IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 25, 2018, accepted December 12, 2018, date of publication December 17, 2018,
date of current version January 11, 2019.

Digital Object Identifier 10.1109/ACCESS.2018.2887085

Energy Efficiency Optimization of Distributed Massive MIMO Systems Under Ergodic QoS and Per-RAU Power Constraints

JIAMIN LI, DONGMING WANG, PENGCHENG ZHU, (Member, IEEE), AND XIAOHU YOU, (Fellow, IEEE)

National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China

Corresponding author: Pengcheng Zhu (p.zhu@seu.edu.cn)

This work was supported in part by National Natural Science Foundation of China (NSFC) under Grant 61501113, Grant 61571120, and Grant 61871122, in part by the Jiangsu Provincial Natural Science Foundation under Grant BK20150630 and Grant BK20180011, and in part by the National Key Special Program under Grant 2018ZX03001008-002.

ABSTRACT Under *ergodic* per-user quality-of-service and per-remote antenna unit (RAU) transmit power constraints, we investigate the problem of maximizing energy efficiency (EE) of distributed massive MIMO systems, which is known to be non-convex. To solve this challenging problem efficiently, we first derive closed-form expressions for the spectral efficiency and the power control parameters (related to per-RAU transmit power constraint) with zero-forcing (ZF) and maximum ratio transmission (MRT) beamforming, and then develop a computationally feasible power allocation algorithm using the tools of fractional programming and sequential convex approximation. The derived closed-form expressions are functions of only slowly changing large-scale fading which enables us to solve the optimization problem over a longer time interval. The proposed power allocation algorithm is guaranteed to converge to the Karush–Kuhn–Tucker points of the original non-convex EE maximization problem. The simulation results demonstrate the accuracy of the derived expressions and the effectiveness of the proposed algorithm. Moreover, some insightful conclusions are arrived at from the EE comparisons between different beamforming schemes (ZF and MRT) and different antenna deployments (distributed and co-located).

INDEX TERMS Energy efficiency, distributed massive MIMO systems, power allocation, fractional programming, sequential convex approximation.

I. INTRODUCTION

With the exponential growth of energy consumption in wireless networks, energy efficiency (EE) receives more and more attention due to economical and societal reasons [1]. Distributed massive multiple-input multiple-output (MIMO), also known as cell-free massive MIMO, large-scale distributed antenna systems (DAS) or massive DAS, combines the advantages of massive MIMO and DAS by dispersing the massive antenna array geographically, thus has great potential to achieve higher EE and spectral efficiency (SE) [2]–[8].

In distributed massive MIMO systems, antennas are dispersed at different remote antenna units (RAUs). The channels have independent yet non-identically distributed entries since the distances from each user to RAUs are different, and thus are typically modeled as composite channels including both small-scale and large-scale fading. It is the key difference compared to co-located antenna systems and brings huge challenges to analyze the system performance and optimize the resource allocation [9]–[13]. Most of the previous works optimized the resource allocation of distributed massive MIMO systems under *instantaneous* (i.e., with respect to small-scale Rayleigh fading) quality of service (QoS) and transmit power constraints [14]–[17]. Since the small-scale fading coefficients change quickly, it is extremely computationally expensive. Considering that users care more about average QoS and small-scale fading has negligible impact on massive MIMO systems due to the effect of channel hardening [18], we can optimize the resource allocation based on only the slowly changing large-scale fading (i.e., under *ergodic* constraints), which is more computationally feasible. With linear beamforming (e.g., MRT and ZF), the closed-form expressions for SE can be derived which are functions of only large-scale fading coefficients. Based on these analytical expressions, resource allocation optimization problems with

the objectives of SE or EE maximization were investigated and computationally efficient resource allocation algorithms were proposed in [19]–[23]. The drawbacks of the studies by [19] and [20] are that no QoS constraint and practical transmit power constraint are considered. Reference [21] investigated the EE maximization problem subject to per-user SE and per-RAU power constraint. However, only non-coherent transmission with conjugate beamforming at single-antenna RAUs are considered where there is no coordination between RAUs and thus the power control parameters related to per-RAU power constraint can be calculated straightforwardly. When coherent transmission with MRT and ZF beamforming (i.e., the beamforming vectors are computed through multiple RAUs) are considered, the power control parameters are related to the norms of the beamforming coefficients at each RAUs. Thus, it is more challenging to obtain ergodic per-RAU power constraints. Reference [22], [23] considered ergodic per-RAU power constraint with ZF beamforming, however, no closed-form expressions for the power control parameters are provided. To the best of our knowledge, under *ergodic* QoS and per-RAU power constraints given in closed-form, no works have investigated the problem of maximizing the EE of distributed massive MIMO systems with ZF and MRT beamforming.

Moreover, it is important to evaluate the EE based on an appropriate power consumption model. Compared with MRT beamforming, ZF beamforming can provide a higher SE [10]. However, it has a higher complexity and will consume more energy [24]. Meanwhile, different from co-located massive MIMO system, more backhaul links are required to connect RAUs to their baseband processing units (BPU) in distributed massive MIMO systems. Thus, although a higher SE can be achieved in distributed massive MIMO systems, the backhaul energy consumption might overwhelm the SE gains and result in a lower EE. Thus, the power consumption of different beamforming schemes and different antenna deployments (co-located or geographically distributed) should be embodied in the adopted power consumption model. Based on the previous works [19], [21], [24], we model the power consumption as the sum of transmit power, the power consumption of transceiver chains, linear processing, and load-dependent backhaul links, which enable the accurate EE comparison between different beamforming schemes and different antenna deployments.

Motivated by the aspects mentioned above, under *ergodic* per-user QoS and per-RAU power constraints, we investigate the problem of maximizing the EE of distributed massive MIMO systems based on a practical power consumption model with ZF and MRT beamforming. Although cell-free scenarios are considered in this paper, as illustrated in Remark 3, our results can be straightforwardly generalized to multi-cell pilot contamination scenarios. Specially, the main contributions of this paper are:

- The EE of distributed massive MIMO systems is formulated based on an appropriate power consumption model, which enables the accurate EE comparison between different beamforming schemes and different antenna deployments.
- The SE and the power control parameters are given in closed-form with ZF and MRT beamforming, which are functions of only large-scale fading coefficients and thus enables us to solve the EE maximization problem over a longer time interval.
- To tackle the non-convexity of the considered EE maximization problem, we first present feasibility conditions in closed-form, and then develop a computationally feasible power allocation algorithm using the tools of fractional programming and sequential convex approximation (SCA), which is guaranteed to converge to Karush-Kuhn-Tucker (KKT) points of the optimization problem.
- We corroborate the analytical analysis by simulation results and arrive at some insightful conclusions from the analysis and comparison of EE.

The remainder of the paper is organized as follows. In Section II, we describe the distributed massive MIMO system model and formulate the EE maximization problem under ergodic per-user QoS and per-RAU transmit power constraints. Section III studies the EE maximization problem where we derive accurate analytical expressions for the SE and power control parameters, provide the feasible conditions of the optimization problem and propose a computationally feasible power allocation algorithm. Numerical results and discussions are given in Section IV and the main results and insights are summarized in Section V.

Notation : Vectors (matrices) are denoted by lower (upper) boldface letters. $(\cdot)^{\mathrm{T}}$, $(\cdot)^{\mathrm{H}}$, $\mathrm{tr}\,(\cdot)$ and $\mathbb{E}\,[\cdot]$ stand for the transpose, Hermitian transpose, trace and expectation operators, respectively. Euclidean norm of a matrix is denoted by $\|\cdot\|$. $\mathcal{CN}(0, \sigma^2)$ represents a circularly symmetric complex Gaussian distribution with mean zero and variance σ^2. $\mathbf{I}_N$ denotes an $N \times N$ identity matrix. $\Gamma(\cdot, \cdot)$ and Nakagami$(\cdot, \cdot)$ denote Gamma distribution and Nakagami distribution with corresponding parameters, respectively.

II. SYSTEM MODEL AND PROBLEM FORMULATION

In this section, we first describe the system model and then formulate the considered EE maximization problem.

A. SYSTEM MODEL

Consider a distributed massive MIMO system which is designed from scratch to uniformly cover a given area with maximal EE. It consists of a BPU, M RAUs and K single-antenna users. Each RAU is equipped with N antennas. The RAUs are connected to BPU through backhaul links and jointly serve the users in the same time-frequency resource. All signal processing and resource management are implemented at BPU.

In this paper, we focus on the downlink. Similar to [19], we assume that users transmit mutually orthogonal pilot sequences with length τ_{p} in the phase of uplink channel estimation. Then, BPU acquires channel state information (CSI)

of users by using the minimum mean-square error (MMSE) estimator, which can be given by [25]

$$\hat{\mathbf{g}}_k = \left[\hat{\mathbf{g}}_{1,k}^{\mathrm{T}}, \cdots, \hat{\mathbf{g}}_{M,k}^{\mathrm{T}}\right]^{\mathrm{T}}, \tag{1}$$

where $\hat{\mathbf{g}}_{m,k} = \sqrt{\beta_{m,k}}\hat{\mathbf{h}}_{m,k}$, $\beta_{m,k} \triangleq \frac{\lambda_{m,k}^2}{\lambda_{m,k}+\sigma_k^2/(p_k^{\mathrm{u}}\tau_{\mathrm{p}})}$, $\hat{\mathbf{h}}_{m,k} \sim \mathcal{CN}(0, \mathbf{I}_N)$, $\lambda_{m,k}$ represents the large-scale fading including path loss and shadowing which changes slowly and is assumed to be known at the BPU, and p_k^{u} is the pilot transmit power of user k. Due to the orthogonality property of MMSE estimate, the channel estimation error $\tilde{\mathbf{g}}_k = \mathbf{g}_k - \hat{\mathbf{g}}_k \sim \mathcal{CN}\Big(0, \mathrm{diag}\big(\left[\eta_{1,k}, \cdots, \eta_{M,k}\right]^{\mathrm{T}}\big) \otimes \mathbf{I}_N\Big)$ is independent of $\hat{\mathbf{g}}_k$, where $\eta_{m,k} \triangleq \lambda_{m,k} - \beta_{m,k}$.

After acquiring the CSI from the uplink channel estimation, the M RAUs user linear beamforming to cooperatively transmit signals to users. The received signal of user k can be written as

$$y_k = \sum_{j=1}^{K} \sqrt{p_j}\mathbf{g}_k^{\mathrm{H}}\vec{\mathbf{w}}_j s_j + z_k, \tag{2}$$

where $\mathbf{g}_k = [\mathbf{g}_{1,k}, \cdots, \mathbf{g}_{M,k}]$ is the channel vector from user k to all of the RAUs, $\mathbf{g}_{m,k} = \sqrt{\lambda_{m,k}}\mathbf{h}_{m,k}$ is the channel vector from user k to RAU m, $\mathbf{h}_{m,k} \sim \mathcal{CN}(0, \mathbf{I}_N)$ is the Rayleigh small-scale fading, p_j is the transmitted power assigned for the j-th user, $\vec{\mathbf{w}}_j$ is the normalized beamforming direction, $s_j \sim \mathcal{CN}(0, 1)$ is the transmitted signal, and $z_k \sim \mathcal{CN}(0, \sigma_k^2)$ indicates the additive noise.

Beamforming direction $\vec{\mathbf{w}}_j$ is a function of the estimated CSI. In this paper, we focus on MRT and ZF beamforming schemes, which gives

$$\vec{\mathbf{w}}_j = \begin{cases} \dfrac{\hat{\mathbf{g}}_j}{\|\hat{\mathbf{g}}_j\|}, & \text{for MRT,} \\ \dfrac{\mathbf{v}_j}{\|\mathbf{v}_j\|}, & \text{for ZF.} \end{cases} \tag{3}$$

where $\mathbf{v}_j$ is the j-th column of $\hat{\mathbf{G}}\big(\hat{\mathbf{G}}^{\mathrm{H}}\hat{\mathbf{G}}\big)^{-1}$ and $\hat{\mathbf{G}} \triangleq \big[\hat{\mathbf{g}}_1, \cdots, \hat{\mathbf{g}}_K\big]$.

B. PROBLEM FORMULATION

The EE is defined as the ratio of the ergodic sum throughput (bit/s) to the total power consumption (Watt) [19], [22], i.e.,

$$\phi(\mathbf{p}) = \frac{B\sum_{k=1}^{K} S_k(\mathbf{p})}{P_{\mathrm{Total}}(\mathbf{p})}, \tag{4}$$

where B is the systems bandwidth, $\mathbf{p} = [p_1, \cdots, p_K]^{\mathrm{T}}$ is the optimization variable, $S_k(\mathbf{p})$ and $P_{\mathrm{Total}}(\mathbf{p})$ are the spectral efficiency of user k and the total power consumption, respectively, which are both functions of power allocation $\mathbf{p}$.

