

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): You, Li; Huang, Yufei; Zhang, Di; Chang, Zheng; Wang, Wenjin; Gao, Xiqi

Title: Energy Efficiency Optimization for Multi-cell Massive MIMO : Centralized and Distributed Power Allocation Algorithms

Year: 2021

Version: Accepted version (Final draft)

Copyright: © 2021, IEEE

Rights: In Copyright

Rights url: <http://rightsstatements.org/page/InC/1.0/?language=en>

Please cite the original version:

You, L., Huang, Y., Zhang, D., Chang, Z., Wang, W., & Gao, X. (2021). Energy Efficiency Optimization for Multi-cell Massive MIMO : Centralized and Distributed Power Allocation Algorithms. *IEEE Transactions on Communications*, 69(8), 5228-5242.
<https://doi.org/10.1109/TCOMM.2021.3081451>

Energy Efficiency Optimization for Multi-cell Massive MIMO: Centralized and Distributed Power Allocation Algorithms

Li You, Yufei Huang, Di Zhang, Zheng Chang, Wenjin Wang, and Xiqi Gao

Abstract—This paper investigates the energy efficiency (EE) optimization in downlink multi-cell massive multiple-input multiple-output (MIMO). In our research, the statistical channel state information (CSI) is exploited to reduce the signaling overhead. To maximize the minimum EE among the neighbouring cells, we design the transmit covariance matrices for each base station (BS). Specifically, optimization schemes for this max-min EE problem are developed, in the centralized and distributed ways, respectively. To obtain the transmit covariance matrices, we first find out the closed-form optimal transmit eigenmatrices for the BS in each cell, and convert the original transmit covariance matrices designing problem into a power allocation one. Then, to lower the computational complexity, we utilize an asymptotic approximation expression for the problem objective. Moreover, for the power allocation design, we adopt the minorization maximization method to address the non-convexity of the ergodic rate, and use Dinkelbach’s transform to convert the max-min fractional problem into a series of convex optimization subproblems. To tackle the transformed subproblems, we propose a centralized iterative water-filling scheme. For reducing the backhaul burden, we further develop a distributed algorithm for the power allocation problem, which requires limited inter-cell information sharing. Finally, the performance of the proposed algorithms are demonstrated by extensive numerical results.

Index Terms—Energy efficiency, statistical CSI, multi-cell MIMO, max-min fairness, distributed processing.

I. INTRODUCTION

AS the communication industry develops at an extremely fast speed, we will witness various emerging applications in every aspect of our life, such as agriculture, traffic, health care, and so on [2]. Consequently, the current wireless networks are facing new challenges to fulfill the demands of these new technologies. For example, the form of information shared by people is changing from texts, images to high-definition videos, which brings new requirements on the speed and capacity of wireless communications. In order to cope with this trend, the cellular network needs to be evolved, i.e.,

Part of this work was presented at the IEEE International Conference on Communications in China, Chongqing, China, Aug. 2020 [1].

Li You, Yufei Huang, Wenjin Wang, and Xiqi Gao are with the National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China, and also with the Purple Mountain Laboratories, Nanjing 211100, China (e-mail: liyou@seu.edu.cn; yufei_huang@seu.edu.cn; wangwj@seu.edu.cn; xqgao@seu.edu.cn).

Di Zhang is with the School of Information Engineering, Zhengzhou University, Zhengzhou 450001, China (e-mail: dr.di.zhang@ieee.org).

Zheng Chang is with the School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China, and also with Faculty of Information Technology, University of Jyväskylä, FI-40014 Jyväskylä, Finland (email: zheng.chang@jyu.fi).

from the existing 5G to beyond 5G (B5G) and even 6G [3], [4]. As a key technology in 5G, massive multiple-input multiple-output (MIMO) is maturing and has been widely adopted. It has the merit of bringing about high energy efficiency (EE) and spectral efficiency (SE) gains [5], and its variant (e.g., cell-free networks, extremely large aperture arrays) will still play an indispensable role in the forthcoming B5G or even 6G eras [3], [6].

In conventional single-cell massive MIMO processing cellular networks, one major obstacle limiting the capacity resides in the inter-cell interference (ICI) resulting from pilot contamination [7]. The cell-edge user terminals (UTs) tend to undergo severe ICI and thus have poor transmission performance. As a solution, the base station (BS) cooperation, also known as the coordinated multi-cell precoding, was proposed [8]. The main idea of this procedure is that adjacent cells form a cooperative cluster where BSs exchange information with each other. It can mitigate the co-channel interference and reach a preferable data rate, and has attracted extensive investigations [8]–[11]. In particular, an overview of the multi-cell MIMO cooperation was shown in [8]. Additionally, downlink beamforming vectors of a multi-cell network were designed in [9] considering two different designing criteria, that is, minimizing the downlink transmit power and maximizing the signal-to-interference-plus-noise ratio. In [10] and [11], joint power allocation approaches in a coordinated multi-cell downlink system were proposed taking SE and EE as the optimization objectives, respectively.

