S}_j,m\in\mathcal{K}\setminus\{k\}\}\bigcup\{\ell,m|\ell\notin\mathcal{S}_j,m\in\mathcal{K}\}} I_{\ell mjk}\mu_{\ell m}$, and $S_{jk} \triangleq \alpha_{jjk}\overline{q}_{\ell jk}\mu_{jk}M$ are respectively the asymptotic receive powers of the coherence ICI, non-coherent ICI plus MUI, and desired signal at UE_{jk}, and $\overline{q}_{\ell jk} = \frac{\alpha_{\ell jk}(\frac{1}{M}\mathrm{tr}(\mathbf{R}_{\ell jk}))^2}{\alpha_{\ell\ell k}^2\frac{1}{M}\mathrm{tr}(\mathbf{R}_{\ell\ell k}(\sum_{i\in\mathcal{S}_\ell}\alpha_{\ell ik}\mathbf{R}_{\ell ik}+\frac{1}{KP_{\mathrm{tr}}T_{\mathrm{tr}}}\mathbf{I}_M)^{-1}\mathbf{R}_{\ell\ell k}^H)}$ is independent from M asymptotically according to the same reason as in Appendix A.

Denote $u^{\max} \triangleq \arg\max_{j,k}\{R_{jk}(P)\}$ and $u^{\min} \triangleq \arg\min_{j,k}\{R_{jk}(P)\}$ as the index of the user achieving the maximal and minimal average data rate among all the LK users in the network, respectively. We can bound the average data rate of UE_{jk} as

$$B\log_2\left(1+\frac{S_{u^{\min}}P}{(I_{u^{\min},P}+I_{u^{\min},nP})P+\sigma^2}\right) \le R_{jk}(P) \le B\log_2\left(1+\frac{S_{u^{\max}}P}{(I_{u^{\max},P}+I_{u^{\max},nP})P+\sigma^2}\right). \quad (18)$$

According to (17), similar to (10), we know that $S_{u^{\min}} \sim O(M)$, $S_{u^{\max}} \sim O(M)$, $I_{u^{\min},nP} \sim O(1)$, and $I_{u^{\max},nP} \sim O(1)$; for the system with PC, $I_{u^{\min},P} \sim O(M)$ and $I_{u^{\max},P} \sim O(M)$; and for the system without PC, $I_{u^{\min},P} = I_{u^{\max},P} = 0$. Substituting $S = S_{u^{\max}}$, $I = I_{u^{\max},P} + I_{u^{\max},nP}$ and $G = \sigma^2$ into (13) and (14), we can obtain the upper bound of the approximate relaxed-EE-maximal transmit power and the approximate relaxed-maximal EE for the system with general channel model and power allocation policy. Similarly, after substituting $S = S_{u^{\min}}$, $I = I_{u^{\min},P} + I_{u^{\min},nP}$ and $G = \sigma^2$ into (13) and (14), the lower bound is obtained.

Because the scaling laws of $S_{u^{\max}}$, $I_{u^{\max},P} + I_{u^{\max},nP}$, $S_{u^{\min}}$ and $I_{u^{\min},P} + I_{u^{\min},nP}$ with M are the same as those in (10), the resulting scaling laws of the upper bound and lower bound obtained from (18) will be the same as those in *Proposition 1* and *2*. Therefore, the resulting scaling laws of the EE-maximal transmit power and maximal EE obtained from the average data rate $R_{jk}(P)$ in (17) will also be the same as those in Table I. In other words, the scaling laws we obtained are valid for general channel models and general power allocation policies.

2) Impacts of Approximations: Now we summarize the approximations used in the analysis and discuss their accuracy.

- In Section III-A, the average rate is approximated by the asymptotic rate. This approximation is accurate for realistic system dimension, e.g., $M = K = 10$, as reported in [23], and becomes more accurate for large values of M and K.
- The relaxed-EE-maximal transmit power is approximated as in (13). According to the derivations in Appendix B, we can find that the approximation is accurate when $P^* \gg \frac{G}{I}$ and $\frac{MP_0}{(1-\frac{KT_{\mathrm{tr}}}{T})\eta} \gg \max(\frac{G}{I}, (\frac{G}{4I}-\frac{G}{4(S+I)}+\frac{G}{S})\ln(1+\frac{S}{I}))$. According to (10), we know that I either increases with M for the system with PC or is independent of M for the system without PC, and S always increases with M. From (12), we know P^* increases with M. Then, the left-hand-side of the above two inequalities increases with M, while the right-hand-side decreases with M for system with PC, and remains constant (the first inequality) or increases with $\ln M$ (the second inequality) for the system without PC. As a consequence, the two inequalities hold for large M, i.e., P^* in (13) and hence EE^* in (14) are accurate. For example, for the system with ULA and one-ring channel model with the angle of arrival (AoA) $\theta = 0^\circ$ and angle of spread (AS) $\Delta = 15^\circ$, $K = 10$, $\frac{\alpha}{\sigma^2} = 0.9$, $\chi = 0.3$, and the power consumption parameters given in Section IV, we find via simulations that

the approximate relaxed-EE-maximal transmit power and approximate relaxed-maximal EE are nearly the same as their accurate values for $M \geq 32$.

- The scaling laws in *Proposition 1* and *Proposition 2* are obtained from (13) and (14), respectively. Through regular derivation, we can find that the scaling laws hold when M is large so that the conditions $M - c\sqrt{\frac{I_{nP}\ln M}{M}} \gg I_{nP}$ and $M\ln M \gg \frac{c^2}{I_{nP}}$ hold for the system without PC and $M \gg \max(\frac{I_{nP}}{\chi(\overline{L}_P-1)} + c\sqrt{\frac{\ln(1+\frac{1}{\chi(\overline{L}_P-1)})}{\chi(\overline{L}_P-1)-\frac{\chi^2(\overline{L}_P-1)^2}{1+\chi(\overline{L}_P-1)}}}, c\sqrt{\frac{\frac{1}{\chi(\overline{L}_P-1)}-\frac{1}{1+\chi(\overline{L}_P-1)}}{\ln(1+\frac{1}{\chi(\overline{L}_P-1)})}})$ holds for the system with PC. Using the parameters listed above, we can compute that the scaling laws are accurate for $M \geq 400$.

F. A Byproduct: SE Loss Analysis

Since massive MIMO can support high SE, a natural concern is whether the operating point of $\tilde{P}^*$ that achieves the maximal EE will cause a SE loss. Such a concern can be addressed as a byproduct of the closed-form expression of $\tilde{P}^*$ we obtained. Specifically, we can analyze the gap between the EE-maximal data rate achieved by $\tilde{P}^*$ and the maximal achievable rate of ZFBF or MRT obtained by setting $P = P_{\max}$. By substituting (13) into (9), the EE-maximal average sum rate of the jth cell with MRT and ZFBF can be obtained as a unified expression, which is

$$\tilde{R}_j^* \triangleq \sum_{k=1}^{K}(\tilde{P}^*) \approx BK\log_2\left(1+\frac{S\tilde{P}^*}{I\tilde{P}^*+G}\right), \tag{19}$$

and the maximal achievable rate of the jth cell with MRT and ZFBF can be obtained as,

$$R_{j,\max} \triangleq \textstyle\sum_{k=1}^{K} R_{jk}(P_{\max}) \approx BK\log_2\left(1+\frac{SP_{\max}}{IP_{\max}+G}\right). \tag{20}$$

The gap between them can be derived as

$$\Delta_R \triangleq R_{j,\max} - \tilde{R}_j^* \approx BK\log_2\left(1+\frac{\frac{G}{\tilde{P}^*}-\frac{G}{P_{\max}}}{I+\frac{G}{P_{\max}}}\right), \tag{21}$$

where $I = I_P^{\mathrm{M}} + I_{nP}^{\mathrm{M}}$ for MRT and $I = I_P^{\mathrm{Z}} + I_{nP}^{\mathrm{Z}}$ for ZFBF, and the approximation comes from $\log_2(1+\gamma) \approx \log_2(\gamma)$ and is accurate for large SINR γ which is true in massive MIMO systems.