Based on the [19], [21], [24], the total power consumption is modeled as

$$P_{\mathrm{Total}}(\mathbf{p}) = P_{\mathrm{TC}} + P_{\mathrm{LP}} + P_{\mathrm{T}} + P_{\mathrm{BH}}, \tag{5}$$

where P_{TC} is the power consumed for transceiver chains, P_{LP} for linear processing at the BPU, P_{T} is the downlink transmit power and P_{BH} accounts for the backhaul-inducing power. The detail definitions of the terms in (5) are given as follows.

1) In distributed massive MIMO systems, we assume that different RAUs have different oscillators and all the antennas at each RAU are connected to a single oscillator. Thus,

$$P_{\mathrm{TC}} = MNP_{\mathrm{TR}} + MP_{\mathrm{SYN}} + KP_{\mathrm{UE}}, \tag{6}$$

where P_{TR} and P_{UE} denote the consumed power per antenna at the transmitter and users, respectively, and P_{SYN} denotes the power consumption of the local oscillator. If single oscillator is used, e.g., in co-located massive MIMO systems, we can easily set the coefficient of P_{SYN} to one.

2) As seen from (3), only the normalization of $\hat{\mathbf{g}}_k$ is needed if MRT is used. However, if ZF is userd, we need to calculate channel matrix inversions. From [24], the power consumption P_{LP} can be approximated as

$$P_{\mathrm{LP}} = \begin{cases} P_{\mathrm{M}} + \dfrac{B}{T}\dfrac{3MNK}{L_{\mathrm{TR}}}, & \text{for MRT} \quad (7a) \\ P_{\mathrm{M}} + \dfrac{B}{T}\left(\dfrac{K^3}{3L_{\mathrm{TR}}} + \dfrac{MNK(3K+1)}{L_{\mathrm{TR}}}\right), & \text{for ZF} \quad (7b) \end{cases}$$

where $P_{\mathrm{M}} = B\frac{T-\tau_{\mathrm{u}}}{T}\frac{2MNK}{L_{\mathrm{TR}}}$ is the power consumption of one matrix-vector multiplication per data symbol, B denotes the system bandwidth and L_{TR} is the computational efficiency at transmitter quantified in terms of the number of arithmetic complex-valued operations per joule.

3) The transmit power can be given by

$$P_{\mathrm{T}} = \frac{T-\tau_{\mathrm{u}}}{T}\frac{1}{\zeta}\sum_{k=1}^{K} p_k, \tag{8}$$

where ζ is the amplifier efficiency.

4) In distributed massive MIMO systems, the backhaul-inducing power is modeled as [19]

$$P_{\mathrm{BH}} = M\Big(P_0 + BP_{\mathrm{BT}}\sum\nolimits_{k=1}^{K} S_k(\mathbf{p})\Big), \tag{9}$$

where P_0 is the traffic-independent power consumption of each backhaul link and P_{BT} is the traffic-dependent power.

In this paper, we aim at optimizing the the power allocation ($\mathbf{p}$) to maximize the EE $\phi(\mathbf{p})$ under *ergodic* per-user QoS and per-RAU transmit power constraints. The optimization problem with variable $\mathbf{p}$ can be formulated as

$$(\mathcal{P}_0): \begin{cases} \max\limits_{\mathbf{p}} \ \phi(\mathbf{p}) \\ \text{s.t. } S_k(\mathbf{p}) \geq S_{\min}, \ \forall\, k, \\ \quad \sum_{k=1}^{K} \omega_{m,k}p_k \leq P_{\max}, \ \forall\, m, \\ \quad p_k \geq 0, \ \forall\, k, \end{cases} \tag{10}$$

where $S_{\min}$ is the minimum spectral efficiency (QoS) required by each user, $P_{\max}$ represents the maximal power constraint at the m-th RAU, the power control parameter

$$\omega_{m,k} = \mathbb{E}\left[\|[\vec{\mathbf{w}}_{m,k}\|^2\right], \tag{11}$$

and $\vec{\mathbf{w}}_{m,k}$ is the $(m-1)N+1 : mN$ elements of the beamforming vectors $\vec{\mathbf{w}}_k$.

Remark 1: It is extremely computationally expensive to maximizes the EE under *instantaneous* QoS and transmit power constraints, since the small-scale fading coefficients change quickly. In optimization problem $(\mathcal{P}_0)$, we consider *ergodic* (long-term) QoS and transmit power constraints which depend on channel statistics (not the instantaneous fading realizations) due to the channel hardening effects in massive MIMO systems. Thus, $(\mathcal{P}_0)$ is more practically appealing since the EE maximization problem can be solved over a longer time period.

The challenge of solving the optimization problem $(\mathcal{P}_0)$ is twofold. First, the closed-form expressions for the SE $S_k(\mathbf{p})$ and power control parameters $\omega_{m,k}$ with ZF and MRT beamforming in distributed massive MIMO systems are unknown. Second, $\mathcal{P}_0$ is non-convex because of the non-convexity of $S_k(\mathbf{p})$ with respect to $\mathbf{p}$. To overcome this dilemma, in the next section, we will first derive accurate closed-form expressions for $S_k(\mathbf{p})$ and $\omega_{m,k}$, and then using the tools of fractional programming and SCA, we will transform the non-convex optimization problem $(\mathcal{P}_0)$ into an equivalent concave optimization problem which can be efficiently solved.

III. POWER ALLOCATION FOR EE MAXIMIZATION

As a necessary basis, accurate closed-form expressions for the SE and the power control parameters with ZF and MRT beamforming are derived firstly in the first and second parts of this section. The third part provides the feasible conditions of the considered EE maximization problem. In the fourth part, with the aid of the fractional programming and SCA techniques, we propose a computationally feasible power allocation algorithm by solving an efficiently solvable concave optimization problem which is equivalent to the original non-convex EE maximization problem.

A. SPECTRAL EFFICIENCY

The downlink signal to user k is assigned a transmit power p_k and a normalized beamforming direction $\vec{\mathbf{w}}_k$. Assuming that users detect the desired signals with the knowledge of channel statistics (i.e., $\mathbb{E}[\mathbf{g}_k^{\mathrm{H}}\mathbf{w}_k]$), the downlink SE of the k-th user can be expressed as [26, Th. 1]

$$S_k(\mathbf{p}) = \frac{T-\tau_{\mathrm{p}}}{T}\log_2\left(1+\frac{p_k\mathcal{U}_k}{p_k\mathcal{V}_k+\sum_{j\neq k}^{K}p_j\mathcal{I}_j+\sigma_k^2}\right), \quad (12)$$

where T is the length of channel coherence interval, and

$$\mathcal{U}_k \triangleq |\mathbb{E}[\mathbf{g}_k^{\mathrm{H}}\vec{\mathbf{w}}_k]|^2, \quad (13)$$
$$\mathcal{V}_k \triangleq \mathrm{var}[\mathbf{g}_k^{\mathrm{H}}\vec{\mathbf{w}}_k], \quad (14)$$
$$\mathcal{I}_j \triangleq \mathbb{E}\big[|\mathbf{g}_k^{\mathrm{H}}\vec{\mathbf{w}}_j|^2\big], \quad (15)$$

are the useful signal power, the beamforming uncertainty and inter-user interference, respectively. The following theorems present the derivations of the closed-form expressions for $\mathcal{U}_k$, $\mathcal{V}_k$ and $\mathcal{I}_k$ with ZF and MRT beamforming.

Theorem 1: When MRT beamforming is employed, $\mathcal{U}_k$, $\mathcal{V}_k$ and $\mathcal{I}_k$ can be given in closed-form by

$$\mathcal{U}_k^{\mathrm{MRT}} = \xi(\hat{\mu}_{k,\mathrm{a}})\hat{\theta}_{k,\mathrm{a}}, \quad (16)$$
$$\mathcal{V}_k^{\mathrm{MRT}} = \hat{\mu}_{k,\mathrm{a}}\hat{\theta}_{k,\mathrm{a}} - \xi(\hat{\mu}_{k,\mathrm{a}})\hat{\theta}_{k,\mathrm{a}} + \frac{1}{MN}\tilde{\mu}_{k,\mathrm{a}}\tilde{\theta}_{k,\mathrm{a}}, \quad (17)$$
$$\mathcal{I}_j^{\mathrm{MRT}} = \frac{1}{MN}\mu_{k,\mathrm{a}}\theta_{k,\mathrm{a}}, \quad (18)$$

where $\xi(x) \triangleq \frac{\Gamma(x+1/2)}{\Gamma(x)}$, and

$$\mu_{k,\mathrm{a}} = \frac{N(\sum_{m=1}^{M}\lambda_{m,k})^2}{\sum_{m=1}^{M}\lambda_{m,k}^2}, \quad \theta_{k,\mathrm{a}} = \frac{\sum_{m=1}^{M}\lambda_{m,k}^2}{\sum_{m=1}^{M}\lambda_{m,k}},$$
$$\hat{\mu}_{k,\mathrm{a}} = \frac{N(\sum_{m=1}^{M}\beta_{m,k})^2}{\sum_{m=1}^{M}\beta_{m,k}^2}, \quad \hat{\theta}_{k,\mathrm{a}} = \frac{\sum_{m=1}^{M}\beta_{m,k}^2}{\sum_{m=1}^{M}\beta_{m,k}},$$
$$\tilde{\mu}_{k,\mathrm{a}} = \frac{N(\sum_{m=1}^{M}\eta_{m,k})^2}{\sum_{m=1}^{M}\eta_{m,k}^2}, \quad \tilde{\theta}_{k,\mathrm{a}} = \frac{\sum_{m=1}^{M}\eta_{m,k}^2}{\sum_{m=1}^{M}\eta_{m,k}}.$$

Proof: See Appendix A. ■

Theorem 2: When ZF beamforming is employed, $\mathcal{U}_k$, $\mathcal{V}_k$ and $\mathcal{I}_k$ can be given in closed-form by

$$\mathcal{U}_k^{\mathrm{ZF}} = \xi(\frac{\chi}{MN}\hat{\mu}_{k,\mathrm{a}})\hat{\theta}_{k,\mathrm{a}}, \quad (19)$$
$$\mathcal{V}_k^{\mathrm{ZF}} = \frac{\chi}{MN}\hat{\mu}_{k,\mathrm{a}}\hat{\theta}_{k,\mathrm{a}} - \xi(\frac{\chi}{MN}\hat{\mu}_{k,\mathrm{a}})\hat{\theta}_{k,\mathrm{a}} + \frac{1}{MN}\tilde{\mu}_{k,\mathrm{a}}\tilde{\theta}_{k,\mathrm{a}}, \quad (20)$$
$$\mathcal{I}_j^{\mathrm{ZF}} = \frac{1}{MN}\tilde{\mu}_{k,\mathrm{a}}\tilde{\theta}_{k,\mathrm{a}}, \quad (21)$$

where $\chi \triangleq MN - K + 1$.

Proof: See Appendix B. ■

B. POWER CONTROL PARAMETERS

In this paper, beamforming vectors are computed through all RAUs since users are served by all RAUs in a coherent way. Thus, when transmitting signals to user k with power p_k, the transmitted power from RAU m is $\omega_{m,k}p_k$. In the following theorems, we derive closed-form expressions for the power control parameters $\omega_{m,k}$ with ZF and MRT beamforming, which have, to the best of authors' knowledge, never been given before.

Theorem 3: With MRT beamforming, for the power control parameters $\omega_{m,k}$, we have

$$\omega_{m,k}^{\mathrm{MRT}} = \frac{c_{m,k}^{N}N}{N+\rho_{m,k}}{}_2\mathrm{F}_1(N+\rho_{m,k}, N+1; N+\rho_{m,k}+1; 1-c_{m,k}), \quad (22)$$

where $\rho_{m,k} = \frac{N(\sum_{i\neq m}^{M}\beta_{i,k})^2}{\sum_{i\neq m}^{M}\beta_{i,k}^2}$, $c_{m,k} = \frac{\sum_{i\neq m}^{M}\beta_{i,k}^2}{\beta_{m,k}\sum_{i\neq m}^{M}\beta_{i,k}}$, and ${}_2\mathrm{F}_1$ is the Gauss hypergeometric function.

Proof: See Appendix C. ■

Theorem 4: With ZF beamforming, for the power control parameters $\omega_{m,k}$, the following limit holds almost surely

$$\lim_{K\to\infty}\omega_{m,k}^{\mathrm{ZF}} = \frac{N}{K}\frac{K\beta_{m,k}+\dot{\Omega}_{m,k}}{\vartheta_k}\frac{1}{\left(1+\sum_{k=1}^{K}\frac{\beta_{m,k}}{\vartheta_k}\right)^2}, \quad (23)$$

where $\vartheta_k = \frac{MN-K+1}{M}\sum_{m=1}^{M}\beta_{m,k}$, $\dot{\Omega}_{m,k}$ is the solution to the linear system $(\mathbf{I}-N\mathbf{Q})\,\dot{\Omega}_m = NK\mathbf{Q}\mathbf{b}_m$, $\dot{\Omega}_m = \mathrm{diag}(\dot{\Omega}_{m,1},\cdots,\dot{\Omega}_{m,K})$, $\mathbf{b}_m = \left(\beta_{m,1},\cdots,\beta_{m,K}\right)^{\mathrm{T}}$, and $\mathbf{Q} = \sum_{m=1}^{M}\frac{1}{\left(1+\sum_{j=1}^{K}\frac{\beta_{m,j}}{\vartheta_j}\right)^2}\mathbf{b}_m\mathbf{b}_m^{\mathrm{T}}\mathrm{diag}\left(\frac{1}{\vartheta_1^2},\cdots,\frac{1}{\vartheta_K^2}\right)$.