Note that in downlink multi-cell MIMO, if the full transmit signals and CSI are shared among cells, the multi-cell network can be equivalently considered as a single cell multi-user system, where the UTs are simultaneously served by a cluster of BSs [12], [13]. Nevertheless, this kind of cooperation will result in an enormous backhaul burden because of the vast data sharing. On the contrary, distributed schemes with limited inter-cell information sharing are more favorable, for they require limited backhaul capacity. To this end, in [14], a distributed beamforming strategy was developed by recasting the downlink beamformer, and the transmit beamforming design was converted to a linear minimum mean square error estimation problem. A distributed EE-oriented transmission design was raised in [15]. By formulating the problem as a non-cooperative game, the research shows that the game reaches a Nash equilibrium and the problem can be tackled in a totally distributed way.

Another critical issue existing in the current massive MIMO

system is the energy consumption. Though the system SE is treated as a more important designing metric in tradition, with the fast growing number of UTs, the power consumption could be remarkably increased. Due to the ecological and economic considerations, EE-oriented investigations have emerged lately [16]–[19]. When it comes to energy-efficient precoding design in multi-cell scenario, there are several optimization objectives, such as the global EE, which is denoted as the ratio of all users’ sum rate to the total consumed power, or the sum of individual EE of each cell. For instance, energy-efficient power allocation for multi-cell massive MIMO transmission was investigated in [11]. A coordinated multi-cell multi-user precoding scheme was proposed in [20], aiming to maximize the weighted sum EE. However, the weighting factors are usually not easy to be adjusted in practice. Moreover, the fairness cannot be guaranteed when different BSs have different transmit power levels [21]. Therefore, we tend to adopt the max-min weighted EE as a design criterion to guarantee the minimum EE performance [22], [23].

Most of the aforementioned EE-based transmission approaches rely on perfectly known instantaneous CSI. However, the acquisition error of instantaneous CSI is usually unavoidable in practice, especially in massive MIMO [24]. Robust transmission designs exploiting imperfect instantaneous CSI were studied in e.g., [25], [26]. On the contrary, the statistical CSI, such as the channel covariance matrix, which reflects the statistical properties of the channels, can be obtained more easily. Moreover, it varies slowly over a much broader time span, which is preferable for power allocation design. Some previous works studied massive MIMO transmission based on statistical CSI [27]–[30]. For instance, in [27], an energy-efficient precoding procedure in the beam domain under the multicast massive MIMO scenario was studied. In [28], the authors extended the EE optimizing question to non-orthogonal unicast and multicast transmission. The resource efficiency optimization transmission strategies were proposed to strike an adaptive EE-SE balance in [29], [30].

Inspired by the above considerations, we aim to investigate the max-min fairness-based EE optimization problem in multi-cell massive MIMO systems exploiting the statistical CSI. The objective function of our optimization problem is the minimum weighted EE among cells. In general, the considered problem is challenging as the objective is non-convex and exhibits a fractional form. The major contributions of our work are summarized as follows:

- We first formulate the max-min EE optimization problem to design the transmit covariance matrices at each BS. By exploiting the channel properties in massive MIMO, we obtain the optimal transmit directions at each BS, thus converting the original complex-matrix-valued precoding design problem into a real-vector-valued beam domain power allocation one.
- We derive an asymptotic approximation, i.e., the deterministic equivalent (DE), of the ergodic user rate. By doing so, we avoid the complicated expectation calculation in the rate expression. We further handle this optimization problem by solving a sequence of convex optimization subproblems based on the minorization maximization

(MM) technique and Dinkelbach’s transform.

- To address the subproblems, we propose two approaches, i.e., the centralized and distributed ones. For each approach, we put forward generalized water-filling schemes to maximize the Lagrangian function.
- Combining all the methods utilized above, we propose low complexity iterative power allocation algorithms with guaranteed convergence for the multi-cell max-min fairness-based EE precoding design. Numerical results illustrate the performance of the proposed algorithms.

The rest of this paper is organized as follows. In Section II, we introduce the considered multi-cell multi-user massive MIMO downlink system. The characteristic of the adopted channel model is demonstrated, and the max-min EE problem is formulated. In Section III, the eigenvectors of the transmit power matrices are determined in closed-form, and the MM method and Dinkelbach’s transform are applied to address the max-min fractional power allocation design. In Section IV, the centralized and distributed algorithms to solve the transformed convex subproblems are described, respectively. We also analyze and compare the complexity of the two approaches. Numerical simulations are conducted in Section V. We conclude this paper in Section VI.

Notations: Throughout this paper, matrices and column vectors are represented by upper and lower case boldface letters, respectively. We adopt $\mathbb{C}^{M \times N}$ to denote $M \times N$ complex-valued vector space and $\mathbb{R}^{M \times N}$ to represent $M \times N$ real-valued one. Moreover, $\mathbf{A} \succeq \mathbf{0}$ indicates a positive semi-definite matrix \mathbf{A} , \mathbf{I}_M represents an identity matrix of size $M \times M$. The notation $\text{diag}\{\mathbf{x}\}$ creates a diagonal matrix with \mathbf{x} along its main diagonal, $[x]^+$ represents $\max\{x, 0\}$. The superscripts $(\cdot)^T$, $(\cdot)^*$ and $(\cdot)^H$ stand for transpose, conjugate and conjugate-transpose operations, respectively. The operators are defined by: $\text{tr}\{\cdot\}$ as the trace operation, $\det\{\cdot\}$ as the determinant operation, $\mathbf{E}\{\cdot\}$ as the expectation operation, \odot as the Hadamard product, and \triangleq for definition, respectively.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. System Model