Proposition 3: For the massive MIMO system with given numbers of antennas and users, the EE-maximal average rate achieved by $\tilde{P}^*$ is close to the maximal average rate achieved by $P_{\max}$ when M is large. When $M \to \infty$, the gap between them approaches to zero.

Proof: See Appendix E. ∎

This implies that supporting the maximal EE of a massive MIMO system with given values of M and K by configuring transmit power will cause a little loss of sum rate compared to the maximal sum rate achieved by ZFBF or MRT. This is sharply in contrast to the SE-EE relationship investigated in [26], which can be interpreted as: *massive MIMO is not energy efficient to achieve very high SE*, because a linear increase of SE leads to an exponential decrease of EE. Such a difference in conclusion comes from the system model. No any interference and maximal transmit power constraint were considered in [26], while we have considered MUI, ICI, PC, and maximal transmit power constraint.

Remark 4: The reasons causing the small SE loss are different for the cases with and without PC. For the system with PC, we have shown that the EE-maximal transmit power is a constant not increasing with M and the coherent ICI I_P increases with M monotonically. In this case, for large M, PC makes the power of ICI much higher than the noise, i.e., $I\tilde{P}^* \gg G$. Therefore, according to (19) and (20), both the EE-maximal average rate $\tilde{R}_j^*$ and the maximal average rate $R_{j,\max}$ approach to $BK\log_2(1+\frac{S}{I})$, i.e., the rate gap converges to zero. For the system without PC, we have shown that when M increases, the EE-maximal transmit power $\tilde{P}^*$ increases and finally reaches $P_{\max}$, making $\tilde{R}_j^* = R_{j,\max}$.

IV. Numerical and Simulation Results

In this section, we validate previous analytical results via simulations.

Consider a massive MIMO system consisting of seven macro cells each with a radius of 250 m, where one central cell is surrounded by six cells. In each cell 10 users are located randomly with a minimum distance of 35 m from the BS. The path loss model is set as $35.3+37.6\log_{10} d$ dB, where d (in m) is the distance between a user and a BS. Both the uplink and downlink noise variances are -174 dBm/Hz [27], and the system bandwidth is 20 MHz. The length of a coherence block is $T = T_cB_c = 380$ channel uses, where $T_c = 3.8$ ms is the coherence time for 60 km/h moving speed and 2 GHz carrier, and $B_c = 100$ kHz is the coherence bandwidth. For uplink training, $P_{\mathrm{tr}} = 200$ mW, $T_{\mathrm{tr}} = 1$ for each user, and MMSE channel estimator are considered. The minimal rate requirement of each user is 5 Mbps and the maximal transmit power of BSs is 40 W (the maximal transmit power for a Long-term Evolution (LTE) macro BS). Since in the literature there are no specific power consumption parameters for a massive MIMO BS, we employ the parameters for a BS in the year 2012 (i.e., $P_c = 0.79$ W and $P_{sp} = 1.9$ mW) and the predicted values for massive MIMO BS in the year 2020 provided by GreenTouch consortium [19] (i.e., $P_c = 0.16$ W and $P_{sp} = 0.12$ mW). $\eta = 2.51$ (obtained from the PA efficiency $\eta_{\mathrm{PA}} = 50\%$ [19]), $\sigma_{\mathrm{DC}} = 6\%$, $\sigma_{\mathrm{cool}} = 9\%$, and $\sigma_{\mathrm{MS}} = 7\%$ (for macro BS [18]) are the same for both years of 2012 and 2020. The performance of the central cell is evaluated. Unless otherwise specified, these simulation setups will be used throughout the simulations.

A. Accuracy of the Approximations

In Fig. 1, we show the impact of the approximations on the closed-form expressions of P^* in (13) and EE* in (14). To this end, we compare the relaxed-EE-maximal transmit powers and relaxed-maximal EEs respectively obtained from the following approaches, where the channel model in (1) with $\alpha = \frac{10^{-3.53}}{250^{3.76}}$,

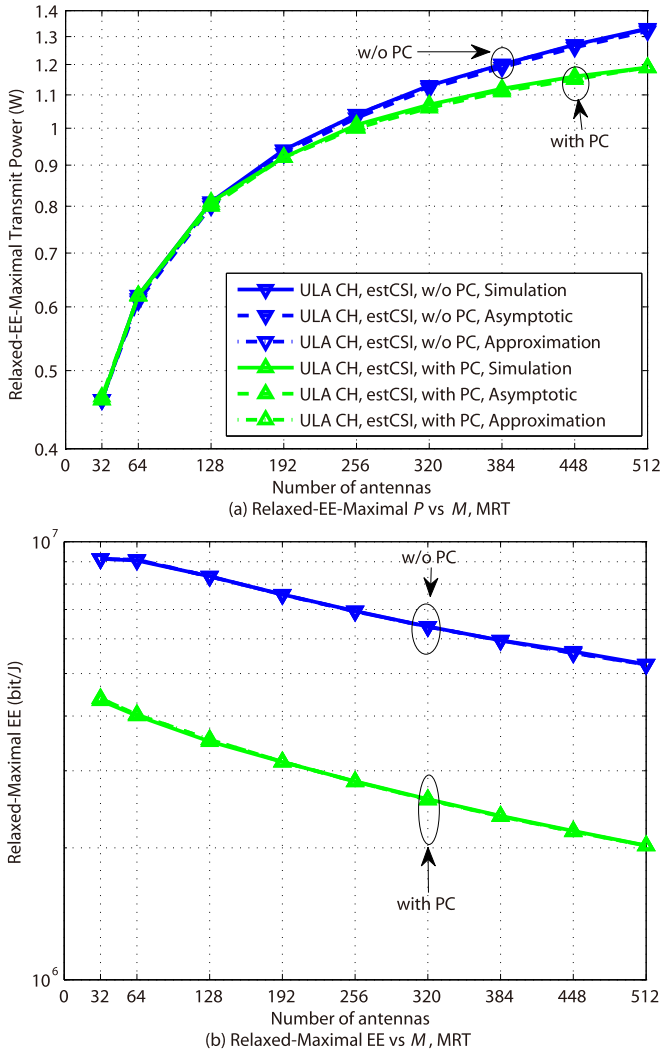


Fig. 1. Evaluation of the approximations, $P_c = 0.16$ W and $P_{sp} = 0.12$ mW (for the year of 2020), with MRT. The three curves for relaxed-EE-maximal transmit power and relaxed-maximal EE in either the case with PC or without PC are overlapped.

$\chi = 0.3$, and one-ring model for ULA with $\theta = 0°$ and $\Delta = 15°$ are used.

a) The optimal solution (with legend "Simulation") is obtained by searching from an EE maximization problem without any approximations, where the average data rate per user is obtained from simulation by averaging over small scale channels.
b) The numerical results (with legend "Asymptotic") are found from optimization problem (11) with no constraints by bisection searching, where the asymptotic data rates are used.
c) The approximated solutions (legend "Approximation") are computed with (13) and (14).

Since lower circuit power consumption will make the approximations less accurate according to Appendix B, we provide the results for the predicted parameters in 2020. Nevertheless, even in this worse case, we can see that for massive MIMO systems with MRT, the approximations are very accurate. The results for the system with ZFBF are similar and are not shown to make the figure clear.

B. Validation of the Scaling Laws With Realistic Channel Model

In order to demonstrate that the analytical results are also valid for more realistic channels, *in the subsequent subsections* we simulate the performance of massive MIMO systems by averaging over both large-scale and small-scale channels, where a three-dimensional (3D) MIMO system with uniform rectangular array in urban macro (UMa) scenario (referred as 3D UMa) is employed [28], considering that such a large number of antennas may not be arranged as a linear array. The lognormal distributed shadowing with 4 dB standard deviation is considered in the 3D UMa channel, and the users are randomly placed in each cell. With such a channel, the large scale channel gains, interference intensity, and channel covariance matrices of multiple users are different, depending on their locations. Besides, such a channel does not belong to the two special cases mentioned in Section III-A. The antenna spacing at the BS is half of the wavelength for both horizontal and vertical directions. The main 3D MIMO channel parameters are listed in Table II.