Proof: See Appendix D. ∎

Remark 2: For user k, $\omega_{m,k}$ equals to zero for some values of m. It means that user k is not associated with RAU m. Thus, the RAU-user association problem is solved implicitly.

Now, the per-user QoS and per-RAU transmit power constraints are both given in closed-form and related to only the large-scale fading coefficients. Thus, we need to solve the EE optimization problem $(\mathcal{P}_0)$ only when the large-scale fading coefficients change. In the following, we present feasibility conditions of $(\mathcal{P}_0)$.

C. FEASIBILITY OF THE EE MAXIMIZATION PROBLEM

The feasibility of the EE maximization problem $(\mathcal{P}_0)$ amounts to providing conditions ensuring that the feasible set is not empty. Given $\mathcal{U}_k$, $\mathcal{V}_k$ and $\mathcal{I}_k$, the necessary and sufficient conditions for $(\mathcal{P}_0)$ to be feasible are given in the following proposition.

Proposition 1: The EE maximization problem $(\mathcal{P}_0)$ is feasible if and only if

$$\rho_{\mathbf{T}} < 1 \quad \text{and} \quad \boldsymbol{\omega}(\mathbf{I}-\mathbf{T})^{-1}\mathbf{s} \le P_{\max}\mathbf{I}_M, \tag{24}$$

where $\rho_{\mathbf{T}}$ denotes the spectral radius of matrix $\mathbf{T}$,

$$[\mathbf{T}]_{k,j} = \begin{cases} 0, & j = k, \\ \dfrac{\mathcal{I}_k\underline{\gamma}_k}{\mathcal{U}_k - \mathcal{V}_k\underline{\gamma}_k}, & j \ne k, \end{cases} \tag{25}$$

$\boldsymbol{\omega} = [\boldsymbol{\omega}_1,\cdots,\boldsymbol{\omega}_M]$, $\boldsymbol{\omega}_m = [\omega_{m,1},\cdots,\omega_{m,K}]$, $\underline{\gamma}_k = 2^{S_{\min}\frac{T}{T-\tau_{\mathrm{u}}}} - 1$, and $\mathbf{s}$ is a $K\times 1$ vector with the $k-$th element given by $[\mathbf{s}]_k = \frac{\underline{\gamma}_k\sigma_k^2}{\mathcal{U}_k - \underline{\gamma}_k\mathcal{V}_k}$

Proof: See Appendix E. ∎

From Proposition 1, we can check whether the EE maximization problem $(\mathcal{P}_0)$ is feasible. If $(\mathcal{P}_0)$ is feasible, we propose a computationally feasible power allocation algorithm to solve $(\mathcal{P}_0)$ in the following subsection.

D. POWER ALLOCATION ALGORITHM

Dividing the numerator and the denominator of $\phi(\mathbf{p})$ by $B\sum_{k=1}^{K}S_k(\mathbf{p})$, it can be found that maximizing $\phi(\mathbf{p})$ is equivalent to maximize

$$\bar{\phi}(\mathbf{p}) = \frac{B\sum_{k=1}^{K}S_k(\mathbf{p})}{P_{\mathrm{IND}} + \frac{T-\tau_{\mathrm{u}}}{\zeta T}\sum_{k=1}^{K}p_k}, \tag{26}$$

where $P_{\mathrm{IND}} = P_{\mathrm{TC}} + P_{\mathrm{LP}} + MP_0$ is the power consumption independent of $\mathbf{p}$. Thus, the optimization problem $(\mathcal{P}_0)$ is equivalent to

$$(\mathcal{P}_1): \begin{cases} \max\limits_{\mathbf{p}} \ \bar{\phi}(\mathbf{p}) & \text{(27a)} \\ \text{s.t.} \ \ S_k(\mathbf{p}) \ge S_{\min}, \ \forall\, k, & \text{(27b)} \\ \quad \sum\limits_{k=1}^{K}\omega_{m,k}p_k \le P_{\max}, \ \forall\, m, & \text{(27c)} \\ \quad p_k \ge 0, \ \forall\, k, & \text{(27d)} \end{cases}$$

The objective function $\bar{\phi}(\mathbf{p})$ of optimization problem $(\mathcal{P}_1)$ is in fraction form. If the numerator of $\bar{\phi}(\mathbf{p})$ is concave, the denominator of $\bar{\phi}(\mathbf{p})$ is convex, and the constraint set of $(\mathcal{P}_1)$ is convex, the fractional programming theory can be used to solve the optimization problem $(\mathcal{P}_1)$ efficiently [27]. Otherwise, no low-complexity optimization methods are available. Unfortunately, the numerator of $\bar{\phi}(\mathbf{p})$ is not concave and the constraint (27b) is not convex due to the non-convexity of $S_k(\mathbf{p})$. To address this difficult, we transform the non-convex optimization problem $(\mathcal{P}_1)$ into an efficiently solvable concave optimization problem by integrating fractional programming and SCA theories. The key issue is to find a suitable convex approximation for $S_k(\mathbf{p})$ which can be addressed by using the following lower bound for $\log_2(1+\gamma_k)$ [28]

$$\log_2(1+\gamma_k) \ge a_k\log_2(\gamma_k) + b_k, \quad \forall\gamma_k,\ \tilde{\gamma}_k \ge 0, \tag{28}$$

where

$$a_k = \frac{\tilde{\gamma}_k}{1+\tilde{\gamma}_k}, \tag{29}$$

$$b_k = \log_2(1+\tilde{\gamma}_k) - \frac{\tilde{\gamma}_k}{1+\tilde{\gamma}_k}\log_2\tilde{\gamma}_k. \tag{30}$$

The lower bound in (28) is tight since both the two sides of (28) and their derivatives with respect to γ_k are equal at $\gamma_k = \tilde{\gamma}_k$.

By applying the lower bound in (28) to $S_k(\mathbf{p})$ in (12) and letting $\tilde{p}_k = \log_2 p_k$, the objective function (27a) of $(\mathcal{P}_1)$ becomes

$$\tilde{\phi}(\tilde{\mathbf{p}}) = \frac{B\sum_{k=1}^{K}\tilde{S}_k(\tilde{\mathbf{p}})}{P_{\mathrm{IND}} + \frac{T-\tau_{\mathrm{u}}}{\zeta T}\sum_{k=1}^{K}2^{\tilde{p}_k}}, \tag{31}$$

where $\tilde{\mathbf{p}} = [\tilde{p}_1,\cdots,\tilde{p}_K]$, and $\tilde{S}_k(\tilde{\mathbf{p}}) = \frac{T-\tau_{\mathrm{u}}}{T}\big(a_k\log_2(\mathcal{U}_k) + a_k\tilde{p}_k + b_k - a_k\log_2\left(2^{\tilde{p}_k}\mathcal{V}_k + \sum_{j\ne k}^{K}2^{\tilde{p}_j}\mathcal{I}_j + \sigma_k^2\right)\big)$. Observe now that the numerator and denominator of $\tilde{\phi}(\tilde{\mathbf{p}})$ are respectivelt concave and convex in $\tilde{\mathbf{p}}$ for any given a_k and b_k.

For the non-convex QoS constraint (27b), we rewrite it as

$$2^{\tilde{p}_k}\left(\mathcal{U}_k - \underline{\gamma}_k\mathcal{V}_k\right) \ge \underline{\gamma}_k\left(\sum_{j\ne k}^{K}2^{\tilde{p}_j}\mathcal{I}_j + \sigma_k^2\right), \tag{32}$$

Taking the logarithm of both sides yields

$$\tilde{p}_k + \log_2\left(\frac{\mathcal{U}_k - \underline{\gamma}_k\mathcal{V}_k}{\underline{\gamma}_k}\right) - \log_2\left(\sum_{j\ne k}^{K}2^{\tilde{p}_j}\mathcal{I}_j + \sigma_k^2\right) \ge 0, \tag{33}$$

which turns out to be convex in $\tilde{\mathbf{p}}$.

Using the above results, we obtain the following optimization problem for given coefficients a_k and b_k

$$(\tilde{\mathcal{P}}_1): \begin{cases} \max\limits_{\tilde{\mathbf{p}}} \ \tilde{\phi}(\tilde{\mathbf{p}}) & (34a) \\ \text{s.t. } (33), \quad \forall\, k, & (34b) \\ \quad \sum\limits_{k=1}^{K} \omega_{m,k} 2^{\tilde{p}_k} \le P_{\max}, \quad \forall\, m, & (34c) \end{cases}$$

which is a concave fractional programming problem and thus can be efficiently solved, e.g., by Dinkelbach's algorithm [29], with affordable polynomial complexity [27]. However, the maximum of $(\tilde{\mathcal{P}}_1)$ is obtained for given coefficients a_k and b_k. Thus, we need to update a_k and b_k further to tighten the lower bound of $S_k(\mathbf{p})$. The overall iterative procedure is summarized in Algorithm 1.

Algorithm 1 Power Allocation for EE Maximization

1: Test feasibility by Proposition 1;
2: **if** Feasible **then**
3: Set the maximum number of iterations $N_{\max}$, and the convergence tolerance $\epsilon > 0$;
4: Set $i = 0$ and select a feasible $\mathbf{p}(0)$;
5: Set $\tilde{\gamma}_k(0) = \gamma_k(\mathbf{p}(0))$ and compute $a_k(0)$, $b_k(0)$ for all k;
6: **repeat**
7: $i = i + 1$;
8: Solve $(\tilde{\mathcal{P}}_1)$, and call the solution $\tilde{\mathbf{p}}(i)$;
9: $\mathbf{p}(i) = 2^{\tilde{\mathbf{p}}(i)}$, $\tilde{\gamma}_k(i) = \gamma_k(\mathbf{p}(i))$ and update $a_k(i)$, $b_k(i)$ for all k;
10: **until** convergence $\left(\max\limits_k \left|\frac{\tilde{\gamma}_k(i)-\tilde{\gamma}_k(i-1)}{\tilde{\gamma}_k(i-1)}\right| < \epsilon \text{ or } i = N_{\max}\right)$
11: **end if**

For completeness, we provide the method of solving $(\tilde{\mathcal{P}}_1)$ by Dinkelbach's algorithm in the following. After converting $(\tilde{\mathcal{P}}_1)$ into a sequence of convex problems

$$(\mathcal{P}_2): \begin{cases} \max\limits_{\tilde{\mathbf{p}}} \ F(\lambda) = f(\tilde{\mathbf{p}}) - \lambda g(\tilde{\mathbf{p}}) & (35a) \\ \text{s.t. } (33), \ \forall\, k, & (35b) \\ \quad \sum\limits_{k=1}^{K} \omega_{m,k} 2^{\tilde{p}_k} \le P_{\max}, \ \forall\, m, & (35c) \end{cases}$$

where $f(\tilde{\mathbf{p}}) = B\sum_{k=1}^{K} \tilde{S}_k(\tilde{\mathbf{p}})$ and $g(\tilde{\mathbf{p}}) = P_{\text{IND}} + \frac{T-\tau_{\text{u}}}{\zeta T} \times \sum_{k=1}^{K} 2^{\tilde{p}_k}$, Dinkelbach's algorithm works by finding the solutions of $(\tilde{\mathcal{P}}_2)$ in each iteration and then updating the parameter λ. This method is summarized in Algorithm 2, which is guaranteed to converge to the global maximum of $(\tilde{\mathcal{P}}_1)$ after a few iterations [29], [30].

Proposition 2: After each iteration, the lower bound of $S_k(\mathbf{p})$ is tightened, the EE value $\tilde{\phi}(\tilde{\mathbf{p}})$ is increased, and Algorithm 1 converges to KKT points of the optimization problem $(\mathcal{P}_1)$.