Consider a U -cell massive MIMO downlink system, where each cell u contains one M -antenna BS which simultaneously serves K_u multi-antenna UTs. Define the sets for all the cells and UTs as $\mathcal{U} \triangleq \{1, \dots, U\}$ and $\mathcal{K} \triangleq \{(k, u) | u = 1, \dots, U, k = 1, \dots, K_u\}$, respectively. Denote the BS in cell u and the k th UT in cell u as BS- u and UT- (k, u) , respectively. Each UT- (k, u) has $N_{k,u}$ receive antennas. With $\mathbf{x}_{k,u} \in \mathbb{C}^{M \times 1}$ and $\mathbf{H}_{k,u,v} \in \mathbb{C}^{N_{k,u} \times M}$ denoting the signal transmitted to UT- (k, u) and the downlink channel matrix from BS- v to UT- (k, u) , respectively, the downlink transmission model is expressed as

$$\mathbf{y}_{k,u} = \mathbf{H}_{k,u,u} \mathbf{x}_{k,u} + \sum_{(i,j) \neq (k,u)} \mathbf{H}_{k,u,j} \mathbf{x}_{i,j} + \mathbf{n}_{k,u}, \quad (1)$$

where $\mathbf{y}_{k,u} \in \mathbb{C}^{N_{k,u} \times 1}$ represents the received signal at UT- (k, u) under an additive Gaussian noise $\mathbf{n}_{k,u} \in \mathbb{C}^{N_{k,u} \times 1}$, which is complex circularly symmetric distributed, zero-mean, and has a covariance matrix $\sigma^2 \mathbf{I}_{N_{k,u}}$. The transmitted vector

$\mathbf{x}_{k,u} (\forall (k,u) \in \mathcal{K})$ satisfies $\mathbb{E}\{\mathbf{x}_{k,u}\} = \mathbf{0}$, $\mathbb{E}\{\mathbf{x}_{k,u}\mathbf{x}_{i,j}^H\} = \mathbf{0} (\forall (i,j) \neq (k,u))$, and its covariance matrix is $\mathbf{Q}_{k,u} = \mathbb{E}\{\mathbf{x}_{k,u}\mathbf{x}_{k,u}^H\} \in \mathbb{C}^{M \times M}$.

In our work, we adopt the widely accepted Weichselberger's channel model [31]. Then, the downlink channel matrix $\mathbf{H}_{k,u,v}$ is characterized as [32]

$$\mathbf{H}_{k,u,v} = \mathbf{U}_{k,u,v} \mathbf{G}_{k,u,v} \mathbf{V}_{k,u,v}^H, \quad (2)$$

where $\mathbf{G}_{k,u,v} \in \mathbb{C}^{N_{k,u} \times M}$ is a random matrix and referred to as the beam domain channel [33], $\mathbf{U}_{k,u,v} \in \mathbb{C}^{N_{k,u} \times N_{k,u}}$ and $\mathbf{V}_{k,u,v} \in \mathbb{C}^{M \times M}$ are deterministic unitary matrices, denoting the eigenvector matrices of the UT and the BS correlation matrix, respectively. Note that the entries in $\mathbf{G}_{k,u,v}$ are statistically uncorrelated [32]. The structure of the downlink MIMO channel model shown in (2) describes the downlink channel spatial correlations. It is proved in [24] that, as $M \rightarrow \infty$, $\mathbf{V}_{k,u,v}$ becomes asymptotically identical, i.e.,

$$\mathbf{V}_{k,u,v} \stackrel{M \rightarrow \infty}{=} \mathbf{V}, \quad (3)$$

where \mathbf{V} only relates to the BS antenna array topologies and is irrelevant to the locations of UTs. For instance, in the case of uniform linear array (ULA) antenna configurations employed at the BS, the discrete Fourier transform (DFT) matrix is a good approximation of \mathbf{V} [24], [34]. Note that for practical cases where M is large and finite, \mathbf{V} can be seen as a good approximation to $\mathbf{V}_{k,u,v}$ [35].

In this work, we design the precoding problem based on the statistical CSI instead of instantaneous CSI. Specifically, we adopt the concept of eigenmode channel coupling matrix in [32], which is defined as

$$\mathbf{\Omega}_{k,u,v} = \mathbb{E}\{\mathbf{G}_{k,u,v} \odot \mathbf{G}_{k,u,v}^*\} \in \mathbb{R}^{N_{k,u} \times M}. \quad (4)$$

Throughout the paper, we assume that only the statistical CSI $\mathbf{\Omega}_{k,u,v}$ is accessible to each BS- v while each UT- (k,u) has access to its own instantaneous CSI through proper pilot design [36].