In Fig. 2, we compare the scaling laws with the simulation results of the EE-maximal transmit power versus M in five scenarios to observe the impacts of channel correlation, channel estimation errors, non-coherent ICI, and coherent ICI, respectively. The scaling law curves are obtained as follows. Take Fig. 2(a) as example, where the scaling law is $O(\sqrt{\frac{M}{\ln M}})$ for the case without PC before reaching the maximal transmit power. We plot the scaling law curve according to $\nu\sqrt{\frac{M}{\ln M}}$, where ν is the positive real scalar and hence does not affect the slope of the curve, which is chosen to make the scaling law curve close to the simulated curves in order to facilitate the comparison. The circuit power parameters for the year of 2020 are considered, while the results of the system with the circuit power parameters for 2012 are similar.

We can see from Fig. 2(a) that for the single-cell system with perfect CSI, the EE-maximal transmit power under i.i.d. and 3D UMa channels are nearly the same, which indicates that spatially correlated channel has little impact on the power scaling law of the system with ZFBF. With the channel estimation errors or non-coherent ICI, the EE-maximal transmit power increases with M in the law of $\sqrt{\frac{M}{\ln M}}$ and increases much slower than the system with perfect channel information, which is consistent with *Remark 1*. With the coherent ICI caused by PC, the EE-maximal transmit power approaches to a constant for large M (which will be obvious if the y-axis is in linear scale), though very slowly in 3D UMa channel due to less interference.

We can see from Fig. 2(b) that the EE-maximal transmit power of the system with MRT increases with M proportional to $\sqrt{\frac{M}{\ln M}}$ for all scenarios except with PC, which approaches a constant slowly. In contrast to the results for ZFBF, the EE-maximal transmit power of the system with i.i.d channel is much lower than that with 3D UMa channel, which means that the spatially correlated channel has large impact on the power scaling law of the system with MRT. This is because the system under the 3D UMa channel suffers less interference than that under the i.i.d. channel.

TABLE II
MAIN PARAMETERS OF 3D MIMO CHANNELS

Mean of azimuth clusters	$U(-60°, 60°)$	Mean of elevation clusters	$U(-45°, 45°)$
Mean of log delay spread (DS) ($\log_{10}([s])$)	-6.62	Variance of log DS	0.32
Mean of log azimuth spread (AS) ($\log_{10}([°])$)	1.25	Variance of log AS	0.42
Mean of log elevation spread (ES) ($\log_{10}[°]$)	$\max[-0.5, -2.1\frac{d}{1000} + 0.9]$	Variance of log ES	0.49
Number of clusters	12	Number of rays per cluster	20

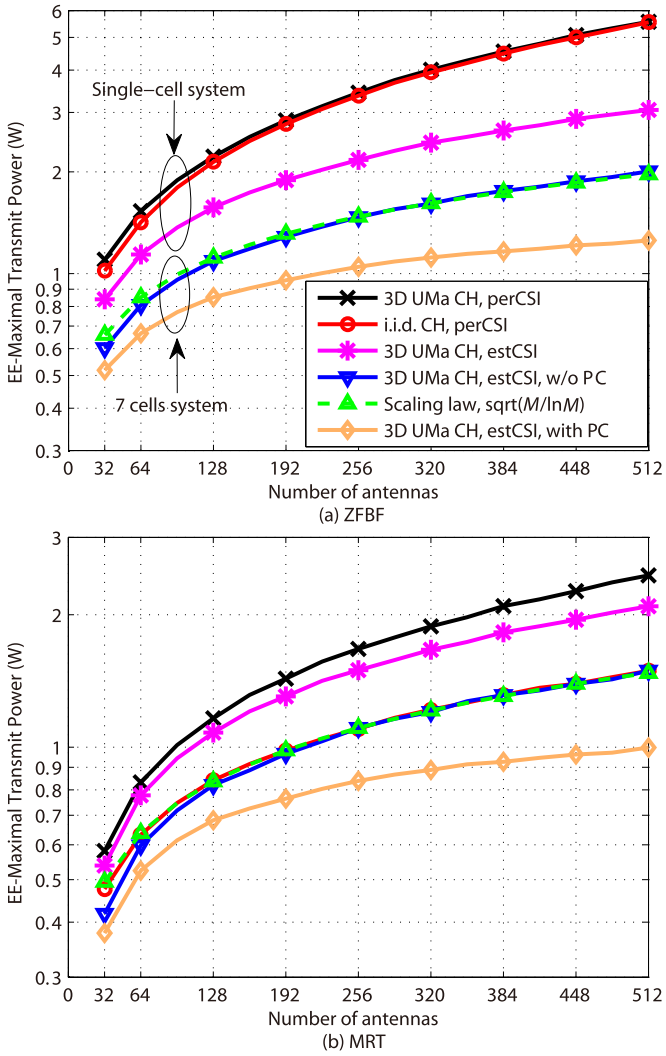


Fig. 2. EE-maximal transmit powers versus M, and the legends "perCSI" and "estCSI" respectively denote the results with perfect and estimated channels. $P_c = 0.16$ W and $P_{sp} = 0.12$ mW (for the year of 2020).

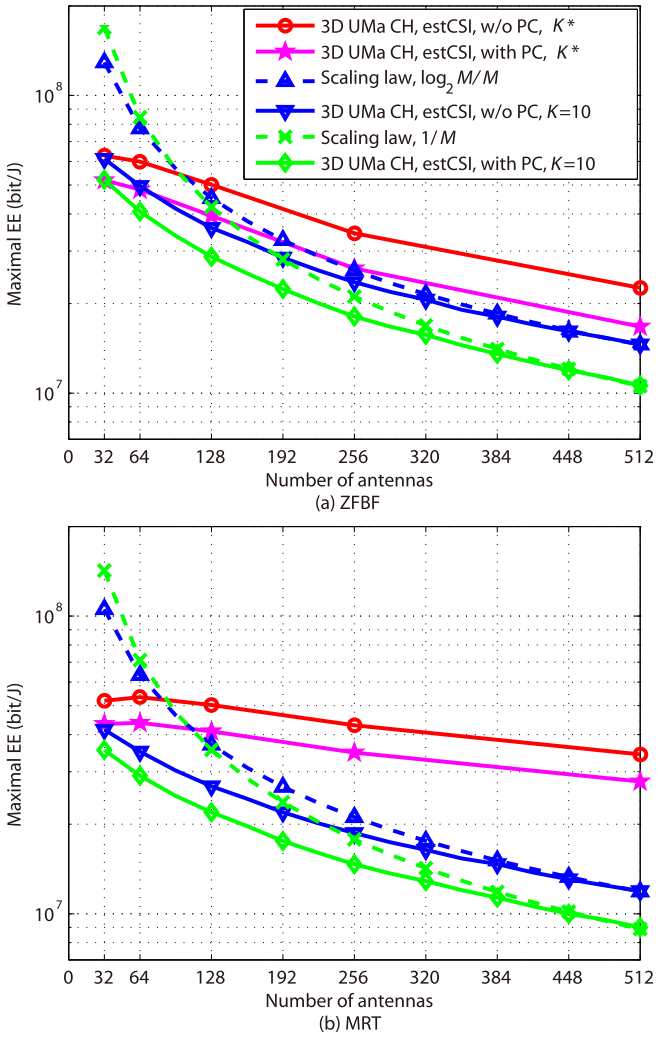


Fig. 3. Maximal EEs versus M under 3D UMa channel model. $P_c = 0.16$ W and $P_{sp} = 0.12$ mW (for the year of 2020).