Proof: See Appendix F. ■

Algorithm 2 Solve $(\tilde{\mathcal{P}}_1)$ Based on Dinkelbach's Algorithm

1: Set $\epsilon > 0$; $j = 0$; $\lambda(j) = 0$;
2: **repeat**
3: Solve $(\mathcal{P}_2)$ and obtain $\tilde{\mathbf{p}}^*(j)$;
4: $F(\lambda(j)) = f(\tilde{\mathbf{p}}^*(j)) - \lambda(j) g(\tilde{\mathbf{p}}^*(j))$;
5: $\lambda(j+1) = f(\tilde{\mathbf{p}}^*(j)) \big/ g(\tilde{\mathbf{p}}^*(j))$;
6: $j = j + 1$;
7: **until** $F(\lambda(j)) < \epsilon$

Remark 3 (Multi-cell extension): In multi-cell with pilot contamination [10] scenarios, the SE of the user k in cell l can be written as

$$S_{l,k}(\mathbf{p}) = \frac{T-\tau_{\text{u}}}{T} \log_2 \times \left(1 + \frac{p_{l,k}\mathcal{U}_{l,k}}{p_{l,k}\mathcal{V}_{l,k} + \sum_{(i,j)\neq(l,k)} p_{i,j}\mathcal{I}_{l,k} + \sigma_{l,k}^2}\right), \quad (36)$$

where $\mathcal{U}_{l,k} \triangleq |\mathbb{E}[\mathbf{g}_{l,l,k}^{\text{H}} \vec{\mathbf{w}}_{l,k}]|^2$, $\mathcal{V}_{l,k} \triangleq \text{var}\left[\mathbf{g}_{l,l,k}^{\text{H}} \vec{\mathbf{w}}_{l,k}\right]$, $\mathcal{I}_{l,k} \triangleq \mathbb{E}\left[|\mathbf{g}_{i,l,k}^{\text{H}} \vec{\mathbf{w}}_{i,j}|^2\right]$, and the power allocation needs to be optimized among L cells by solving the following global EE maximization problem

$$(\mathcal{P}_{\text{mul}}): \begin{cases} \max\limits_{\mathbf{p}} \ \dfrac{B\sum_{l=1}^{L}\sum_{k=1}^{K} S_{l,k}(\mathbf{p})}{P_{\text{Total}}(\mathbf{p})} \\ \text{s.t. } S_{l,k}(\mathbf{p}) \ge S_{\min}, \quad \forall\, l, k, \\ \quad \sum\limits_{k=1}^{K} \omega_{l,m,k} p_{l,k} \le P_{\max}, \quad \forall\, l, m, \\ \quad p_{l,k} \ge 0, \quad \forall\, l, k, \end{cases} \quad (37)$$

where $\mathbf{p} = [\mathbf{p}_1^{\text{T}}, \cdots, \mathbf{p}_L^{\text{T}}]^{\text{T}}$, $\mathbf{p}_l = [p_{l,1}, \cdots, p_{l,K}]^{\text{T}}$, $P_{\text{Total}}(\mathbf{p}) = LP_{\text{IND}} + \frac{T-\tau_{\text{u}}}{\zeta T}\sum_{l=1}^{L}\sum_{k=1}^{K} p_{l,k} + P_{\text{BT}}MB \times \sum_{l=1}^{L}\sum_{k=1}^{K} S_{l,k}(\mathbf{p})$ is the total power consumption of L cells. Based on the method used in this paper, we can obtain the closed-form expressions for $S_{l,k}(\mathbf{p})$ and $\omega_{l,m,k}$ where pilot contamination [31] makes the derivation process more difficult (similar derivation for $S_{l,k}(\mathbf{p})$ please refer to [10]). Moreover, the method of solving the optimization problem $(\mathcal{P}_0)$ can also be used to solve the $(\mathcal{P}_{\text{mul}})$. Thus, the analytic framework and the power allocation algorithm can be applied in multi-cell scenarios.

In conclusion of this section, after deriving the analytical expressions for the SE $S_k(\mathbf{p})$ and power control parameters $\omega_{m,k}$, we solved the EE maximization problem under ergodic per-user QoS and per-RAU power constraints by integrating fractional programming and SCA theories and proposed a computationally feasible power allocation algorithm with guaranteed convergence to the KKT points of the original non-convex EE maximization problem.

IV. NUMERICAL RESULTS

In this section, the theoretical analysis presented in previous sections are verified by numerical results.

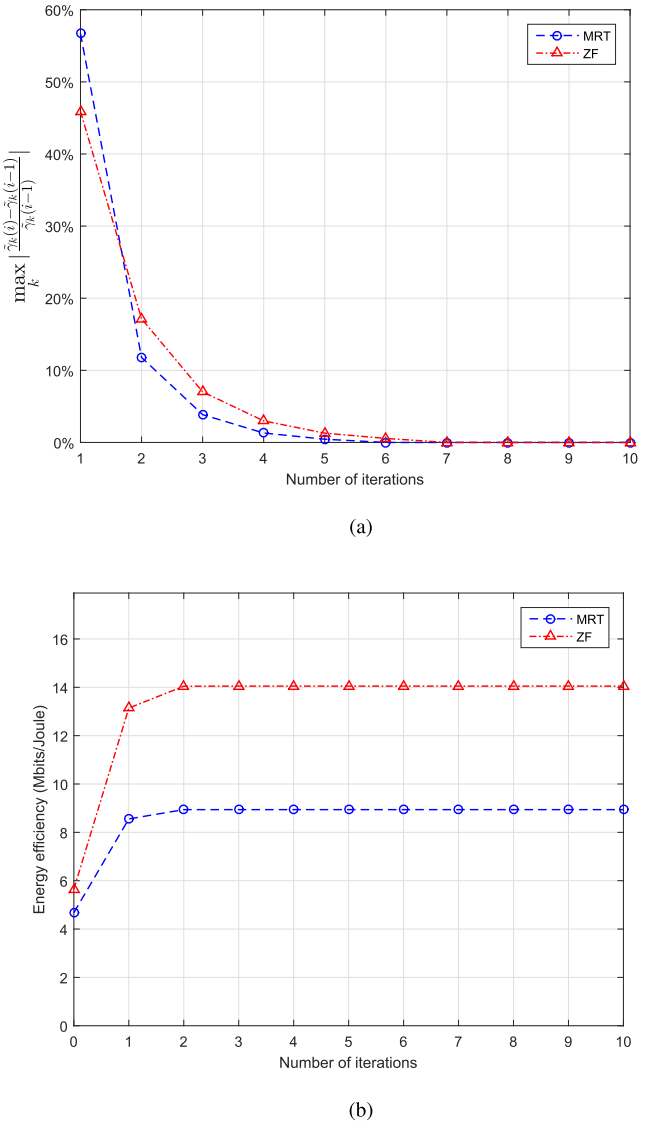


FIGURE 1. The convergence procedure of Algorithm 1, $N = 8$, $K = 10$. (a) EE versus the number of iterations.

A. SIMULATION SETUP

Consider a circular coverage area with a radius 1 km where there are seven RAUs jointly serving K users. To ensure the reproducibility, we assume that one RAU is deployed in the center and the other six RAUs are uniformly deployed on the circle of radius $(3-\sqrt{3})/2$. The K users are distributed uniformly. The distance between RAUs and users is not less than 50 m. The large-scale fading coefficient between RAU m and user k is modeled using the COST Hata model [23] as $\lambda_{m,k} = 10^{-13.6-3.5\log_{10}(d_{m,k})+s_{m,k}/10}$, where $d_{m,k}$ is the corresponding distance and $s_{m,k}$ represents the independent log-normal shadowing with a standard deviation of 8 dB. We choose system bandwidth $B = 20$ MHz, minimal spectral efficiency $S_{\min} = 1$ bit/s/Hz, coherence interval in symbols $T = 196$, power of pilot sequences $p^{\mathrm{u}}_{l,k} = 0.2$ W, length of pilot sequences $\tau_{\mathrm{p}} = K$, and a noise figure equals to 9 dB [19], [21]. The maximum transmitted power of each antenna is set to 0.2 W, and thus the maximal transmit power of each RAU $P_{\max} = 0.2 \times N$ W. Unless otherwise stated, the other parameters are chosen as in Table 1 [19], [24].

TABLE 1. Power Consumption Parameters.

Parameters	Value
Power consumption/antenna at transmitters, P_{TR}	0.2 W
Power consumption/antenna at users, P_{UE}	0.1 W
Power consumption/local oscillator, P_{SYN}	2 W
Computational efficiency at transmitters, L_{TR}	12.8 Gflops/W
Power amplifier efficiency, ζ	0.4
Fixed power consumption/backhaul, P_0	0.825 W
Traffic-dependent backhaul power, P_{BT}	0.25 W/(Gbits/s)

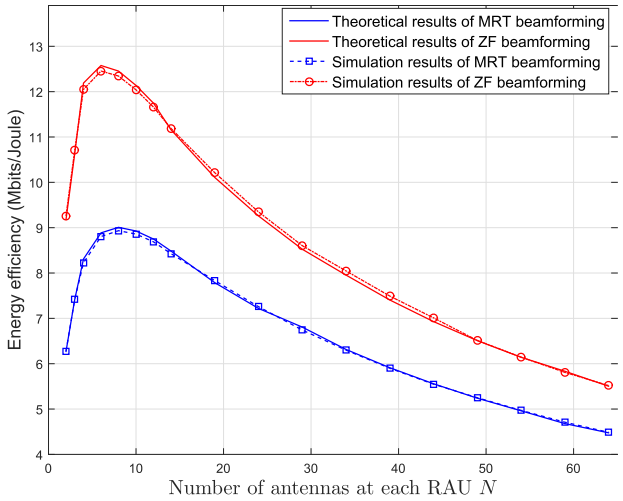


FIGURE 2. Energy efficiency obtained by simulations and analytical expressions against the number of antennas at each RAU, $K = 4$, and $M = 7$.

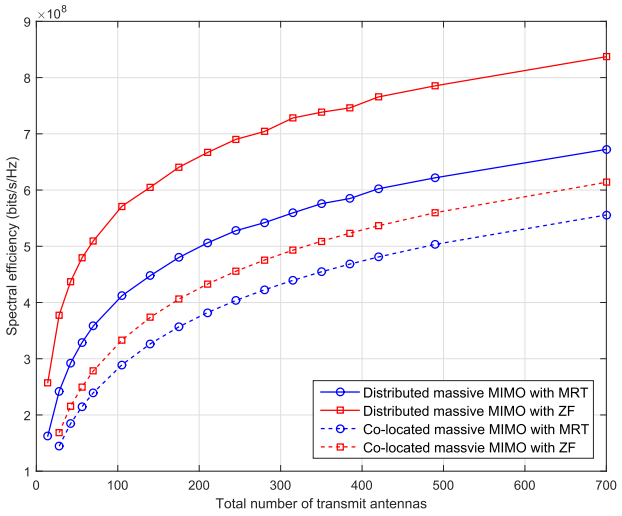
B. RESULTS AND DISCUSSIONS

1) PERFORMANCE OF CONVERGENCE

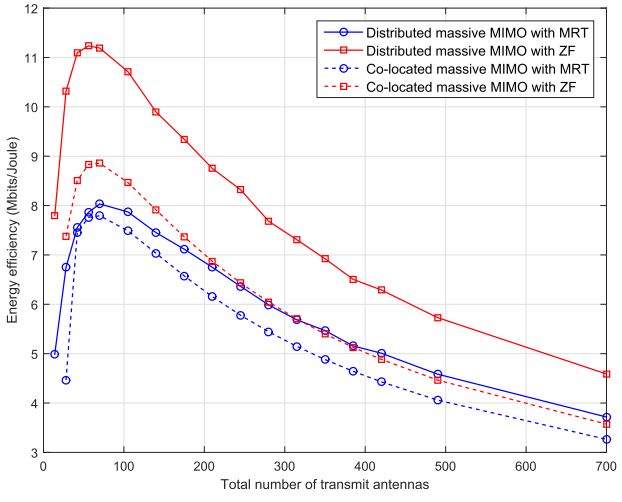
We first provide insight on the convergence procedure of Algorithm 1. The flag value $\max_k |\frac{\tilde{\gamma}_k(i)-\tilde{\gamma}_k(i-1)}{\tilde{\gamma}_k(i-1)}|$ reflects the accuracy of the convex approximation (28) for spectral efficiency. The smaller flag value, the higher approximate accuracy. Fig. 1(a) presents the flag value versus the number of iterations. When the flag value turns into zero, the obtained power allocation (by Dinkelbach's algorithm) stays the same for every iteration, which indicates the convergence of Algorithm 1. From Fig. 1(a), it can be seen that Algorithm 1 converges rapidly (about 7 iterations for both beamforming), and MRT has a faster rate of convergence. Fig. 1(b) shows the EE versus the number of iterations. It can be observed that the EE remains nearly constant after 3 iterations, which means that the convex approximation (28) is accurate enough for our EE maximization problem after 3 iterations.

2) ACCURACY OF THE ANALYTICAL RESULTS

In Fig. 2, we verify the accuracy of the analytical results obtained with the closed-form expressions for the SE and



(a)



(b)

FIGURE 3. SE and EE comparisons between distributed massive MIMO and co-located massive MIMO systems as a function of the total number of transmit antennas, $K = 5$. (a) SE comparison. (b) EE comparison.

the power control parameters derived in Theorems 1, 2, 3, 4. First, it is clearly seen that the analytical results are very tight in all cases despite the number of antennas at each RAU. Second, the energy efficiency increases and then turns to decrease as the number of antennas at each RAU N increases. This is because that as N increase, the spectral efficiency first increases largely and then tends to be constant while the energy consumption increases always. The optimal N is observed, e.g., $N = 8$ and $N = 6$ for ZF and MRT beamforming, respectively. Third, ZF achieves a larger EE than MRT since the spectral efficiency of ZF is much larger than MRT [10]. However, the improvement decreases as N increases since ZF consumes more power as shown in (7a) and (7b). The improvement equals to about 47.5% at $N = 2$ and about 22.9% at $N = 64$.