B. Problem Formulation

We treat $\mathbf{n}'_{k,u} = \sum_{(i,j) \neq (k,u)} \mathbf{H}_{k,u,j} \mathbf{x}_{i,j} + \mathbf{n}_{k,u}$ as the aggregate interference-plus-noise, whose covariance matrix $\mathbf{K}_{k,u} \in \mathbb{C}^{N_{k,u} \times N_{k,u}}$ is also accessible at each UT- (k,u) . For a worst-case design¹, $\mathbf{n}'_{k,u}$ is supposed to be a Gaussian noise with covariance [37]

$$\mathbf{K}_{k,u} = \sigma^2 \mathbf{I}_{N_{k,u}} + \sum_{(i,j) \neq (k,u)} \mathbb{E}\{\mathbf{H}_{k,u,j} \mathbf{Q}_{i,j} \mathbf{H}_{k,u,j}^H\}. \quad (5)$$

Under the above assumptions, we calculate the ergodic data rate of UT- (k,u) as in [38], [39], i.e.,

$$\begin{aligned} R_{k,u} &= \mathbb{E}\{\log \det(\mathbf{K}_{k,u} + \mathbf{H}_{k,u,u} \mathbf{Q}_{k,u} \mathbf{H}_{k,u,u}^H)\} \\ &\quad - \log \det(\mathbf{K}_{k,u}) \\ &= \mathbb{E}\left\{\log \det\left(\tilde{\mathbf{K}}_{k,u} + \mathbf{G}_{k,u,u} \mathbf{V}^H \mathbf{Q}_{k,u} \mathbf{V} \mathbf{G}_{k,u,u}^H\right)\right\} \end{aligned}$$

¹Note that the worst-case design means that for a fixed noise covariance, the worst case is that the noise follows a Gaussian distribution in the sense that the corresponding rate is a lower bound of the exact rate [37].

$$- \log \det(\tilde{\mathbf{K}}_{k,u}), \quad (6)$$

where the second equality can be obtained through first rewriting $\mathbf{H}_{k,u,u}$ by (2) and (3) as $\mathbf{H}_{k,u,u} = \mathbf{U}_{k,u,u} \mathbf{G}_{k,u,u} \mathbf{V}^H$ and further exploiting Sylvester's determinant identity, i.e., $\det(\mathbf{I} + \mathbf{WZ}) = \det(\mathbf{I} + \mathbf{ZW})$. Moreover, $\tilde{\mathbf{K}}_{k,u} \in \mathbb{C}^{N_{k,u} \times N_{k,u}}$ in (6) is expressed as

$$\begin{aligned} \tilde{\mathbf{K}}_{k,u} &\triangleq \mathbf{U}_{k,u,u}^H \mathbf{K}_{k,u} \mathbf{U}_{k,u,u} \\ &= \sigma^2 \mathbf{I}_{N_{k,u}} + \sum_{(i,j) \neq (k,u)} \underbrace{\mathbb{E}\{\mathbf{G}_{k,u,j} \mathbf{V}^H \mathbf{Q}_{i,j} \mathbf{V} \mathbf{G}_{k,u,j}^H\}}_{\triangleq \mathbf{\Pi}_{k,u,j}(\mathbf{V}^H \mathbf{Q}_{i,j} \mathbf{V})}. \end{aligned} \quad (7)$$

The matrix-valued function $\mathbf{\Pi}_{k,u,j}(\mathbf{X})$ defined in (7) can be proved diagonal by applying the uncorrelated properties of the elements in $\mathbf{G}_{k,u,v}$. Moreover, the n th diagonal element of $\mathbf{\Pi}_{k,u,j}(\mathbf{X})$ can be calculated as

$$[\mathbf{\Pi}_{k,u,j}(\mathbf{X})]_{n,n} = \text{tr}\left\{\text{diag}\left\{\left([\mathbf{\Omega}_{k,u,j}]_{n,:}\right)^T\right\} \mathbf{X}\right\}. \quad (8)$$

In this context, the EE of each cell u is defined as

$$\begin{aligned} \text{EE}_u &= \frac{WR_u}{P_u} \\ &= \frac{W \sum_{k=1}^{K_u} R_{k,u}}{\xi_u \sum_{k=1}^{K_u} \text{tr}\{\mathbf{Q}_{k,u}\} + MP_{c,u} + P_{s,u}}, \quad \forall u \in \mathcal{U}, \end{aligned} \quad (9)$$

where W represents the communication bandwidth, R_u is the sum rate of cell u and P_u is the power consumption required to operate cell u . Note that this power consumption model is widely adopted [16], [40]. Specifically, the constants $\xi_u (> 1)$, $P_{s,u}$, and $P_{c,u}$ are the inverse of the power amplifier efficiency, the dynamic power consumed per BS antenna and the basic power consumption of BS- u , respectively, and $\text{tr}\{\mathbf{Q}_{k,u}\}$ denotes the transmit power of UT- k in BS- u .

The main focus of our work is to investigate the multi-cell fairness-based EE problem. To this end, we maximize the minimum weighted EE among all cells, which is characterized as

$$\begin{aligned} \mathcal{P} : \quad &\max_{\mathbf{Q}_{k,u}, \forall (k,u)} \min_u \{w_u \text{EE}_u\} \\ &\text{s.t.} \quad \sum_{k=1}^{K_u} \text{tr}\{\mathbf{Q}_{k,u}\} \leq P_{\max,u}, \quad \forall u \in \mathcal{U} \\ &\quad \mathbf{Q}_{k,u} \succeq \mathbf{0}, \quad \forall (k,u) \in \mathcal{K}, \end{aligned} \quad (10)$$

where the weighting factor w_u indicates the priority of cell u and $P_{\max,u}$ is the maximum value of the transmit power of BS- u .