Comparing Figs. 2(a) and (b), we can find that the EE-maximal transmit power $\tilde{P}^*$ of the system with ZFBF is much higher than that with MRT. For both MRT and ZFBF, $\tilde{P}^* < 6$ W, which is much lower than the maximal transmit power of LTE macro BSs, 40 W. This indicates that maximal transmit power constraint in problem (11) does not affect the scaling laws of EE-maximal transmit power with M and hence does not affect the scaling laws of maximal EE as well for systems with moderate number of antennas. Besides, the EE-maximal average rate per user $\tilde{R}^*_{jk}$ in the five scenarios are higher than 22 Mbps and 36 Mbps for the systems using MRT and ZFBF with $M \geq 32$, respectively, which is higher than the minimal rate requirement per user of 5 Mbps. Therefore, the minimal rate requirement does not affect the scaling laws.

In Fig. 3, we compare the scaling laws with the simulation results of the maximal EE versus M, where the power consumption parameters in 2020 are used (the results with the parameters in 2012 are similar). The scaling law curves in Fig. 3 are plotted by the same method with Fig. 2. Since there may exist large number of users in a macro cell in practice and we can employ an optimal user scheduler to improve EE, we also simulate the maximal EE of the system with the optimal number of users K^*, which is obtained by jointly searching the optimal (K, P) to maximize the EE satisfying minimal rate requirement per user and maximal transmit power constraint when the number of antennas M is given.

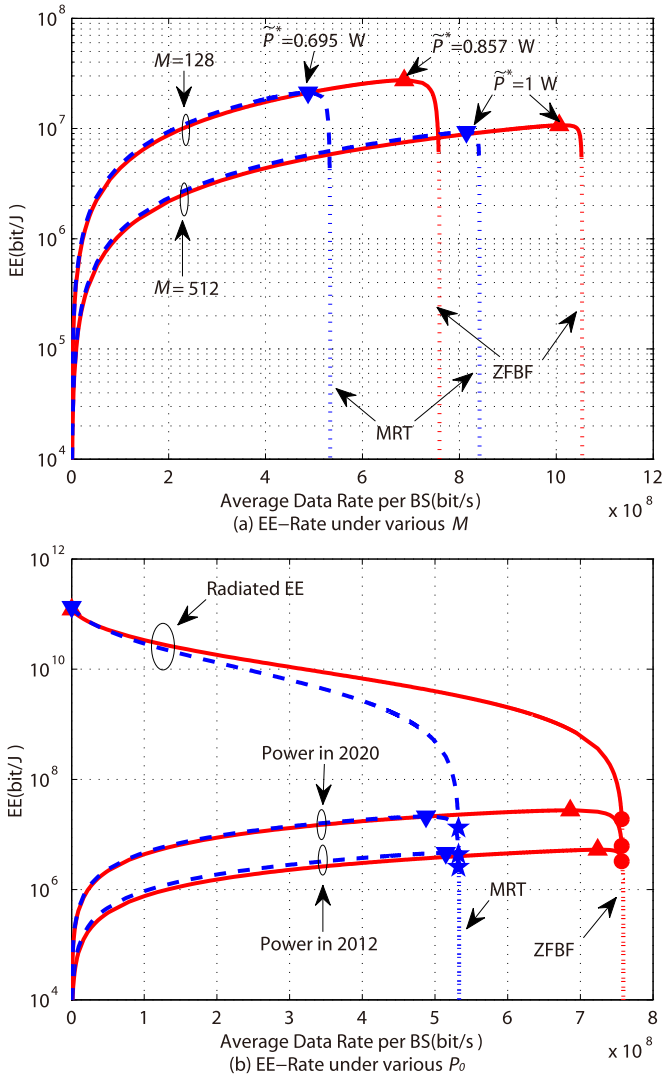


Fig. 4. EE versus sum rate per cell. The dashed and solid lines represent the results for MRT and ZFBF, and the maximal EEs for MRT and ZFBF are marked by lower and upper triangle, respectively. In Fig. 4(a), $P_c = 0.16$ W and $P_{sp} = 0.12$ mW (for the year of 2020). In Fig. 4(b), $M = 128$, the EEs achieved by $P = 40$ W are marked by stars for MRT and bullets for ZFBF. In both Figs. 4(a) and (b), $L = 7$, and 3D UMa channel model is employed.

It can be observed that for a given number of users, the maximal EE decreases with M in proportional to $\frac{\log_2 M}{M}$ for the system without PC and $\frac{1}{M}$ for the system with PC for large M (e.g., $M > 384$), which are consistent with the analytical results. When the number of users is optimized, the maximal EE still decreases with M, but the decreasing speed is much slower for the system with MRT. This is because the optimal number of users in the system with MRT increases with M faster and is much larger than the system with ZFBF, owing to the fact that the signal processing power increases with K in the system with MRT while increases with K^2 in the system with ZFBF as given in (7).

C. The Byproduct: EE versus Rate

In Fig. 4, we validate *Proposition 3*. The average sum rate per cell and the corresponding EE of the system are obtained from simulations, where P increases from 0 W without limit and $\tilde{P}^*$ is found by exhaustive searching. To show the feasible region of the rate achievable by a macro BS with a maximal transmit power of 40 W, the results with $P > 40$ W are plotted with dotted curves.

In Fig. 4(a), we show the impact of the number of transmit antennas M, where the power consumption parameters for the year 2020 are considered (the results with the parameters for 2012 are similar). Due to the MUI, non-coherent ICI, and coherent ICI, the data rates of the massive MIMO systems are limited to finite values when $P \to \infty$. The maximal EEs are achieved approximately at the maximal achievable rates of ZFBF and MRT, which agrees with *Proposition 3*. Before the sharp reduction, the EEs achieved by MRT and ZFBF are very close, which can be interpreted by their similar level of circuit powers (with only difference in P_{sp}). The maximal EE achieved by ZFBF exceeds that by MRT due to the higher achievable rate. Again, we can see that $\tilde{P}^* \ll 40$ W, and increasing the transmit power beyond $\tilde{P}^*$ will reduce EE significantly. In Fig. 4(b), we show the impact of P_c and P_{sp}. It is shown that with the increase of circuit power consumption, $\tilde{R}_j^*$ moves closer to $R_{j,\max}$, which agrees with the analysis. Besides, our results are quite distinct from the radiated EE (without considering the circuit power) in the EE-rate relation.

V. Conclusions

In this paper, we investigated the scaling laws of the EE-maximal transmit power and maximal EE for downlink multicell massive MIMO system with and without pilot contamination under spatially correlated channel model. We derived the closed-form expressions of approximate maximal EEs of the massive MIMO systems with MRT and ZFBF, and then analyzed the scaling laws of the EE-maximal transmit power and maximal EE with the number of transmit antennas M. Analytical results showed that the maximal EE of massive MIMO systems asymptotically decreases with M as long as the circuit power consumption is a constant, no matter how low it is. The pilot contamination is the dominating factor in determining the scaling laws. The power scaling law suggests that the EE-maximal transmit power should be configured to increase with M in order to maximize the EE until reaching the maximal transmit power of the BSs for the system without pilot contamination, but should be configured as a constant independent from M for the system with pilot contamination. Channel correlation has large impact on the power scaling law for massive MIMO with MRT, and has minor impact on that with ZFBF. Simulation results under realistic channel model validated the analytical analysis.

Appendix A
Proof of the Asymptotic Behaviors of S, I_P, and I_{nP} With M in (10)

By using the same approaches as in [23], for systems with MRT and arbitrary covariance matrix $\mathbf{R}$, we can derive that

$$S^{\mathrm{M}} = M + \frac{\overline{q(\lambda - q)}}{\overline{q}^2}, \quad I_P^{\mathrm{M}} = \left(M\chi + \frac{\overline{q(\lambda - \chi q)}}{\overline{q}^2} \right) (\overline{L}_P - 1), \tag{A.1a}$$

$$I_{nP}^{\mathrm{M}} = (K\overline{L} - \overline{L}_P)\frac{\overline{q\lambda}}{\overline{q}^2}, \tag{A.1b}$$

where $\overline{q} = \frac{1}{M}\sum_{i=1}^{N} q_i$, $\overline{q(\lambda-q)} = \frac{1}{M}\sum_{i=1}^{N} q_i(\lambda_i - q_i)$, $\overline{q(\lambda-\chi q)} = \frac{1}{M}\sum_{i=1}^{N} q_i(\lambda_i - \chi q_i)$, $\overline{q\lambda} = \frac{1}{M}\sum_{i=1}^{N} q_i\lambda_i$, and $\overline{L} \triangleq \sum_{\ell=1}^{L}\chi_{\ell j} = 1 + \chi(L-1)$.