3) DISTRIBUTED MASSIVE MIMO VERSUS CO-LOCATED MASSIVE MIMO

Fig. 3 shows the SE and EE comparisons between the distributed massive MIMO system and co-located massive MIMO system (MN antennas are co-located at the center). From this figure, we get the following findings. First, compared to the co-located systems, distributed massive MIMO systems improve the SE significantly due to the decreased average access distance of the users and the multiplexing and diversity gain offered by coordination among RAUs. The SE improvement is 36.4% with ZF beamforming and 20.9% with MRT beamforming at $MN = 700$. Second, under the setting as in Table 1, distributed massive MIMO systems are more energy efficient than co-located systems. However, distributed massive MIMO systems consume more power due to more oscillators and backhaul-inducing power as defined in Section II.B. Thus, the EE improvement is only 28.4% with ZF beamforming and 13.8% with MRT beamforming at $MN = 700$. Moreover, in the simulations, the traffic-dependent backhaul power $P_{\mathrm{BT}} = 0.25$W/(Gbits/s). If the backhaul link is less power efficient which means more backhauling power consumed, the EE of distributed massive MIMO systems will decrease and even less than that of co-located systems [19]. These motivate the works considering RAU selection schemes to reduce the backhaul-induced power, which can improve the EE [15].

V. CONCLUSIONS

In this paper, we have investigated power allocation optimization for EE maximization in distributed massive MIMO systems under *ergodic* per-user QoS and per-RAU transmit power constraints, which is more practically appealing since it is related to only large-scale fading and can be solved over a longer time period. We first derived accurate closed-form expressions for the SE and power control parameters with ZF and MRT beamforming. Based on these expressions, we formulated the EE maximization problem based on a practical power consumption model. By integrating fractional programming and SCA theories, we developed a computationally feasible power allocation algorithm to solve the non-convex optimization problem, which is guaranteed to converge to the KKT points of the original problem. Based on these analytical results, we analyzed and compared the performance of different beamforming schemes and different antenna deployments. From the numerical results, it was observed that ZF achieves a larger EE than MRT, however, the improvement decrease as the number of transmit antennas increases since ZF has a higher complexity and consumes more power. Moreover, distributed massive MIMO systems have great potential to improve the SE and EE.

APPENDIX A
PROOF OF THEOREM 1

The closed-form expressions for (13)-(15) can be obtained by characterizing the distribution of the powers of non-isotropic vector $\mathbf{g}_k = [\sqrt{\lambda_{1,k}}\mathbf{h}_{1,k}, \cdots, \sqrt{\lambda_{M,k}}\mathbf{h}_{M,k}]$ projected onto

an s-dimensional beamforming subspace. The beamforming vector $\vec{\mathbf{w}}_k^{\mathrm{MRT}}$ is a function of channel estimation $\hat{\mathbf{g}}_k$ and independent of $\hat{\mathbf{g}}_j$ for $j \neq k$ and the estimation error $\tilde{\mathbf{g}}_j$ for $\forall j$. The projection power of $\hat{\mathbf{g}}_j$ onto the beamforming subspace obtained by $\vec{\mathbf{w}}_k^{\mathrm{MRT}}$ is distributed as $\Gamma(\frac{s}{MN}\hat{\mu}_{k,a}, \hat{\theta}_{k,a})$ where $s = MN$ for $j = k$ and $s = 1$ for $j \neq k$, and the projection power of $\tilde{\mathbf{g}}_j$ is distributed as $\Gamma(\frac{1}{MN}\tilde{\mu}_{k,a}, \tilde{\theta}_{k,a})$ for $\forall j$ [10, Lemma 3].

For the term $\mathcal{U}_k$, we have $|\mathbb{E}[\mathbf{g}_k^{\mathrm{H}}\vec{\mathbf{w}}_k^{\mathrm{MRT}}]|^2 = |\mathbb{E}[\|\hat{\mathbf{g}}_k\|]|^2$ due to the independence of channel estimation error $\tilde{\mathbf{g}}_k$ and channel estimate $\hat{\mathbf{g}}_k$. Since $\|\hat{\mathbf{g}}_k\|^2 = |\hat{\mathbf{g}}_k^{\mathrm{H}}\vec{\mathbf{w}}_k^{\mathrm{MRT}}|^2 \sim \Gamma(\hat{\mu}_{k,a}, \hat{\theta}_{k,a})$, we have $\|\hat{\mathbf{g}}_k\| \sim$ Nakagami$(\hat{\mu}_{k,a}, \hat{\mu}_{k,a}\hat{\theta}_{k,a})$ and $|\mathbb{E}[\|\hat{\mathbf{g}}_k\|]|^2 = \xi(\hat{\mu}_{k,a})\hat{\theta}_{k,a}$ based on the relationship between Nakagami distribution and Gamma distribution.

Considering the term $\mathcal{V}_k$, due to the independence of $\hat{\mathbf{g}}_k$ and $\tilde{\mathbf{g}}_k$, we have $\mathrm{var}[\mathbf{g}_k^{\mathrm{H}}\vec{\mathbf{w}}_k] = \mathbb{E}[|\hat{\mathbf{g}}_k^{\mathrm{H}}\vec{\mathbf{w}}_k^{\mathrm{MRT}}|^2] + \mathbb{E}[|\tilde{\mathbf{g}}_k^{\mathrm{H}}\vec{\mathbf{w}}_k^{\mathrm{MRT}}|^2] - |\mathbb{E}[\hat{\mathbf{g}}_k^{\mathrm{H}}\vec{\mathbf{w}}_k^{\mathrm{MRT}}]|^2$. From the above analysis, $|\hat{\mathbf{g}}_k^{\mathrm{H}}\vec{\mathbf{w}}_k^{\mathrm{MRT}}|^2$ and $|\tilde{\mathbf{g}}_k^{\mathrm{H}}\vec{\mathbf{w}}_k^{\mathrm{MRT}}|^2$ are distributed as $\Gamma(\hat{\mu}_{k,a}, \hat{\theta}_{k,a})$ and $\Gamma(\frac{1}{MN}\tilde{\mu}_{k,a}, \tilde{\theta}_{k,a})$, respectively. Thus, we have $\mathbb{E}[|\hat{\mathbf{g}}_k^{\mathrm{H}}\vec{\mathbf{w}}_k^{\mathrm{MRT}}|^2] = \hat{\mu}_{k,a}\hat{\theta}_{k,a}$ and $\mathbb{E}[|\tilde{\mathbf{g}}_k^{\mathrm{H}}\vec{\mathbf{w}}_k^{\mathrm{MRT}}|^2] = \frac{1}{MN}\tilde{\mu}_{k,a}\tilde{\theta}_{k,a}$. combined with $|\mathbb{E}[\hat{\mathbf{g}}_k^{\mathrm{H}}\vec{\mathbf{w}}_k^{\mathrm{MRT}}]|^2 = \mathcal{U}_k = \xi(\hat{\mu}_{k,a})\hat{\theta}_{k,a}$, we obtain the expression for $\mathcal{V}_k$ as in (17).

Similarly, for the term $\mathcal{I}_j$, we have $\mathbb{E}\big[|\mathbf{g}_k^{\mathrm{H}}\vec{\mathbf{w}}_j|^2\big] = \frac{1}{MN}\mu_{k,a}\theta_{k,a}$ since $|\mathbf{g}_k^{\mathrm{H}}\vec{\mathbf{w}}_j|^2 \sim \Gamma(\frac{1}{MN}\mu_{k,a}, \theta_{k,a})$.

APPENDIX B
PROOF OF THEOREM 2

For the term $\mathcal{U}_k$, we have

$$|\mathbb{E}[\mathbf{g}_k^{\mathrm{H}}\vec{\mathbf{w}}_k^{\mathrm{ZF}}]|^2 \overset{(a)}{=} |\mathbb{E}\left[1/\|\mathbf{v}_k\|\right]|^2 \overset{(b)}{=} \left|\mathbb{E}\Big[\big([(\hat{\mathbf{G}}^{\mathrm{H}}\hat{\mathbf{G}})^{-1}]_{k,k}\big)^{-1/2}\Big]\right|^2, \quad \text{(B.1)}$$

where (a) is obtained because $\tilde{\mathbf{g}}_k$ is independent of $\vec{\mathbf{w}}_k^{\mathrm{ZF}}$ and $\hat{\mathbf{g}}_k^{\mathrm{H}}\vec{\mathbf{w}}_k^{\mathrm{ZF}} = 1/\|\mathbf{v}_k\|$, (b) results from $\|\mathbf{v}_k\|^2 = [(\hat{\mathbf{G}}^{\mathrm{H}}\hat{\mathbf{G}})^{-1}]_{k,k}$. When the channel vectors are isotropic, $[(\hat{\mathbf{G}}^{\mathrm{H}}\hat{\mathbf{G}})^{-1}]_{k,k} \sim \Gamma(\chi, \hat{\theta}_{k,\mathrm{a}})$ [32]. Based on the lemma for the distribution of the projection power of non-isotropic channel [10, Lemma 3], we can approximate the distribution of $[(\hat{\mathbf{G}}^{\mathrm{H}}\hat{\mathbf{G}})^{-1}]_{k,k}$ as $\Gamma(\frac{\chi}{MN}\hat{\mu}_{k,\mathrm{a}}, \hat{\theta}_{k,\mathrm{a}})$. Thus, we have $|\mathbb{E}[\mathbf{g}_k^{\mathrm{H}}\vec{\mathbf{w}}_k^{\mathrm{ZF}}]|^2 = \xi(\frac{\chi}{MN}\hat{\mu}_{k,\mathrm{a}})\hat{\theta}_{k,\mathrm{a}}$. Moreover, the closed-form expressions for $\mathcal{V}_k$ and $\mathcal{I}_k$ can be obtained based on the similar analysis in proof of Theorem 1 and the main difference is that the dimension of projection is changed from MN to χ with ZF beamforming when the channel and beamforming vectors are correlated.

APPENDIX C
PROOF OF THEOREM 3

Substituting $\vec{\mathbf{w}}_k = \frac{\hat{\mathbf{g}}_k}{\|\hat{\mathbf{g}}_k\|}$ into (11), we have

$$\omega_{m,k} = \mathbb{E}\left[\frac{\|\hat{\mathbf{g}}_{m,k}\|^2}{\|\hat{\mathbf{g}}_k\|^2}\right] \overset{(a)}{=} \mathbb{E}\left[\frac{\|\hat{\mathbf{g}}_{m,k}\|^2}{\|\hat{\mathbf{g}}_{m,k}\|^2 + \sum_{i\neq m}\|\hat{\mathbf{g}}_{i,k}\|^2}\right], \quad \text{(C.2)}$$

where (a) results from (1). From [10, Lemma 2], we have $\|\hat{\mathbf{g}}_{m,k}\|^2 \sim \Gamma(N, \beta_{m,k})$ and $\sum_{i\neq m}\|\hat{\mathbf{g}}_{i,k}\|^2 \sim \Gamma(\rho_{m,k}, \kappa_{m,k})$ where $\rho_{m,k} = \frac{N\left(\sum_{i\neq m}\beta_{i,k}\right)^2}{\sum_{i\neq m}\beta_{i,k}^2}$, $\kappa_{m,k} = \frac{\sum_{i\neq m}\beta_{i,k}^2}{\sum_{i\neq m}\beta_{i,k}}$. Defining $X \sim \Gamma(N, 1)$, $Y \sim \Gamma(\rho_{m,k}, 1)$, (C.2) can be rewritten as

$$\omega_{m,k} = \mathbb{E}\left[\frac{\beta_{m,k}X}{\beta_{m,k}X + \kappa_{m,k}Y}\right] \quad \text{(C.3)}$$

$$\overset{(a)}{=} \mathbb{E}\left[\frac{Z}{Z + c_{m,k}}\right], \quad \text{(C.4)}$$

where (a) is obtained by dividing the numerator and the denominator of (C.3) by $\beta_{m,k}Y$, $c_{m,k} = \frac{\kappa_{m,k}}{\beta_{m,k}}$ and $Z = \frac{X}{Y} \sim \mathrm{B2}(N, \rho_{m,k})$, where $\mathrm{B2}(\cdot,\cdot)$ is the beta type 2 distribution. The probability density function (p.d.f) of Z is

$$f(z) = \frac{z^{N-1}(z+1)^{-(N+\rho_{m,k})}}{\mathrm{B}(N, \rho_{m,k})}, \quad x \in [0, \infty), \quad \text{(C.5)}$$

where $\mathrm{B}(\cdot,\cdot)$ denotes the Beta function. Defining $W = \frac{Z}{Z+c_{m,k}}$, we can obtain the p.d.f of W from the p.d.f. of Z with the change of variables theorem

$$f(w) = \frac{c_{m,k}^N}{\mathrm{B}(N, \rho_{m,k})}\frac{w^{N-1}(1-w)^{\rho_{m,k}-1}}{(1+(c_{m,k}-1)w)^{N+\rho_{m,k}}}, \quad w \in (0,1), \quad \text{(C.6)}$$

Thus, (C.4) can be calculated by

$$\begin{aligned}\omega_{m,k} &= \int_0^1 wf(w)dw \\ &= \frac{c_{m,k}^N}{\mathrm{B}(N, \rho_{m,k})}\int_0^1 \frac{w^N(1-w)^{\rho_{m,k}-1}}{(1+(c_{m,k}-1)w)^{N+\rho_{m,k}}}dw \\ &\overset{(a)}{=} \frac{c_{m,k}^N}{\mathrm{B}(N, \rho_{m,k})}\frac{\Gamma(N+1)\Gamma(\rho_{m,k})}{\Gamma(N+\rho_{m,k}+1)} \times \\ &\quad \times {}_2\mathrm{F}_1(N+\rho_{m,k}, N+1;\ N+\rho_{m,k}+1;\ 1-c_{m,k}) \\ &\overset{(b)}{=} c_{m,k}^N\frac{N}{N+\rho_{m,k}} \times \\ &\quad \times {}_2\mathrm{F}_1(N+\rho_{m,k}, N+1;\ N+\rho_{m,k}+1;\ 1-c_{m,k}),\end{aligned} \quad \text{(C.7)}$$

where (a) is obtained by expressing the integration with hypergeometric function ${}_2\mathrm{F}_1$, and (b) results from $\mathrm{B}(N, \rho_{m,k}) = \frac{\Gamma(N)\Gamma(\rho_{m,k})}{\Gamma(N+\rho_{m,k})}$.