III. ENERGY EFFICIENT PRECODING DESIGN

In this section, we investigate the precoding design for the minimum weighted EE maximization problem \mathcal{P} in (10). Directly handling \mathcal{P} is of high complexity, since $\mathbf{Q}_{k,u}, \forall (k,u)$ is a high dimensional complex-valued matrix. Therefore, we start by decomposing the transmit covariance matrix as $\mathbf{Q}_{k,u} = \mathbf{\Psi}_{k,u} \mathbf{\Lambda}_{k,u} \mathbf{\Psi}_{k,u}^H$ based on the eigenvalue decomposition. By doing so, we derive the transmit directions of the signals, which are represented by the columns of $\mathbf{\Psi}_{k,u}$. In addition, the

power allocated to each transmit direction is represented by the diagonal elements of $\mathbf{\Lambda}_{k,u}$, i.e., the eigenvalues of $\mathbf{Q}_{k,u}$.

A. Optimal Transmit Direction

For the transmit eigenmatrix $\mathbf{\Psi}_{k,u}$, we apply the similar technique adopted in [17], [27], and obtain the following proposition.

Proposition 1: The optimal transmit eigenmatrix $\mathbf{\Psi}_{k,u}$ for problem \mathcal{P} in (10) is given as $\mathbf{\Psi}_{k,u}^{\text{opt}} = \mathbf{V}$, where \mathbf{V} is the transmit correlation matrix of the channel, and the transmit covariance matrices become

$$\mathbf{Q}_{k,u}^{\text{opt}} = \mathbf{V}\mathbf{\Lambda}_{k,u}\mathbf{V}^H, \forall (k,u) \in \mathcal{K}. \quad (11)$$

From *Proposition 1*, we can obtain the optimal transmit eigenmatrix $\mathbf{\Psi}_{k,u}$ in closed-form for problem \mathcal{P} in (10). Therefore, we can reduce the corresponding optimization complexity by converting the problem \mathcal{P} into a beam domain power allocation problem. With a slight abuse of notations, we define a matrix set $\mathbf{\Lambda} \triangleq \{\mathbf{\Lambda}_{1,1}, \dots, \mathbf{\Lambda}_{K_U,U}\}$, and the corresponding power allocation problem is formulated as

$$\begin{aligned} \mathcal{P}_1 : \max_{\mathbf{\Lambda}} \quad & \min_u \left\{ \frac{w_u W \sum_{k=1}^{K_u} R_{k,u}(\mathbf{\Lambda})}{\xi_u \sum_{k=1}^{K_u} \text{tr}\{\mathbf{\Lambda}_{k,u}\} + MP_{c,u} + P_{s,u}} \right\} \\ \text{s.t.} \quad & \sum_{k=1}^{K_u} \text{tr}\{\mathbf{\Lambda}_{k,u}\} \leq P_{\max,u}, \forall u \in \mathcal{U} \\ & \mathbf{\Lambda}_{k,u} \succeq \mathbf{0}, \mathbf{\Lambda}_{k,u} \text{ diagonal}, \forall (k,u) \in \mathcal{K}, \end{aligned} \quad (12)$$

where

$$\begin{aligned} R_{k,u}(\mathbf{\Lambda}) = & \mathbb{E} \left\{ \underbrace{\log \det \left(\tilde{\mathbf{K}}_{k,u}(\mathbf{\Lambda}) + \mathbf{G}_{k,u,u} \mathbf{\Lambda}_{k,u} \mathbf{G}_{k,u,u}^H \right)}_{\triangleq A_{k,u}(\mathbf{\Lambda})} \right. \\ & \left. - \underbrace{\log \det \left(\tilde{\mathbf{K}}_{k,u}(\mathbf{\Lambda}) \right)}_{\triangleq B_{k,u}(\mathbf{\Lambda})} \right\}. \end{aligned} \quad (13)$$

Note that $\tilde{\mathbf{K}}_{k,u}(\mathbf{\Lambda})$ in (13) is a matrix-valued function defined as

$$\tilde{\mathbf{K}}_{k,u}(\mathbf{\Lambda}) \triangleq \sigma^2 \mathbf{I}_{N_{k,u}} + \sum_{(i,j) \neq (k,u)} \mathbf{\Pi}_{k,u,j}(\mathbf{\Lambda}_{i,j}), \quad (14)$$

and the definition of $\mathbf{\Pi}_{k,u,j}(\mathbf{X})$ is given in (8).

Note that the number of variables in problem \mathcal{P}_1 , compared with problem \mathcal{P} , is reduced from $M^2 \times \left(\sum_{u=1}^U K_u \right)$ to $M \times \left(\sum_{u=1}^U K_u \right)$. This reduction is especially significant for a massive MIMO system with large M .