Next we prove the asymptotic behaviors of S^{M}, I_P^{M}, and I_{nP}^{M} in (10) by showing that $\overline{q}$, $\overline{q(\lambda-q)}$, $\overline{q(\lambda-\chi q)}$, and $\overline{q\lambda}$ in (A.1) are independent from M asymptotically in two special cases.

In the first special case, $\Lambda = \frac{M}{N}\mathbf{I}_N$ and $\frac{M}{N} = \beta$ with a fixed β. This in fact is the channel correlation model considered in [23]. Considering that both $\lambda_i = \beta$ and $q_i = \frac{\beta^2}{\overline{L}_P\beta + \frac{1}{\gamma_{\mathrm{tr}}}} \triangleq q$ are constants, after regular manipulations, we can derive that $\overline{q} = \frac{q}{\beta}$, $\overline{q(\lambda-q)} = \frac{q(\beta-q)}{\beta}$, $\overline{q(\lambda-\chi q)} = \frac{q(\beta-\chi q)}{\beta}$, and $\overline{q\lambda} = q$ are all constants and independent from M.

In the second special case, the systems with ULA and one-ring channel model are considered, where the mth row and nth column of the channel covariance matrix for users with AoA θ and AS Δ is $[\mathbf{R}]_{m,n} = \frac{1}{2\Delta}\int_{\theta-\Delta}^{\theta+\Delta} e^{-j2\pi D(m-n)\sin(x)}dx$ [29] and D is the normalized antenna spacing over wavelength. Noting that $\mathbf{R}$ is in Toeplitz form, according to *Fact 1* in [29], for any continuous function $f(x)$ defined over $[x_1, x_2]$, we have

$$\lim_{M\to\infty}\frac{1}{M}\sum_{m=1}^{M} f(\lambda_m(\mathbf{R})) = \int_{-\frac{1}{2}}^{\frac{1}{2}} f(S(\xi))d\xi, \tag{A.2}$$

where $\lambda_m(\mathbf{R})$ is the m-th eigenvalue of $\mathbf{R}$, and $S(\xi) \in [x_1, x_2]$ is a function depending on the AoA θ, AS Δ, and the normalized antenna space D. For $-\frac{\pi}{2} \le \theta - \Delta < \theta + \Delta \le \frac{\pi}{2}$, $S(\xi) = \frac{1}{2\Delta}\sum_{m\in[D\sin(-\Delta+\theta)+\xi, D\sin(\Delta+\theta)+\xi]}\frac{1}{\sqrt{D^2-(m-\xi)^2}}$.

From the MMSE channel estimation, we have $q_i = \frac{\lambda_i^2}{\overline{L}_P\lambda_i + \frac{1}{\gamma_{\mathrm{tr}}}}$. By defining function $f_q(x) = \frac{x^2}{\overline{L}_P x + \frac{1}{\gamma_{\mathrm{tr}}}}$, $q_i = f_q(\lambda_i)$ holds. According to (A.2), we know that

$$\begin{aligned}\lim_{M\to\infty}\overline{q} &= \lim_{M\to\infty}\frac{1}{M}\sum_{m=1}^{N} f_q(\lambda_i)\\ &= \lim_{M\to\infty}\frac{1}{M}\sum_{m=1}^{M} f_q(\lambda_i)\\ &= \int_{-\frac{1}{2}}^{\frac{1}{2}} f_q(S(\xi))d\xi = \int_{-\frac{1}{2}}^{\frac{1}{2}}\frac{S(\xi)^2}{\overline{L}_P S(\xi) + \frac{1}{\gamma_{\mathrm{tr}}}}d\xi,\end{aligned} \tag{A.3}$$

which is independent from M. Similarly, by defining functions $f_{q,1}(x) = f_q(x)(x - f_q(x))$, $f_{q,2}(x) = f_q(x)(x - \chi f_q(x))$ and $f_{q,3}(x) = x f_q(x)$, we can obtain that $\lim_{M\to\infty}\overline{q(\lambda-q)} = \int_{-\frac{1}{2}}^{\frac{1}{2}} f_{q,1}(S(\xi))d\xi$, $\lim_{M\to\infty}\overline{q(\lambda-\chi q)} = \int_{-\frac{1}{2}}^{\frac{1}{2}} f_{q,2}(S(\xi))d\xi$ and $\lim_{M\to\infty}\overline{q\lambda} = \int_{-\frac{1}{2}}^{\frac{1}{2}} f_{q,3}(S(\xi))d\xi$, which are also independent from M.

By using the same approaches as in [23] and [24], for systems with ZFBF and any covariance matrix $\mathbf{R}$, the receive powers can be derived as

$$S^{\mathrm{Z}} = M - \frac{M}{N}K + (1-\frac{K}{N})\frac{\delta' - \mu}{\delta^2}, \tag{A.4a}$$

$$I_P^{\mathrm{Z}} = (\overline{L}_P - 1)\left((M - \frac{M}{N}K)\chi + (1-\frac{K}{N})\frac{\delta' - \chi\mu}{\delta^2}\right), \tag{A.4b}$$

$$I_{nP}^{\mathrm{Z}} = (1-\frac{K}{N})\left((K\overline{L} - \overline{L}_P)\frac{\delta'}{\delta^2} - (1+\chi(\overline{L}_P - 1))(K-1)\frac{\delta''}{\delta^2}\right), \tag{A.4c}$$

where $\delta' = \frac{\frac{\delta^2}{K}\sum_{i=1}^{N}\frac{\lambda_i}{q_i}(\frac{K}{M}+\frac{\delta}{\alpha q_i})^{-2}}{\frac{M}{K} - \frac{1}{M}\sum_{i=1}^{N}(\frac{K}{M}+\frac{\delta}{\alpha q_i})^{-2}}$, $\delta'' = \frac{\delta^2\frac{1}{K}\sum_{i=1}^{N}\left(\frac{K}{M}+\frac{\delta}{\alpha q_i}\right)^{-2}}{\frac{M}{K} - \frac{1}{M}\sum_{i=1}^{N}\left(\frac{K}{M}+\frac{\delta}{\alpha q_i}\right)^{-2}}$, $\mathbf{T}' = \left(\alpha\Lambda + \frac{K}{M}\frac{\alpha\mathbf{Q}\delta'}{\delta^2}\right)\left(\mathbf{I}_N + \frac{K}{M}\frac{\alpha\mathbf{Q}}{\delta}\right)^{-2}$, $\mu = \frac{1}{M}\mathrm{tr}(\alpha^2\mathbf{Q}\mathbf{V}\Lambda\mathbf{T}')$, and δ can be found from the equation $\frac{1}{M}\sum_{i=1}^{N}\frac{\alpha q_i}{\frac{K}{M}\alpha q_i + \delta} = 1$.

We next prove the asymptotic behaviors of S^{Z}, I_P^{Z}, and I_{nP}^{Z} with M in (10) by proving that δ, δ', δ'' and μ in (A.4) are independent from M asymptotically in the two special cases.

In the first special case, when $M \gg K$ in massive MIMO systems, we can derive that $\delta = \frac{\alpha q}{\beta}$, $\delta' = \alpha^2 q$, and $\delta'' = \mu = \frac{\alpha^2 q^2}{\beta}$ are all independent from M asymptotically.