APPENDIX D
PROOF OF THEOREM 4

Based on the random matrix theory (see [33] and [34]) and the approximation method for the projection of non-isotropic channel vectors (see [10]), we give the following proof.

With ZF beamforming $\vec{\mathbf{w}}_k = \frac{\mathbf{v}_k}{\|\mathbf{v}_k\|}$, for the power control parameters defined in (11), we have

$$\begin{aligned}\omega_{m,k} &= \mathbb{E}\left[\frac{\|\mathbf{v}_{m,k}\|^2}{\|\mathbf{v}_k\|^2}\right] \\ &= \mathbb{E}\Big[\mathrm{tr}\big(\mathbf{\Lambda}_m(\hat{\mathbf{G}}^+)^{\mathrm{H}}\mathbf{\Theta}^{\frac{1}{2}}\mathbf{\Delta}_k\mathbf{\Theta}^{\frac{1}{2}}\hat{\mathbf{G}}^+\mathbf{\Lambda}_m\big)\Big],\end{aligned} \quad \text{(D.8)}$$

where $\mathbf{\Lambda}_m$ is an all-zeros diagonal matrix except that the elements from $(m-1)N+1$ to mN on the main diagonal equal to one, corresponding to the m-th RAU, $\hat{\mathbf{G}}^{+} = (\hat{\mathbf{G}}^{\mathrm{H}}\hat{\mathbf{G}})^{-1}\hat{\mathbf{G}}^{\mathrm{H}}$, $\mathbf{\Theta} = \mathrm{diag}(\vartheta_1, \cdots, \vartheta_K)$, $\vartheta_k = 1/\|\mathbf{v}_k\|^2 = 1/[(\hat{\mathbf{G}}^{\mathrm{H}}\hat{\mathbf{G}})^{-1}]_{k,k}$, $\mathbf{\Delta}_k$ denotes a K-dimensional all-zeros diagonal matrix except that the k-th main diagonal element equals to one, corresponding to the k-th user.

Defining $\mathbf{\Phi}_m \triangleq \frac{1}{\sqrt{K}}\left[\hat{\mathbf{h}}_{m,1}, \cdots, \hat{\mathbf{h}}_{m,K}\right] \sim \mathcal{CN}(0, \frac{1}{K}\mathbf{I}_N)$ and $\mathbf{D}_m \triangleq \mathrm{diag}\left(\beta_{m,1}^{1/2}, \cdots, \beta_{m,K}^{1/2}\right)$, the non-zero submatrix of $\mathbf{\Lambda}_m\hat{\mathbf{G}}$ which corresponds to the channel from the m-th RAU to all of the users can be given by

$$\left[\hat{\mathbf{g}}_{m,1}, \cdots, \hat{\mathbf{g}}_{m,K}\right] = \sqrt{K}\mathbf{\Phi}_m\mathbf{D}_m, \tag{D.9}$$

and thus,

$$\hat{\mathbf{G}}^{\mathrm{H}}\hat{\mathbf{G}} = \sum_{m=1}^{M}\mathbf{D}_m\mathbf{\Phi}_m^{\mathrm{H}}\mathbf{\Phi}_m\mathbf{D}_m. \tag{D.10}$$

With the definition of $\mathbf{A} \triangleq \frac{1}{K}\hat{\mathbf{G}}^{\mathrm{H}}\hat{\mathbf{G}}$, (D.8) can be rewritten as

$$\begin{aligned}\omega_{m,k} &= \mathbb{E}\left[\frac{1}{K}\mathrm{tr}\left(\mathbf{\Phi}_m\mathbf{D}_m\mathbf{A}^{-1}\mathbf{\Theta}^{\frac{1}{2}}\mathbf{\Delta}_k\mathbf{\Theta}^{\frac{1}{2}}\mathbf{A}^{-1}\mathbf{D}_m\mathbf{\Phi}_m^{\mathrm{H}}\right)\right]\\ &= \mathbb{E}\left[\frac{1}{K}\sum_{n=1}^{N}\boldsymbol{\phi}_{m,n}\mathbf{D}_m\mathbf{A}^{-1}\mathbf{\Theta}^{\frac{1}{2}}\mathbf{\Delta}_k\mathbf{\Theta}^{\frac{1}{2}}\mathbf{A}^{-1}\mathbf{D}_m\boldsymbol{\phi}_{m,n}^{\mathrm{H}}\right]\\ &\overset{(a)}{\to} \frac{N}{K}\frac{\psi(\mathbf{D}_m\mathbf{A}^{-1}\mathbf{C}_k\mathbf{A}^{-1}\mathbf{D}_m)}{\left(1+\psi(\mathbf{D}_m\mathbf{A}^{-1}\mathbf{D}_m)\right)^2},\end{aligned} \tag{D.11}$$

where $\boldsymbol{\phi}_{m,n}$ is the n-th row of $\mathbf{\Phi}_m$, (a) results from [33, Lemma 3], $\mathbf{C}_k \triangleq \mathbf{\Theta}^{\frac{1}{2}}\mathbf{\Delta}_k\mathbf{\Theta}^{\frac{1}{2}}$, and $\psi(\cdot) \triangleq \lim_{K\to\infty}\frac{1}{K}\mathrm{tr}(\cdot)$.

In (D.11), there are two limiting traces that need to be computed. For the term in the denominator, we have

$$\begin{aligned}\psi(\mathbf{D}_m\mathbf{A}^{-1}\mathbf{D}_m) &= \lim_{K\to\infty}\frac{1}{K}\mathrm{tr}(\mathbf{D}_m\mathbf{A}^{-1}\mathbf{D}_m)\\ &\overset{(a)}{=} \lim_{K\to\infty}\frac{1}{K}\mathrm{tr}\left(K(\hat{\mathbf{G}}^{\mathrm{H}}\hat{\mathbf{G}})^{-1}\mathbf{D}_m^2\right)\\ &\overset{(b)}{=} \lim_{K\to\infty}\sum_{k=1}^{K}\frac{\beta_{m,k}}{\vartheta_k},\end{aligned} \tag{D.12}$$

where (a) is obtained by substituting $\mathbf{A} = \frac{1}{K}\hat{\mathbf{G}}^{\mathrm{H}}\hat{\mathbf{G}}$, (b) results from $\vartheta_k = 1/[(\hat{\mathbf{G}}^{\mathrm{H}}\hat{\mathbf{G}})^{-1}]_{k,k}$ and $\mathbf{D}_m = \mathrm{diag}\left(\beta_{m,1}^{1/2}, \cdots, \beta_{m,K}^{1/2}\right)$.

For the numerator of (D.11), we have

$$\begin{aligned}&\psi(\mathbf{D}_m\mathbf{A}^{-1}\mathbf{C}_k\mathbf{A}^{-1}\mathbf{D}_m)\\ &\overset{(a)}{=} \lim_{t\downarrow 0}\frac{-d}{dt}\psi\left((t\mathbf{D}_m^2+\mathbf{A})^{-1}\mathbf{C}_k\right) &\text{(D.13)}\\ &= \lim_{t\downarrow 0}\frac{-d}{dt}\psi\left((t\mathrm{I}+\mathbf{D}_m^{-1}\mathbf{A}\mathbf{D}_m^{-1})^{-1}\mathbf{D}_m^{-1}\mathbf{C}_k\mathbf{D}_m^{-1}\right), &\text{(D.14)}\end{aligned}$$

where (a) is obtained from the identity $\frac{-d}{dt}\mathrm{tr}((t\mathbf{D}_m^2+\mathbf{A})^{-1}\mathbf{C}_k) = \mathrm{tr}(\mathbf{D}_m(t\mathbf{D}_m^2+\mathbf{A})^{-1}\mathbf{C}_k(t\mathbf{D}_m^2+\mathbf{A})^{-1}\mathbf{D}_m)$. For the term $(t\mathbf{I}+\mathbf{D}_m^{-1}\mathbf{A}\mathbf{D}_m^{-1})^{-1}$, we have

$$\begin{aligned}&(t\mathbf{I}+\mathbf{D}_m^{-1}\mathbf{A}\mathbf{D}_m^{-1})^{-1}\\ &= \frac{1}{t}\left(\mathbf{I}+\frac{1}{t}\mathbf{D}_m^{-1}\mathbf{A}\mathbf{D}_m^{-1}\right)^{-1}\\ &= \frac{1}{t}\left(\mathbf{I}+\frac{1}{t}\mathbf{D}_m^{-1}\frac{1}{K}\hat{\mathbf{G}}^{\mathrm{H}}\hat{\mathbf{G}}\mathbf{D}_m^{-1}\right)^{-1}\\ &= \frac{1}{t}\left(\mathbf{I}+\frac{1}{t}\left(\frac{1}{\sqrt{K}}\hat{\mathbf{G}}\mathbf{D}_m^{-1}\right)^{\mathrm{H}}\left(\frac{1}{\sqrt{K}}\hat{\mathbf{G}}\mathbf{D}_m^{-1}\right)\right)^{-1}.\end{aligned} \tag{D.15}$$

From (D.9), we have

$$\hat{\mathbf{G}}^{\mathrm{H}} = \left[\mathbf{D}_1\mathbf{W}_1^{\mathrm{H}}, \cdots, \mathbf{D}_M\mathbf{W}_M^{\mathrm{H}}\right]^{\mathrm{T}}, \tag{D.16}$$

and thus,

$$\mathbf{D}_m^{-1}\hat{\mathbf{G}}^{\mathrm{H}} = \left[\mathbf{D}_m^{-1}\mathbf{D}_1\mathbf{W}_1^{\mathrm{H}}, \cdots, \mathbf{D}_m^{-1}\mathbf{D}_M\mathbf{W}_M^{\mathrm{H}}\right]. \tag{D.17}$$

The variance profile function [34] of $\frac{1}{\sqrt{K}}\mathbf{D}_m^{-1}\hat{\mathbf{G}}^{\mathrm{H}}$ can be given by

$$v_m(x,y) = \frac{\beta_{i,k}}{\beta_{m,k}}, \quad (x,y) \in \left[\frac{k-1}{K}, \frac{k}{K}\right) \times \left[\frac{i-1}{M}, \frac{i}{M}\right), \tag{D.18}$$

then, letting $s = 1/t$ and using [34, Lemma 2.51], we obtain

$$\begin{aligned}&\frac{1}{K}\left[\left(\mathbf{I}+s\left(\frac{1}{\sqrt{K}}\hat{\mathbf{G}}\mathbf{D}_m^{-1}\right)^{\mathrm{H}}\left(\frac{1}{\sqrt{K}}\hat{\mathbf{G}}\mathbf{D}_m^{-1}\right)\right)^{-1}\right]_{k,k}\\ &\qquad\to \int_{\frac{k-1}{K}}^{\frac{k}{K}}\Psi_m(x,s)dx,\end{aligned} \tag{D.19}$$

where

$$\Psi_m(x,s) = \frac{1}{1+\frac{MN}{K}s\mathbb{E}\left[v_m(x,Y)\Upsilon_m(Y,s)\right]}, \tag{D.20}$$

$$\Upsilon_m(y,s) = \frac{1}{1+s\mathbb{E}\left[v_m(X,y)\Psi_m(X,s)\right]}. \tag{D.21}$$

Both $\Psi_m(x,s)$ and $\Upsilon_m(y,s)$ are piecewise constant functions since $v_m(x,y)$ is piecewise constant as seen from (D.18). Denoting the q-th value of of $\Psi_m(x,s)$ by $\Psi_{m,q}(s)$ for $q = 1, \cdots, K$ and the i-th value of $\Upsilon_m(y,s)$ by $\Upsilon_{m,i}(s)$ for $i = 1, \cdots, M$, we have

$$\Psi_{m,q}(s) = \frac{1}{1+\frac{N}{K}s\sum_{i=1}^{M}\frac{\beta_{i,k}}{\beta_{m,k}}\Upsilon_{m,i}(s)}, \tag{D.22}$$

$$\Upsilon_{m,i}(s) = \frac{1}{1+\frac{1}{K}s\sum_{q=1}^{K}\frac{\beta_{i,q}}{\beta_{m,k}}\Psi_{m,q}(s)}. \tag{D.23}$$