B. Power Allocation Design

Next, we focus on the optimization strategies for problem \mathcal{P}_1 . Note that the objective of \mathcal{P}_1 is the minimum of several fractional functions, where the denominator of each fractional function is linear over $\mathbf{\Lambda}$ and the numerator exhibits a non-concave form. In line with this, we put forward a power allocation strategy to maximize the minimum EE by means of fractional programming and iterative optimization.

Firstly, we can see that for each UT- (k,u) , direct calculation of the ergodic rate, $R_{k,u}(\mathbf{\Lambda})$, which involves the expectation operation on large-dimensional matrices, would incur high computational complexity. To obtain a tractable approximation for the ergodic rate, we apply the large-dimensional matrix theory [41], [42] and replace the ergodic rate with its DE expression. The main idea of DE is to introduce several auxiliary variables to iteratively calculate the object function, avoiding the high-dimensional expectation operation. Note that the DE expression is proved to be an asymptotically accurate approximation of the objective and especially suitable in massive MIMO systems [43]. Specially, the DE of $A_{k,u}(\mathbf{\Lambda})$ in (13) is computed by

$$\begin{aligned} \bar{A}_{k,u}(\mathbf{\Lambda}) = & \log \det(\mathbf{I}_M + \mathbf{\Gamma}_{k,u} \mathbf{\Lambda}_{k,u}) \\ & + \log \det \left(\tilde{\mathbf{\Gamma}}_{k,u} + \tilde{\mathbf{K}}_{k,u}(\mathbf{\Lambda}) \right) - \text{tr} \left\{ \mathbf{I}_{N_{k,u}} - \tilde{\mathbf{\Phi}}_{k,u}^{-1} \right\}, \end{aligned} \quad (15)$$

where the auxiliary variables are represented as

$$\mathbf{\Gamma}_{k,u} = \mathbf{T}_{k,u} \left(\tilde{\mathbf{\Phi}}_{k,u}^{-1} \left(\tilde{\mathbf{K}}_{k,u}(\mathbf{\Lambda}) \right)^{-1} \right), \quad (16)$$

$$\tilde{\mathbf{\Gamma}}_{k,u} = \mathbf{F}_{k,u} \left(\mathbf{\Phi}_{k,u}^{-1} \mathbf{\Lambda}_{k,u} \right), \quad (17)$$

$$\tilde{\mathbf{\Phi}}_{k,u} = \mathbf{I}_{N_{k,u}} + \tilde{\mathbf{\Gamma}}_{k,u} \left(\tilde{\mathbf{K}}_{k,u}(\mathbf{\Lambda}) \right)^{-1}, \quad (18)$$

$$\mathbf{\Phi}_{k,u} = \mathbf{I}_M + \mathbf{\Gamma}_{k,u} \mathbf{\Lambda}_{k,u}. \quad (19)$$

The matrix-valued functions $\mathbf{T}_{k,u}(\mathbf{X})$ and $\mathbf{F}_{k,u}(\mathbf{Y})$ are defined by $\mathbf{T}_{k,u}(\mathbf{X}) \triangleq \mathbb{E} \left\{ \mathbf{G}_{k,u,u}^H \mathbf{X} \mathbf{G}_{k,u,u} \right\} \in \mathbb{C}^{M \times M}$ and $\mathbf{F}_{k,u}(\mathbf{Y}) \triangleq \mathbb{E} \left\{ \mathbf{G}_{k,u,u} \mathbf{Y} \mathbf{G}_{k,u,u}^H \right\} \in \mathbb{C}^{N_{k,u} \times N_{k,u}}$, respectively. They are both diagonal functions since the elements of $\mathbf{G}_{k,u,u}$ are zero-mean and statistically uncorrelated, and the corresponding diagonal elements are given by

$$[\mathbf{T}_{k,u}(\mathbf{X})]_{m,m} = \text{tr} \left\{ \text{diag} \left\{ [\mathbf{\Omega}_{k,u,u}]_{:,m} \right\} \mathbf{X} \right\}, \quad (20)$$

$$[\mathbf{F}_{k,u}(\mathbf{Y})]_{n,n} = \text{tr} \left\{ \text{diag} \left\{ \left([\mathbf{\Omega}_{k,u,u}]_{n,:} \right)^T \right\} \mathbf{Y} \right\}, \quad (21)$$

respectively. Then, with the aid of $\bar{A}_{k,u}(\mathbf{\Lambda})$, we turn to consider the following problem

$$\begin{aligned} \mathcal{P}_2 : \max_{\mathbf{\Lambda}} \quad & \min_u \left\{ \frac{w_u W \sum_{k=1}^{K_u} \{ \bar{A}_{k,u}(\mathbf{\Lambda}) - B_{k,u}(\mathbf{\Lambda}) \}}{\xi_u \sum_{k=1}^{K_u} \text{tr}\{\mathbf{\Lambda}_{k,u}\} + MP_{c,u} + P_{s,u}} \right\} \\ \text{s.t.} \quad & \sum_{k=1}^{K_u} \text{tr}\{\mathbf{\Lambda}_{k,u}\} \leq P_{\max,u}, \forall u \in \mathcal{U} \\ & \mathbf{\Lambda}_{k,u} \succeq \mathbf{0}, \mathbf{\Lambda}_{k,u} \text{ diagonal}, \forall (k,u) \in \mathcal{K}. \end{aligned} \quad (22)$$