In the second special case, in order to analyze the properties of δ', δ'', and μ, we first derive the approximate expression of δ from the equation $\frac{1}{M}\sum_{i=1}^{N}\frac{\alpha q_i}{\frac{K}{M}\alpha q_i + \delta} = 1$, which can be rewritten as $\frac{1}{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\frac{K}{M}\alpha q_i + \delta}} = \frac{N\delta}{N-K}$, i.e., the harmonic mean of the variables $\frac{K}{M}\alpha q_i + \delta$. According to the relationship between the harmonic mean and arithmetic mean, we have $\frac{N\delta}{N-K} = \frac{1}{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\frac{K}{M}\alpha q_i + \delta}} \le \frac{1}{N}\sum_{i=1}^{N}(\frac{K}{M}\alpha q_i + \delta)$, where $\bar{q}$ is defined in (9) and the inequality holds with equality when the variables $\frac{K}{M}\alpha q_i + \delta$ are identical for $i = 1, \cdots, N$. Since $\lim_{M\to\infty}\frac{K}{M}\alpha q_i + \delta = \delta$, the inequality holds with equality, and hence we have $\delta = \frac{N-K}{N}\alpha\bar{q}$, which approaches to $\alpha\bar{q}$ for large M considering the fact that N (i.e., the spatial degrees of freedom of a massive MIMO system) increases with M typically. Recalling that $\overline{q}$ is independent from M, we can find that δ is also independent from M asymptotically.

Substituting the asymptotic expression of $\delta = \alpha\bar{q}$ into (A.4), we can derive $\frac{\delta'}{\delta^2}$ as

$$\begin{aligned}\lim_{M\to\infty}\frac{\delta'}{\delta^2} &= \lim_{M\to\infty}\frac{\frac{1}{K}\sum_{i=1}^{N}\frac{\lambda_i q_i}{(\frac{K}{M}q_i + \overline{q})^2}}{\frac{M}{K} - \frac{1}{M}\sum_{i=1}^{N} q_i^2(\frac{K}{M}q_i + \overline{q})^{-2}}\\ &\overset{(a)}{=} \lim_{M\to\infty}\frac{\frac{M}{K}\overline{q\lambda}(\overline{q})^{-2}}{\frac{M}{K} - \overline{q^2}(\overline{q})^{-2}} = \frac{\overline{q\lambda}}{(\overline{q})^2},\end{aligned} \tag{A.5}$$

where (a) comes from $\lim_{M\to\infty}\frac{K}{M}q_i + \overline{q} = \overline{q}$ and $\lim_{M\to\infty}\overline{q^2} = \int_{-\frac{1}{2}}^{\frac{1}{2}}\left(f_q(S(\xi))\right)^2 d\xi$. Then, we have $\lim_{M\to\infty}\delta' = \alpha^2\overline{q\lambda}$, which is independent from M. By employing the same approach as deriving (A.5), we can derive that $\lim_{M\to\infty}\delta'' = \lim_{M\to\infty}\mu = \alpha^2\overline{q^2}$, which is also independent from M.

$$\begin{aligned}
&\frac{SG(P^*+\frac{MP_0}{(1-\frac{KT_{\mathrm{tr}}}{T})\eta})}{((S+I)P^*+G)(IP^*+G)}-\ln\left(1+\frac{SP^*}{IP^*+G}\right)\\
&=\frac{\frac{MP_0}{(1-\frac{KT_{\mathrm{tr}}}{T})\eta}-\frac{G}{S+I}}{P^*+\frac{G}{S+I}}-\frac{\frac{MP_0}{(1-\frac{KT_{\mathrm{tr}}}{T})\eta}-\frac{G}{I}}{P^*+\frac{G}{I}}-\ln\left(1+\frac{G}{(S+I)P^*}\right)+\ln\left(1+\frac{G}{IP^*}\right)-\ln\left(1+\frac{S}{I}\right)\\
&\overset{(a)}{\approx}\frac{\frac{MP_0}{(1-\frac{KT_{\mathrm{tr}}}{T})\eta}-\frac{G}{S+I}}{P^*+\frac{G}{S+I}}-\frac{\frac{MP_0}{(1-\frac{KT_{\mathrm{tr}}}{T})\eta}-\frac{G}{I}}{P^*+\frac{G}{I}}-\frac{\frac{G}{(S+I)}}{P^*+\frac{G}{2(S+I)}}+\frac{\frac{G}{I}}{P^*+\frac{G}{2I}}-\ln\left(1+\frac{S}{I}\right)\\
&\overset{(b)}{\approx}\frac{\frac{MP_0}{(1-\frac{KT_{\mathrm{tr}}}{T})\eta}-\frac{2G}{S+I}}{P^*+\frac{G}{S+I}}-\frac{\frac{MP_0}{(1-\frac{KT_{\mathrm{tr}}}{T})\eta}-\frac{2G}{I}}{P^*+\frac{G}{I}}-\ln\left(1+\frac{S}{I}\right)\overset{(c)}{\approx}\frac{\frac{MP_0}{(1-\frac{KT_{\mathrm{tr}}}{T})\eta}}{P^*+\frac{G}{S+I}}-\frac{\frac{MP_0}{(1-\frac{KT_{\mathrm{tr}}}{T})\eta}}{P^*+\frac{G}{I}}-\ln\left(1+\frac{S}{I}\right)=0, \qquad \text{(B.1)}
\end{aligned}$$

From the analysis above, we can conclude (10) for the systems using MRT and ZFBF in the two special cases.

Appendix B
Approximated Relaxed-EE-Maximal Transmit Power P^*

The condition in (12) can be expressed as (B.1), shown at the top of this page, where (a) employs the approximation of $\ln(1+x)=2\mathrm{atanh}(\frac{x}{2+x})\approx\frac{2x}{2+x}$ according to the first order taylor expansion of $\mathrm{atanh}(x)\approx x$. We can obtain that when $x<2$, the relative approximation error $\frac{|\mathrm{atanh}(\frac{x}{2+x})-\frac{x}{2+x}|}{\mathrm{atanh}(\frac{x}{x+2})}$ is less than 0.09 and reduces with the decrease of x. We know $\lim_{M\to\infty}\frac{S}{M}=1$ from (10), and P^* increases with M monotonically from (12). Therefore, for large M such that the corresponding S and P^* make $\frac{G}{(S+I)P^*}<\frac{G}{IP^*}\ll 2$ hold, approximation (a) will become very accurate. The condition for approximation (a) being accurate is equivalent to $P^*\gg\frac{G}{2I}>\frac{G}{2(S+I)}$, which also makes approximation (b) accurate, where $P^*+\frac{G}{2(S+I)}\approx P^*+\frac{G}{(S+I)}$ and $P^*+\frac{G}{2I}\approx P^*+\frac{G}{I}$ are employed in (b). We approximate $\frac{MP_0}{(1-\frac{KT_{\mathrm{tr}}}{T})\eta}-\frac{2G}{I}\approx\frac{MP_0}{(1-\frac{KT_{\mathrm{tr}}}{T})\eta}$ and $\frac{MP_0}{(1-\frac{KT_{\mathrm{tr}}}{T})\eta}-\frac{2G}{S+I}\approx\frac{MP_0}{(1-\frac{KT_{\mathrm{tr}}}{T})\eta}$ in (c), which are accurate for $\frac{MP_0}{(1-\frac{KT_{\mathrm{tr}}}{T})\eta}\gg\frac{2G}{I}>\frac{2G}{S+I}$. Because the first term of the inequality increases with M while $\frac{2G}{I}$ is constant or decreases with M according to (10), we know that approximation (c) in (B.1) is accurate for large M. Furthermore, when P_0 is large, $\frac{MP_0}{(1-\frac{KT_{\mathrm{tr}}}{T})\eta}$ is large and P^* is also large according to (12). Then the approximations (b) and (c) in (B.1) are more accurate for large P_0.