Moreover, the nonzero diagonal elements of $\mathbf{D}_m^{-1}\mathbf{C}_k\mathbf{D}_m^{-1}$ equals to $\vartheta_k\beta_{m,k}^{-1}$. Combined with (D.14), (D.15) and (D.19), we arrive at

$$\begin{aligned}\psi\left(\mathbf{D}_m\mathbf{A}^{-1}\mathbf{C}_k\mathbf{A}^{-1}\mathbf{D}_m\right)&=\lim_{K\to\infty}\frac{1}{K}\frac{1}{t}\int_{\frac{k-1}{K}}^{\frac{k}{K}}\Psi_m(x,1/t)dx\vartheta_k\beta_{m,k}^{-1})\\&=\lim_{K\to\infty}\frac{1}{K}\frac{1}{t}\Psi_{m,k}(1/t)\vartheta_k\beta_{m,k}^{-1}.\end{aligned}\tag{D.24}$$

Defining $S_{m,q}(t)\triangleq\frac{1}{t\beta_{m,q}}\Psi_{m,q}(1/t)$, $G_{m,i}(t)\triangleq\Upsilon_{m,i}(1/t)$, we can rewrite (D.22), (D.23) as

$$S_{m,q}(t)=\frac{1}{t\beta_{m,q}+\frac{N}{K}\sum_{i=1}^{M}\beta_{i,q}G_{m,i}(t)},\tag{D.25}$$

$$G_{m,i}(t)=\frac{1}{1+\frac{1}{K}\sum_{q=1}^{K}\beta_{i,q}S_{m,q}(t)},\tag{D.26}$$

respectively, and (D.24) as

$$\psi\left(\mathbf{D}_m\mathbf{A}^{-1}\mathbf{C}_k\mathbf{A}^{-1}\mathbf{D}_m\right)=\lim_{K\to\infty}\frac{1}{K}S_{m,k}(t)\vartheta_k.\tag{D.27}$$

Taking the derivative of (D.27), we obtain

$$\lim_{t\downarrow0}\frac{-d}{dt}\psi\left((t\mathbf{D}_m^2+\mathbf{A})^{-1}\mathbf{C}_k\right)=\lim_{K\to\infty}\frac{1}{K}\dot{S}_{m,k}(0)\vartheta_k,\tag{D.28}$$

where

$$\begin{aligned}\dot{S}_{m,k}(0)&\triangleq\frac{-d}{dt}S_{m,k}(t)|_{t=0}\\&=\frac{\beta_{m,k}+\frac{N}{K}\sum_{i=1}^{M}\beta_{i,k}\dot{G}_{m,i}(0)}{\left(\frac{N}{K}\sum_{i=1}^{M}\beta_{i,q}G_{m,i}(0)\right)^2},\end{aligned}\tag{D.29}$$

$$\begin{aligned}\dot{G}_{m,i}(0)&\triangleq\frac{-d}{dt}G_{m,l}(t)|_{t=0}\\&=\frac{\frac{1}{K}\sum_{q=1}^{K}\beta_{i,q}\dot{S}_{m,q}(0)}{\left(1+\frac{1}{K}\sum_{q=1}^{K}\beta_{i,q}S_{m,q}(0)\right)^2}.\end{aligned}\tag{D.30}$$

The expressions for $S_{m,q}(0)$ and $G_{m,i}(0)$ can be obtained directly by bringing $t=0$ into (D.25) and (D.26) as

$$S_{m,q}(0)=\frac{1}{\frac{N}{K}\sum_{i=1}^{M}\beta_{i,q}G_{m,i}(0)},\tag{D.31}$$

$$G_{m,i}(0)=\frac{1}{1+\frac{1}{K}\sum_{q=1}^{K}\beta_{i,q}S_{m,q}(0)}.\tag{D.32}$$

Define $U_{m,q}=N\sum_{i=1}^{M}\beta_{i,q}G_{m,i}(0)$. By substituting (D.31) and (D.32) into the definition, we obtain the fixed point of $U_{m,q}$ as

$$U_{m,q}=N\sum_{i=1}^{M}\frac{\beta_{i,q}}{1+\sum_{k=1}^{K}\frac{\beta_{i,k}}{U_{m,k}}}.\tag{D.33}$$

Since the variance profile function of $\frac{1}{N}\hat{\mathbf{G}}^{\mathrm{H}}\hat{\mathbf{G}}$ can be given by

$$v(x,y)=\beta_{m,k},\quad(x,y)\in\left[\frac{m-1}{M},\frac{m}{M}\right)\times\left[\frac{k-1}{K},\frac{k}{K}\right),\tag{D.34}$$

from [33, Th. 4, Corollary 1], we obtain the fixed point of ϑ_k

$$\vartheta_k=\frac{1}{\left[\left(\hat{\mathbf{G}}^{\mathrm{H}}\hat{\mathbf{G}}\right)^{-1}\right]_{k,k}}=N\sum_{m=1}^{M}\frac{\beta_{m,k}}{1+\sum_{q=1}^{K}\frac{\beta_{m,q}}{\vartheta_q}}.\tag{D.35}$$

From (D.33) and (D.35), it can be found that $U_{m,q}=\vartheta_q$. Thus, (D.31) can be rewritten as

$$S_{m,q}(0)=\frac{1}{\frac{N}{K}\sum_{i=1}^{M}\beta_{i,q}G_{m,i}(0)}=\frac{K}{\vartheta_q},\tag{D.36}$$

and (D.29) becomes

$$\dot{S}_{m,k}(0)=\frac{K^2\beta_{m,q}+K\dot{\Omega}_{m,q}}{(\vartheta_q)^2},\tag{D.37}$$

where

$$\begin{aligned}\dot{\Omega}_{m,q}&=N\sum_{i=1}^{M}\beta_{i,q}\dot{G}_{m,i}(0)\\&\overset{(a)}{=}N\sum_{i=1}^{M}\beta_{i,q}\frac{\frac{1}{K}\sum_{q=1}^{K}\beta_{i,q}\dot{S}_{m,q}(0)}{\left(1+\frac{1}{K}\sum_{q=1}^{K}\beta_{i,q}S_{m,q}(0)\right)^2}\\&\overset{(b)}{=}N\sum_{i=1}^{M}\beta_{i,q}\frac{\frac{1}{K}\sum_{q'=1}^{K}\beta_{i,q'}\frac{K^2\beta_{m,q'}+K\dot{\Omega}_{m,q'}}{(\vartheta_{q'})^2}}{(1+\frac{1}{K}\sum_{q'=1}^{K}\beta_{i,q'}\frac{K}{\vartheta_{q'}})^2}\\&=KN\sum_{q'=1}^{K}\left(\sum_{i=1}^{M}\frac{\beta_{i,q}\beta_{i,q'}}{(1+\sum_{q'=1}^{K}\beta_{i,q'}/\vartheta_{q'})^2}\right)\\&\quad\times\frac{1}{(\vartheta_{q'})^2}\left(\beta_{m,q'}+\frac{\dot{\Omega}_{m,q'}}{K}\right)\end{aligned}\tag{D.38}$$

where (a) and (b) are obtained by substituting (D.30) and (D.37), respectively. It can be found that (D.38) is a system of K linear equations in the K unknow $\{\dot{\Omega}_{m,q}:q=1,\cdots,K\}$ which can be solved explicitly by

$$[\mathbf{I}-N\mathbf{Q}]\dot{\Omega}_m=NK\mathbf{Q}\mathbf{b}_m.\tag{D.39}$$

After obtaining the solutions to the system (D.39) and substituting (D.37) into (D.28), we obtain

$$\lim_{t\downarrow0}\frac{-d}{dt}\psi\left((t\mathbf{D}_m^2+\mathbf{A})^{-1}\mathbf{C}_k\right)=\frac{K\beta_{m,k}+\dot{\Omega}_{m,k}}{\vartheta_k}.\tag{D.40}$$

Combing (D.40), (D.12) and (D.11), we obtain the results shown in (23). The parameter ϑ_k was calculated by the fixed point (D.35). However, it needs iterations. In the following, we derive the closed-form expression for ϑ_k

$$\begin{aligned}\vartheta_k&=\frac{1}{\left[\left(\hat{\mathbf{G}}^{\mathrm{H}}\hat{\mathbf{G}}\right)^{-1}\right]_{k,k}}\\&\overset{(a)}{\sim}\Gamma(\frac{MN-K+1}{MN}\hat{\mu}_{k,\mathrm{a}},\hat{\theta}_{k,\mathrm{a}})\\&\overset{(b)}{=}X\left(\frac{1}{X}\sum_{i=1}^{X}\gamma_i\right)\\&\overset{(c)}{\to}X\mathbb{E}[\gamma_i]\\&\overset{(d)}{=}\frac{MN-K+1}{M}\sum_{m=1}^{M}\beta_{m,k}\end{aligned}\tag{D.41}$$

where (a) results from [10, Lemma 3], (b) is obtained by defining $\gamma_i \sim \Gamma(1, \hat{\theta}_{k,\mathrm{a}})$ and $X = \frac{MN-K+1}{MN}\hat{\mu}_{k,\mathrm{a}}$, (c) results from central limiting theorem since $\hat{\mu}_{k,\mathrm{a}} \to \infty$ as $MN \to \infty$ and (d) is because $\mathbb{E}[\gamma_i] = \hat{\theta}_{k,\mathrm{a}}$ and $\hat{\mu}_{k,\mathrm{a}}\hat{\theta}_{k,\mathrm{a}} = N\sum_{m=1}^{M}\beta_{m,k}$.

Combining all results concludes the proof.

APPENDIX E
PROOF OF PROPOSITION 1

First, we prove the sufficiency of (24). From the spectral efficiency constraint, we have

$$\frac{p_k\mathcal{U}_k}{p_k\mathcal{V}_k + \sum_{j\neq k} p_j\mathcal{I}_k + \sigma_k^2} \geq \underline{\gamma}_k, \tag{E.42}$$

which leads to

$$p_k - \sum_{j\neq k}\frac{\underline{\gamma}_k\mathcal{I}_k}{\mathcal{U}_k - \underline{\gamma}_k\mathcal{V}_k}p_j \geq \frac{\underline{\gamma}_k\sigma_k^2}{\mathcal{U}_k - \underline{\gamma}_k\mathcal{V}_k}. \tag{E.43}$$

From (E.43), we obtain

$$(\mathbf{I} - \mathbf{T})\mathbf{p} \geq \mathbf{s}. \tag{E.44}$$

Since $\mathbf{T}$ is irreducible and non-negative, from [15, Th. 2.1], it follows that a solution $\mathbf{p} \geq \mathbf{0}$ to (E.44) exists for any $\mathbf{s} \geq \mathbf{0}$ if and only if $\rho_{\mathbf{T}} < 1$. Thus, $\rho_{\mathbf{T}} < 1$ ensures that there are feasible solutions meeting the ergodic per-user QoS constraint. Moreover, the solutions need to meet the ergodic per-RAU transmit power constraints further. From the second constraint of $(\mathcal{P}_0)$, we have

$$\boldsymbol{\omega}\mathbf{p} \leq P_{\max}\mathbf{I}_M. \tag{E.45}$$

Substituting $\mathbf{p} \geq (\mathbf{I}-\mathbf{T})^{-1}\mathbf{s}$ obtained from (E.44) into (E.45), we have

$$\boldsymbol{\omega}(\mathbf{I} - \mathbf{T})^{-1}\mathbf{s} \leq P_{\max}\mathbf{I}_M. \tag{E.46}$$

This proves the sufficiency of (24).

For the necessity, assume that there is a vector $\tilde{\mathbf{p}} = [\tilde{p}_1, \cdots, \tilde{p}_K]$ satisfying

$$\gamma_k(\tilde{\mathbf{p}}) = \tilde{\gamma}_k \geq \underline{\gamma}_k, \tag{E.47}$$

and

$$\sum_{k=1}^{K}\omega_{m,k}\tilde{p}_k \leq P_{\max}. \tag{E.48}$$

From (E.47), we obtain

$$\rho_{\tilde{\mathbf{T}}} < 1, \tag{E.49}$$

where

$$[\tilde{\mathbf{T}}]_{k,j} = \begin{cases} 0, & j = k, \\ \dfrac{\mathcal{I}_k\tilde{\gamma}_k}{\mathcal{U}_k - \mathcal{V}_k\tilde{\gamma}_k}, & j \neq k. \end{cases} \tag{E.50}$$

Since $\tilde{\gamma}_k \geq \underline{\gamma}_k$, we have $\mathbf{T} \leq \tilde{\mathbf{T}}$, and thus, $\rho_{\mathbf{T}} \leq \rho_{\tilde{\mathbf{T}}} < 1$ [35]. From (E.48), we have

$$\boldsymbol{\omega}(\mathbf{I} - \tilde{\mathbf{T}})^{-1}\tilde{\mathbf{s}} \leq P_{\max}\mathbf{I}_M, \tag{E.51}$$

where $[\tilde{\mathbf{s}}]_k = \frac{\tilde{\gamma}_k\sigma_k^2}{\mathcal{U}_k - \tilde{\gamma}_k\mathcal{V}_k}$. Since $\tilde{\gamma}_k \geq \underline{\gamma}_k$, we have $\mathbf{s} \leq \tilde{\mathbf{s}}$. Combined with $\boldsymbol{\omega} \geq \mathbf{0}$, $\mathbf{s} \geq \mathbf{0}$, it is easy to obtain

$$\boldsymbol{\omega}(\mathbf{I} - \mathbf{T})^{-1}\mathbf{s} \leq \boldsymbol{\omega}(\mathbf{I} - \tilde{\mathbf{T}})^{-1}\tilde{\mathbf{s}} \leq P_{\max}\mathbf{I}_M. \tag{E.52}$$

This proves the necessary of (24).