For iterative convex optimization methods, we resort to the MM procedure [44]. The main idea of MM procedure is to convert the original non-convex program into a series of solvable subproblems. Following the MM procedure, in each subproblem, we find a surrogate function to approximate the ergodic rate. Specially, we replace $B_{k,u}(\mathbf{\Lambda})$ in (13), the negative term of the rate, with its first-order Taylor expansion. Then, the numerator becomes a concave function of $\mathbf{\Lambda}$. Therefore, problem \mathcal{P}_2 is solved through iteratively solving

the subproblems as follows

$$\begin{aligned}
& \mathcal{P}_3^{(\ell)} : \mathbf{\Lambda}^{(\ell+1)} = \\
& \arg \max_{\mathbf{\Lambda}} \min_u \left\{ \frac{w_u W \sum_{k=1}^{K_u} \left\{ \bar{A}_{k,u}(\mathbf{\Lambda}) - \Delta B_{k,u}^{(\ell)}(\mathbf{\Lambda}) \right\}}{\xi_u \sum_{k=1}^{K_u} \text{tr} \{ \mathbf{\Lambda}_{k,u} \} + MP_{c,u} + P_{s,u}} \right\} \\
& \text{s.t.} \quad \sum_{k=1}^{K_u} \text{tr} \{ \mathbf{\Lambda}_{k,u} \} \leq P_{\max,u}, \quad \forall u \in \mathcal{U} \\
& \quad \mathbf{\Lambda}_{k,u} \succeq \mathbf{0}, \quad \mathbf{\Lambda}_{k,u} \text{ diagonal}, \quad \forall (k,u) \in \mathcal{K}, \quad (23)
\end{aligned}$$

where $\mathbf{\Lambda}^{(\ell)} \triangleq \{ \mathbf{\Lambda}_{1,1}^{(\ell)}, \dots, \mathbf{\Lambda}_{K_U,U}^{(\ell)} \}$ and ℓ denotes the iteration index. The first-order Taylor expansion term is expressed as

$$\begin{aligned}
\Delta B_{k,u}^{(\ell)}(\mathbf{\Lambda}) &= B_{k,u}(\mathbf{\Lambda}^{(\ell)}) \\
&+ \sum_{(i,j) \neq (k,u)} \text{tr} \left\{ \left(\frac{\partial B_{k,u}(\mathbf{\Lambda}^{(\ell)})}{\partial \mathbf{\Lambda}_{i,j}} \right)^T (\mathbf{\Lambda}_{i,j} - \mathbf{\Lambda}_{i,j}^{(\ell)}) \right\}. \quad (24)
\end{aligned}$$

Moreover, the derivative $\frac{\partial B_{k,u}(\mathbf{\Lambda}^{(\ell)})}{\partial \mathbf{\Lambda}_{i,j}}$ is derived as

$$\begin{aligned}
\mathbf{D}_{k,u,j}^{(\ell)} &= \frac{\partial B_{k,u}(\mathbf{\Lambda}^{(\ell)})}{\partial \mathbf{\Lambda}_{i,j}} \\
&= \sum_{n=1}^{N_{k,u}} \frac{\hat{\mathbf{R}}_{k,u,j,n}}{\sigma^2 + \text{tr} \{ \mathbf{\Lambda}_{\setminus(k,u)}^{(\ell)} \hat{\mathbf{R}}_{k,u,q,n} \}}, \quad (i,j) \neq (k,u), \quad (25)
\end{aligned}$$

where $\mathbf{\Lambda}_{\setminus(k,u)}^{(\ell)} = \sum_{(p,q) \neq (k,u)} \mathbf{\Lambda}_{p,q}^{(\ell)}$ and $\hat{\mathbf{R}}_{k,u,j,n} = \text{diag} \{ \omega_{k,u,j,n} \}$ with $\omega_{k,u,j,n}^T$ being the n th row of $\mathbf{\Omega}_{k,u,j}$. Note that $\mathbf{D}_{k,u,j}^{(\ell)}$ is a diagonal matrix with its t th diagonal entry given by

$$\begin{aligned}
& \left[\mathbf{D}_{k,u,j}^{(\ell)} \right]_{t,t} \\
&= \sum_{n=1}^{N_{k,u}} \frac{[\mathbf{\Omega}_{k,u,j}]_{n,t}}{\sigma^2 + \sum_{(p,q) \neq (k,u)} \sum_{m=1}^M [\mathbf{\Omega}_{k,u,q}]_{n,m} [\mathbf{\Lambda}_{p,q}^{(\ell)}]_{m,m}}. \quad (26)
\end{aligned}$$