The approximate solution of (12) (i.e., the solution of (B.1)) is

$$\begin{aligned}
P^*&\approx\left(\frac{G}{I}-\frac{G}{S+I}\right)\sqrt{\frac{\frac{MP_0}{(1-\frac{KT_{\mathrm{tr}}}{T})\eta}}{(\frac{G}{I}-\frac{G}{S+I})\ln(1+\frac{S}{I})}+\frac{1}{4}}-\frac{1}{2}\left(\frac{G}{I}+\frac{G}{S+I}\right)\\
&\overset{(d)}{\approx}\sqrt{\frac{\frac{MP_0}{(1-\frac{KT_{\mathrm{tr}}}{T})\eta}(\frac{G}{I}-\frac{G}{S+I})}{\ln(1+\frac{S}{I})}}-\frac{1}{2}\left(\frac{G}{I}+\frac{G}{S+I}\right)\\
&\overset{(e)}{\approx}\sqrt{\frac{MP_0K\sigma^2}{(1-\frac{KT_{\mathrm{tr}}}{T})\eta\alpha\overline{q}}}\sqrt{\frac{\frac{1}{I}-\frac{1}{S+I}}{\ln(1+\frac{S}{I})}}, \qquad \text{(B.2)}
\end{aligned}$$

where $\frac{1}{4}$ is ignored in (d) and $\frac{1}{2}\left(\frac{G}{I}+\frac{G}{S+I}\right)$ is ignored in (e). The accuracy of the approximations is discussed as follows.

The condition for (e) being accurate is $\sqrt{\frac{\frac{MP_0}{(1-\frac{KT_{\mathrm{tr}}}{T})\eta}(\frac{G}{I}-\frac{G}{S+I})}{\ln(1+\frac{S}{I})}}\gg\frac{1}{2}\left(\frac{G}{I}+\frac{G}{S+I}\right)$, which is equivalent to

$$\frac{MP_0}{(1-\frac{KT_{\mathrm{tr}}}{T})\eta}\gg(\frac{G}{4I}-\frac{G}{4(S+I)}+\frac{G}{S})\ln(1+\frac{S}{I}). \qquad \text{(B.3)}$$

We can find that because $\frac{G}{I}+\frac{G}{S+I}>\frac{G}{I}-\frac{G}{S+I}$, when the approximate condition of (e) holds, $\frac{\frac{MP_0}{(1-\frac{KT_{\mathrm{tr}}}{T})\eta}}{(\frac{G}{I}-\frac{G}{S+I})\ln(1+\frac{S}{I})}\gg\frac{1}{4}$ also holds, i.e., approximation (d) is also accurate. From (10), we have $I\sim\mathcal{O}(M)$ and $I\sim\mathcal{O}(1)$ for the system with and without PC, respectively. Therefore, the right-hand-side of (B.3) decreases with $\frac{1}{M}$ for the system with PC while increases with $\ln M$ for the system without PC. However, because of the left-hand-side of (B.3) increases linearly with M, when M is large, the inequality holds for both systems with and without PC, i.e., approximations (d) and (e) are accurate. We can observe that when P_0 is large, the left-hand-side of (B.3) is large and hence approximations (d) and (e) are more accurate.

In summary, the approximate relaxed-EE-maximal transmit power in (B.2) is accurate when M is large and becomes more accurate for large P_0.

Appendix C
Proof of Proposition 1

For the system without PC, $L_P=1$ and $I_P=0$. From (10), we know that $\lim_{M\to\infty}\frac{S}{M}=1$ and hence $\lim_{M\to\infty}(\frac{1}{I_{nP}}-\frac{1}{S+I_{nP}})=\frac{1}{I_{nP}}$ and $\lim_{M\to\infty}\frac{\ln(1+\frac{S}{I_{nP}})}{\ln(M)}=1$. Considering (13), we can derive that

$$\lim_{M\to\infty}\frac{P^*}{\sqrt{\frac{M}{\ln M}}}=\sqrt{\frac{P_0K\sigma^2}{(1-\frac{KT_{\mathrm{tr}}}{T})\eta\alpha\overline{q}I_{nP}}}. \qquad \text{(C.1)}$$

For the system with PC, from (10), we have $\lim_{M\to\infty}\frac{I}{\chi(\overline{L}_P-1)M}=1$. In this case, $\lim_{M\to\infty}(\frac{M}{I}-\frac{M}{S+I})=$

$\frac{1}{\chi(\overline{L}_P-1)} - \frac{1}{1+\chi(\overline{L}_P-1)}$ and $\lim_{M\to\infty}\ln(1+\frac{S}{I}) = \ln(1+\frac{1}{\chi(\overline{L}_P-1)})$. Upon substituting the results into (13), we can derive that

$$\lim_{M\to\infty} P^* = \sqrt{\frac{P_0 K\sigma^2}{(1-\frac{KT_{\rm tr}}{T})\eta\alpha\overline{q}} \frac{\frac{1}{\chi(\overline{L}_P-1)} - \frac{1}{1+\chi(\overline{L}_P-1)}}{\ln(1+\frac{1}{\chi(\overline{L}_P-1)})}}. \quad \text{(C.2)}$$

Combining (C.1) and (C.2), we can obtain the scaling law of P^* with respect to M in (15).

Appendix D
Proof of Proposition 2

For the system without PC, $I_P = 0$ and $I = I_{nP}$. From (10), we can obtain $\frac{M}{I_{nP}} \gg \frac{M}{S+I_{nP}}$. Considering the expression of $f_{\rm EE}(M)$ in (14), we can derive $\lim_{M\to\infty}\frac{f_{\rm EE}(M)}{\sqrt{\frac{M}{I_{nP}\ln(M)}}} = 1$, i.e., $f_{\rm EE}(M)$ in (14) scales as $f_{\rm EE}(M) \sim O\left(\sqrt{\frac{M}{I_{nP}\ln(M)}}\right)$. Therefore, $\lim_{M\to\infty}\log_2\left(1+\frac{S}{I_{nP}+\frac{c}{f_{\rm EE}(M)}}\right)/\log_2(M) = 1$ and $\lim_{M\to\infty}\frac{M+c\cdot f_{\rm EE}(M)}{M} = 1$. According to (14), we can derive that

$$\lim_{M\to\infty}\frac{{\rm EE}^*\cdot M}{\log_2(M)} = \frac{(1-\frac{KT_{\rm tr}}{T})BK}{P_0}\lim_{M\to\infty}\frac{\log_2\left(1+\frac{S}{I+\frac{c}{f_{\rm EE}(M)}}\right)}{M+cf_{\rm EE}(M)}\frac{M}{\log_2(M)} = \frac{(1-\frac{KT_{\rm tr}}{T})BK}{P_0}. \quad \text{(D.1)}$$

For the system with PC, based on $\lim_{M\to\infty}\frac{I}{\chi(\overline{L}_P-1)M} = 1$ given in (10), we know that $\lim_{M\to\infty}\frac{S}{I} = \frac{1}{\chi(\overline{L}_P-1)}$ and $\lim_{M\to\infty}\frac{M}{S+I} = \frac{1}{1+\chi(\overline{L}_P-1)}$ hold. Upon substituting the results into $f_{\rm EE}(M)$, we can derive that $\lim_{M\to\infty} f_{\rm EE}(M) = \sqrt{\frac{\frac{1}{\chi(\overline{L}_P-1)}-\frac{1}{1+\chi(\overline{L}_P-1)}}{\ln(1+\frac{1}{\chi(\overline{L}_P-1)})}}$, which is independent of M. Therefore, $\lim_{M\to\infty}\log_2(1+\frac{S}{I+\frac{c}{f_{\rm EE}(M)}}) = \lim_{M\to\infty}\log_2(1+\frac{S}{I}) = \log_2(1+\frac{1}{\chi(\overline{L}_P-1)})$. Then, we can derive that

$$\lim_{M\to\infty}{\rm EE}^*\cdot M = \frac{(1-\frac{KT_{\rm tr}}{T})BK}{P_0}\lim_{M\to\infty}\frac{\log_2(1+\frac{S}{I+\frac{c}{f_{\rm EE}(M)}})}{M+cf_{\rm EE}(M)}\cdot M = \frac{(1-\frac{KT_{\rm tr}}{T})BK\log_2(1+\frac{1}{\chi(\overline{L}_P-1)})}{P_0}. \quad \text{(D.2)}$$

Combining (D.1) and (D.2), the scaling laws in (16) are obtained.