APPENDIX F
PROOF OF PROPOSITION 2

Motivated by [27] and [36], we provide the following proof which is divided into two steps.

First, we prove the convergence of Algorithm 1. Recall that $\tilde{p}_k = \log_2 p_k$, we have the following chain inequalities after the i-th iteration of Algorithm 1,

$$\bar{\phi}\big(\mathbf{p}^{(i)}\big) \overset{(a)}{=} \tilde{\phi}\big(\tilde{\mathbf{p}}^{(i)}\big) \overset{(b)}{\leq} \tilde{\phi}\big(\tilde{\mathbf{p}}^{(i+1)}\big) \overset{(c)}{\leq} \bar{\phi}\big(\mathbf{p}^{(i+1)}\big), \tag{F.53}$$

where (a) is because the both sides of (28) are equal with given $\tilde{\gamma}_k^{(}i)$, (b) results from that $\tilde{\mathbf{p}}_k^{(i+1)}$ is the global maximizer of $\tilde{\phi}$ since the problem $(\tilde{\mathcal{P}}_1)$ is convex, (c) is because $\tilde{\phi}$ is the lower bound of $\bar{\phi}$ from (28). Consequently, Algorithm 1 must converge since the value of $\bar{\phi}$ increases after each iteration and is upper-bounded by the constraints of the problem $(\tilde{\mathcal{P}}_1)$.

Second, we prove that Algorithm 1 converges to the KKT points of $(\mathcal{P}_1)$. Denoting by $\tilde{\mathbf{p}}^*$ the power allocation vector at convergence which fulfills the KKT conditions of $(\tilde{\mathcal{P}}_1)$, we have $\tilde{\phi}(\tilde{\mathbf{p}}^*) = \bar{\phi}(\mathbf{p}^*)$ and $\nabla\tilde{\phi}(\tilde{\mathbf{p}}^*) = \nabla\bar{\phi}(\mathbf{p}^*)$ which results from (28). Moreover, the optimization problem $(\mathcal{P}_1)$ has the same constraints as $(\tilde{\mathcal{P}}_1)$. Hence, $\mathbf{p}^*$ must fulfills the KKT conditions of $(\mathcal{P}_1)$.

This completes the proof.

REFERENCES

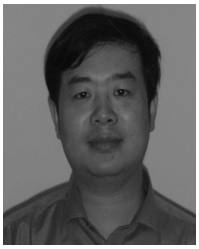
[1] S. Buzzi, C.-L. I, T. E. Klein, H. V. Poor, C. Yang, and A. Zappone, "A survey of energy-efficient techniques for 5G networks and challenges ahead," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 697–709, Apr. 2016.
[2] H. Zhu, "Performance comparison between distributed antenna and microcellular systems," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 6, pp. 1151–1163, Jun. 2011.
[3] L. Dai, "A comparative study on uplink sum capacity with co-located and distributed antennas," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 6, pp. 1200–1213, Jun. 2011.
[4] X. You, D. Wang, P. Zhu, and B. Sheng, "Cell edge performance of cellular mobile systems," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 6, pp. 1139–1150, Jun. 2011.
[5] J. Wang, H. Zhu, and N. J. Gomes, "Distributed antenna systems for mobile communications in high speed trains," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 4, pp. 675–683, May 2012.
[6] L. Dai, "An uplink capacity analysis of the distributed antenna system (DAS): From cellular DAS to DAS with virtual cells," *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2717–2731, May 2014.
[7] J. Wang and L. Dai, "Downlink rate analysis for virtual-cell based large-scale distributed antenna systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 1998–2011, Mar. 2016.
[8] J. Li, D. Wang, P. Zhu, and X. You, "Spectral efficiency analysis of large-scale distributed antenna system in a composite correlated Rayleigh fading channel," *IET Commun.*, vol. 9, no. 5, pp. 681–688, Apr. 2015.
[9] H. Zhu and J. Wang, "Radio resource allocation in multiuser distributed antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 10, pp. 2058–2066, Oct. 2013.
[10] J. Li, D. Wang, P. Zhu, J. Wang, and X. You, "Downlink spectral efficiency of distributed massive MIMO systems with linear beamforming under pilot contamination," *IEEE Trans. Veh. Technol.*, vol. 67, no. 2, pp. 1130–1145, Feb. 2018.

[11] D. Wang, J. Wang, X. You, Y. Wang, M. Chen, and X. Hou, "Spectral efficiency of distributed MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 10, pp. 2112–2127, Oct. 2013.
[12] H. Zhu and J. Wang, "Chunk-based resource allocation in OFDMA systems—Part II: Joint chunk, power and bit allocation," *IEEE Trans. Commun.*, vol. 60, no. 2, pp. 499–509, Feb. 2012.
[13] D. Wang, Y. Zhang, H. Wei, X. You, X. Gao, and J. Wang, "An overview of transmission theory and techniques of large-scale antenna systems for 5G wireless communications," *Sci. China Inf. Sci.*, vol. 59, no. 8, p. 081301, Aug. 2016.
[14] B. Dai and W. Yu, "Energy efficiency of downlink transmission strategies for cloud radio access networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 1037–1050, Apr. 2016.
[15] J. Joung, Y. K. Chia, and S. Sun, "Energy-efficient, large-scale distributed-antenna system (L-DAS) for multiple users," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 954–965, Oct. 2014.
[16] D. W. K. Ng and R. Schober, "Secure and green SWIPT in distributed antenna networks with limited backhaul capacity," *IEEE Trans. Wireless Commun.*, vol. 14, no. 9, pp. 5082–5097, Sep. 2015.
[17] C. He, B. Sheng, P. Zhu, X. You, and G. Li, "Energy- and spectral-efficiency tradeoff for distributed antenna systems with proportional fairness," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 5, pp. 894–902, May 2013.
[18] T. Van Chien, E. Björnson, and E. G. Larsson, "Joint power allocation and user association optimization for massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 6384–6399, Sep. 2016.
[19] J. Zuo, J. Zhang, C. Yuen, W. Jiang, and W. Luo, "Energy-efficient downlink transmission for multicell massive DAS with pilot contamination," *IEEE Trans. Veh. Technol.*, vol. 66, no. 2, pp. 1209–1221, Feb. 2017.
[20] J. Zuo, J. Zhang, C. Yuen, W. Jiang, and W. Luo, "Energy efficient user association for cloud radio access networks," *IEEE Access*, vol. 4, pp. 2429–2438, 2016.
[21] H. Q. Ngo, L.-N. Tran, T. Q. Duong, M. Matthaiou, and E. G. Larsson, "On the total energy efficiency of cell-free massive MIMO," *IEEE Trans. Green Commun. Netw.*, vol. 2, no. 1, pp. 25–39, Mar. 2018.
[22] L. D. Nguyen, T. Q. Duong, H. Q. Ngo, and K. Tourki, "Energy efficiency in cell-free massive MIMO with zero-forcing precoding design," *IEEE Commun. Lett.*, vol. 21, no. 8, pp. 1871–1874, Aug. 2017.
[23] E. Nayebi, A. Ashikhmin, T. L. Marzetta, H. Yang, and B. D. Rao, "Precoding and power optimization in cell-free massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4445–4459, Jul. 2017.
[24] 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.
[25] J. Li, D. Wang, P. Zhu, and X. You, "Downlink spectral efficiency of multi-cell multi-user large-scale DAS with pilot contamination," in *Proc. IEEE Int. Conf. Commun. (ICC)*, London, U.K., Jun. 2015, pp. 2011–2016.
[26] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.
[27] A. Zappone, L. Sanguinetti, G. Bacci, E. Jorswieck, and M. Debbah, "Energy-efficient power control: A look at 5G wireless technologies," *IEEE Trans. Signal Process.*, vol. 64, no. 7, pp. 1668–1683, Apr. 2016.
[28] J. Papandriopoulos and J. S. Evans, "Low-complexity distributed algorithms for spectrum balancing in multi-user DSL networks," in *Proc. IEEE Int. Conf. Commun.*, Istanbul, Turkey, Jun. 2006, pp. 3270–3275.
[29] W. Dinkelbach, "On nonlinear fractional programming," *Manage. Sci.*, vol. 13, no. 7, pp. 492–498, Mar. 1967.
[30] H. Ren, N. Liu, C. Pan, and C. He, "Energy efficiency optimization for MIMO distributed antenna systems," *IEEE Trans. Veh. Technol.*, vol. 66, no. 3, pp. 2276–2288, Mar. 2017.
[31] J. Li, D. Wang, P. Zhu, and X. You, "Uplink spectral efficiency analysis of distributed massive MIMO with channel impairments," *IEEE Access*, vol. 5, pp. 5020–5030, 2017.
[32] A. Khansefid and H. Minn, "Achievable downlink rates of MRC and ZF precoders in massive MIMO with uplink and downlink pilot contamination," *IEEE Trans. Commun.*, vol. 63, no. 12, pp. 4849–4864, Dec. 2015.
[33] H. Huh, A. M. Tulino, and G. Caire, "Network MIMO with linear zero-forcing beamforming: Large system analysis, impact of channel estimation, and reduced-complexity scheduling," *IEEE Trans. Inf. Theory*, vol. 14, no. 9, pp. 2911–2934, May 2015.
[34] A. M. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications* (Foundations and Trends in Communications and Information Theory). Boston, MA, USA: Now, 2004.
[35] G. Bacci, E. V. Belmega, and L. Sanguinetti, "Distributed energy-efficient power optimization in cellular relay networks with minimum rate constraints," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Florence, Italy, Jul. 2014, pp. 7014–7018.
[36] F. Tan, T. Lv, and S. Yang, "Power allocation optimization for energy-efficient massive MIMO aided multi-pair decode-and-forward relay systems," *IEEE Trans. Commun.*, vol. 65, no. 6, pp. 2368–2381, Jun. 2017.

JIAMIN LI received the B.S. and M.S. degrees in communication and information systems from Hohai University, Nanjing, China, in 2006 and 2009, respectively, and the Ph.D. degree in information and communication engineering from Southeast University, Nanjing, in 2014. He has been a Lecturer with the National Mobile Communications Research Laboratory, Southeast University, since 2014. His research interests include massive MIMO, distributed antenna systems, and cooperative communications.

DONGMING WANG received the B.S. degree from the Chongqing University of Posts and Telecommunications, Chongqing, China, the M.S. degree from the Nanjing University of Posts and Telecommunications, Nanjing, China, and the Ph.D. degree from Southeast University, Nanjing, in 1999, 2002, and 2006, respectively. He joined the National Mobile Communications Research Laboratory, Southeast University, in 2006, where he has been an Associate Professor, since 2010. His research interests include turbo detection, channel estimation, distributed antenna systems, and large-scale MIMO systems.

PENGCHENG ZHU received the B.S and M.S. degrees in electrical engineering from Shandong University, Jinan, China, in 2001 and 2004, respectively, and the Ph.D. degree in communication and information science from Southeast University, Nanjing, China, in 2009. He has been a Lecturer with the National Mobile Communications Research Laboratory, Southeast University, China, since 2009. His research interests lie in the areas of communication and signal processing, including limited feedback techniques, and distributed antenna systems.



XIAOHU YOU received the B.S., M.S., and Ph.D. degrees in electrical engineering from the Nanjing Institute of Technology, Nanjing, China, in 1982, 1985, and 1989, respectively. From 1987 to 1989, he was with the Nanjing Institute of Technology as a Lecturer. Since 1990, he has been with Southeast University, as an Associate Professor and then as a Professor. His research interests include mobile communications, adaptive signal processing, and artificial neural networks with applications to communications and biomedical engineering. He is the Chief of the Technical Group of China 3G/B3G Mobile Communication R & D Project. He received the Excellent Paper Prize from the China Institute of Communications, in 1987, and the Elite Outstanding Young Teacher Awards from Southeast University, in 1990, 1991, and 1993. He was also a recipient of the 1989 Young Teacher Award of Fok Ying Tung Education Foundation and the State Education Commission of China.