Appendix E
Proof of Proposition 3

For the system without PC, according to (10) and the scaling law of $\tilde{P}^*$ in Table I, $I = I_{nP}$ is a constant and $\tilde{P}^*$ firstly increases with M and then becomes the constant $P_{\max}$. In this case, when M increases, $\frac{1}{\tilde{P}^*} - \frac{1}{P_{\max}}$ reduces to zero. Considering (21), with the increase of M, the rate gap $\Delta_R = BK\log_2\left(1+\frac{\frac{1}{\tilde{P}^*}-\frac{1}{P_{\max}}}{\frac{I_{nP}}{G}+\frac{1}{P_{\max}}}\right)$ reduces accordingly and finally reaches zero.

For the system with PC, according to (10) and Table I, when M increases, $I = I_P + I_{nP}$ increases monotonically and $\tilde{P}^*$ firstly increases and then converges to a constant. In this case, with the increase of M, the rate gap $\Delta_R = BK\log_2\left(1+\frac{\frac{1}{\tilde{P}^*}-\frac{1}{P_{\max}}}{\frac{I_P+I_{nP}}{G}+\frac{1}{P_{\max}}}\right)$ converges to zero as a result of $\frac{I}{G}+\frac{1}{P_{\max}}$ increasing with M and $\frac{1}{\tilde{P}^*}-\frac{1}{P_{\max}}$ approaching to a constant.

Therefore, for the massive MIMO system with or without PC, the EE-maximal average rate $\tilde{R}_j^*$ is close to the maximal average rate $R_{j,\max}$, and the gap between them reduces to zero with the increase of M.

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] F. Rusek *et al.*, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[3] G. Y. Li *et al.*, "Energy-efficient wireless communications: Tutorial, survey, and open issues," *IEEE Wireless Commun.*, vol. 18, no. 6, pp. 28–35, Dec. 2011.

[4] X. Zhang *et al.*, "Energy-efficient multimedia transmissions through base station cooperation over heterogeneous cellular networks exploiting user behavior," *IEEE Wireless Commun.*, vol. 21, no. 4, pp. 54–61, Aug. 2014.

[5] F. D. Cardoso and L. M. Correia, "MIMO gain and energy efficiency in LTE," in *Proc. IEEE WCNC*, Apr. 2012, pp. 2593–2597.

[6] Z. Xu, C. Yang, G. Y. Li, S. Zhang, and S. Xu, "Energy-efficient configuration of spatial and frequency resources in MIMO-OFDMA systems," *IEEE Trans. Commun.*, vol. 61, no. 2, pp. 564–575, Feb. 2013.

[7] 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.

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

[9] S. K. Mohammed, "Impact of transceiver power consumption on the energy efficiency of zero-forcing detector in massive MIMO systems," *IEEE Trans. Commun.*, vol. 62, no. 11, pp. 3874–3890, Nov. 2014.

[10] D. Ha, K. Lee, and J. Kang, "Energy efficiency analysis with circuit power consumption in massive MIMO systems," in *Proc. IEEE PIMRC*, Sep. 2013, pp. 938–942.

[11] Y. Xin, D. Wang, J. Li, H. Zhu, J. Wang, and X. You, "Area spectral efficiency and area energy efficiency of massive MIMO cellular systems," *IEEE Trans. Veh. Technol.*, vol. 65, no. 5, pp. 3243–3254, May 2016.

[12] 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*, Jun. 2015, pp. 1776–1782.

[13] H. Yang and T. L. Marzetta, "Energy efficient design of massive MIMO: How many antennas?," in *Proc. IEEE VTC*, May 2015, pp. 1–5.

[14] 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.

[15] C. Kong, C. Zhong, A. K. 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.

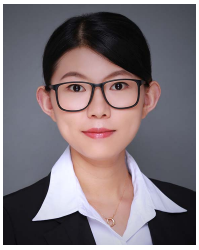
[16] E. Björnson, L. Sanguinetti, J. Hoydis, and M. Debbah, "Optimal design of energy-efficient multi-user MIMO systems: Is massive MIMO the answer?," *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3059–3075, Jun. 2015.

[17] Z. Chen and C. Yang, "Pilot decontamination in wideband massive MIMO systems by exploiting channel sparsity," *IEEE Trans. Wireless Commun.*, vol. 15, no. 7, pp. 5087–5100, Jul. 2016.

[18] G. Auer *et al.*, "How much energy is needed to run a wireless network?," *IEEE Trans. Wireless Commun.*, vol. 18, no. 5, pp. 40–49, Oct. 2011.

[19] C. Desset, B. Debaillie, and F. Louagie, "Modeling the hardware power consumption of large scale antenna systems," in *Proc. IEEE GreenComm*, Nov. 2014, pp. 1–6.

[20] F. D. Cardoso *et al.*, "Energy efficient transmission techniques for LTE," *IEEE Commun. Mag.*, vol. 51, no. 10, pp. 182–190, Oct. 2013.
[21] H. Yang and T. L. Marzetta, "Total energy efficiency of cellular large scale antenna system multiple access mobile networks," in *Proc. IEEE Online Conf. Green Commun.*, Oct. 2013, pp. 27–32.
[22] R. Couillet and M. Debbah, *Random Matrix Methods for Wireless Communications*, 1st ed. New York, NY, USA: Cambridge Univ. Press, 2011.
[23] 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.
[24] S. Wagner, R. Couillet, M. Debbah, and D. T. M. Slock, "Large system analysis of linear precoding in correlated MISO broadcast channels under limited feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4509–4537, Jul. 2012.
[25] R. M. Corless, 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. 1, pp. 329–359, 1996.
[26] Z. Xu, S. Han, Z. Pan, and C.-L. Lin, "EE-SE relationship for large-scale antenna systems," in *Proc. IEEE ICC*, Jun. 2014, pp. 38–42.
[27] 3GPP, "Further Advancements for E-UTRA Physical Layer Aspects (Release 1169 9)", document 3GPP TR 36.814, 3rd Generation Partnership Project, 2010.
[28] 3GPP, Technical Specification Group, "Radio Access Network; Study on 3D Channel Model for LTE (Release 12)", document 3GPP TR 36.873, 3rd Generation Partnership Project, 2013.
[29] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing—The large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.



Wenjia Liu received the B.S. degree from the School of Electronics and Information Engineering, Beihang University (BUAA), Beijing, China, in 2011. She is currently pursuing the Ph.D. degree with the School of Electronics and Information Engineering, BUAA. From January to July 2015, she was a Visiting Student with the Communications Laboratory, Technische Universität Dresden, Dresden, Germany. Her research interests lie in the area of energy efficient, massive MIMO, and small cell network.

Shengqian Han (S'05–M'12) received the B.S. degree in communication engineering and the Ph.D. degree in signal and information processing from Beihang University, Beijing, China, in 2004 and 2010, respectively. From 2010 to 2012, he held a post-doctoral research position with the School of Electronics and Information Engineering, Beihang University. Since 2012, he has been a Lecturer with the School of Electronics and Information Engineering, Beihang University. From 2015 to 2016, he was a Visiting Scholar with the Department of Electrical Engineering, The University of Southern California, Los Angeles, USA. His research interests are cooperative communication, full-duplex communications, and energy efficient transmission in the areas of wireless communications and signal processing.

Chenyang Yang (SM'08) received the Ph.D. degree in electrical engineering from Beihang University, Beijing, China, in 1997. Since 1999, she has been a Full Professor with the School of Electronic and Information Engineering, Beihang University. She has filed over 60 patents. She has authored over 200 international journal and conference papers in the fields of energy efficient transmission, coordinated multipoint, interference management, and so on. Her recent research interests include green radio, local caching, and other emerging techniques for next generation wireless networks.

Prof. Yang was the Chair of the Beijing chapter of the IEEE Communications Society from 2008 to 2012. She has served as a Technical Program Committee Member for numerous IEEE conferences. She served as an Associate Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and a Guest Editor for the IEEE JOURNAL ON SELECTED TOPICS IN SIGNAL PROCESSING and IEEE JOURNAL OF SELECTED AERAS OF COMMUNICATIONS. She is currently an Associate Editor-in-Chief of the *Chinese Journal of Communications* and the *Chinese Journal of Signal Processing*. She was supported by the First Teaching and Research Award Program for Outstanding Young Teachers of Higher Education Institutions by the Ministry of Education.