Dinkelbach's transform can obtain the global optimal solution of a max-min concave-linear fractional programming by solving a sequence of convex problems [16], [45]. Specifically, for each max-min fractional problem $\mathcal{P}_3^{(\ell)}$ in (23), it can be equivalently handled by solving a series of concave maximization subproblems as follows

$$\begin{aligned}
& \mathcal{P}_4^{(\ell),[t]} : \mathbf{\Lambda}^{(\ell),[t+1]} = \\
& \arg \max_{\mathbf{\Lambda}} \min_u \left\{ w_u W \bar{R}_u^{(\ell)}(\mathbf{\Lambda}) - \eta^{(\ell),[t]} P_u(\mathbf{\Lambda}) \right\} \\
& \text{s.t.} \quad \sum_{k=1}^{K_u} \text{tr} \{ \mathbf{\Lambda}_{k,u} \} \leq P_{\max,u}, \quad \forall u \in \mathcal{U} \\
& \quad \mathbf{\Lambda}_{k,u} \succeq \mathbf{0}, \quad \mathbf{\Lambda}_{k,u} \text{ diagonal}, \quad \forall (k,u) \in \mathcal{K}, \quad (27)
\end{aligned}$$

where $\mathbf{\Lambda}^{(\ell),[t+1]} \triangleq \{ \mathbf{\Lambda}_{1,1}^{(\ell),[t+1]}, \dots, \mathbf{\Lambda}_{K_U,U}^{(\ell),[t+1]} \}$, t is the iteration

index of Dinkelbach's transform and

$$\bar{R}_u^{(\ell)}(\mathbf{\Lambda}) = \sum_{k=1}^{K_u} \left\{ \bar{A}_{k,u}(\mathbf{\Lambda}) - \Delta B_{k,u}^{(\ell)}(\mathbf{\Lambda}) \right\}, \quad (28)$$

$$P_u(\mathbf{\Lambda}) = \xi_u \sum_{k=1}^{K_u} \text{tr} \{ \mathbf{\Lambda}_{k,u} \} + MP_{c,u} + P_{s,u}, \quad (29)$$

$$\eta^{(\ell),[t]} = \min_u \left\{ \eta_u^{(\ell),[t]} \right\} = \min_u \left\{ \frac{w_u W \bar{R}_u^{(\ell)}(\mathbf{\Lambda}^{(\ell),[t]})}{P_u(\mathbf{\Lambda}^{(\ell),[t]})} \right\}. \quad (30)$$

Particularly, we summarize Dinkelbach's transform-based max-min EE maximization power allocation procedure in **Algorithm 1**.

Algorithm 1 Max-Min EE Power Allocation Algorithm

Input: Initial power allocation matrix $\mathbf{\Lambda}^{(0)}$, channel statistics

$\mathbf{\Omega}_{k,u,v}$, iteration thresholds ϵ_1, ϵ_2 .

Output: Power allocation matrix $\mathbf{\Lambda}$.

- 1: Initialization: $\ell = 0$, $\min_u \left\{ w_u \overline{\text{EE}}_u^{(-1)} \right\} = 0$.
- 2: For $\forall u \in \mathcal{U}$, calculate

$$\overline{\text{EE}}_u^{(\ell)} = \frac{W \sum_{k=1}^{K_u} \left\{ \bar{A}_{k,u}(\mathbf{\Lambda}^{(\ell)}) - \Delta B_{k,u}^{(\ell)}(\mathbf{\Lambda}^{(\ell)}) \right\}}{P_u(\mathbf{\Lambda}^{(\ell)})}. \quad (31)$$

- 3: **while** $\left| \min_u \left\{ w_u \overline{\text{EE}}_u^{(\ell)} \right\} - \min_u \left\{ w_u \overline{\text{EE}}_u^{(\ell-1)} \right\} \right| \geq \epsilon_1$ **do**
 - 4: Initialization: $t = 0$, $\mathbf{\Lambda}^{(\ell),[-1]} = \mathbf{\Lambda}^{(\ell)}$, $\eta^{(\ell),[-1]} = 0$.
 - 5: **while** $|\eta^{(\ell),[t]} - \eta^{(\ell),[t-1]}| \geq \epsilon_2$ **do**
 - 6: Set $t = t + 1$.
 - 7: Solve problem $\mathcal{P}_4^{(\ell),[t-1]}$ in (27) with $\eta^{(\ell),[t-1]}$ and obtain the solution $\mathbf{\Lambda}^{(\ell),[t]}$.
 - 8: Update $\eta^{(\ell),[t]}$ with (30).
 - 9: **end while**
 - 10: Set $\ell = \ell + 1$.
 - 11: Set $\mathbf{\Lambda}^{(\ell)} = \mathbf{\Lambda}^{(\ell-1),[t]}$.
 - 12: Update $\overline{\text{EE}}_u^{(\ell)}$ with (31).
 - 13: **end while**
 - 14: **return** $\mathbf{\Lambda} = \mathbf{\Lambda}^{(\ell)}$.
-

IV. ENERGY-EFFICIENT POWER ALLOCATION ALGORITHM

It is worth noting that each subproblem $\mathcal{P}_4^{(\ell),[t]}$ in (27) is convex and can be solved via utilizing the classical convex optimization methods [46]. However, directly solving it might lead to high computational complexity, especially in the massive MIMO scenario. We therefore focus on developing low-complexity algorithms for $\mathcal{P}_4^{(\ell),[t]}$ in this section.

Generally, power allocation design in the multi-cell scenario is conducted in a centralized way, where the BSs jointly design the power allocation matrices to suppress the ICI and reach a favorable performance. The centralized approach is feasible in many scenarios. However, in massive MIMO transmissions, the exchange of CSI between BSs might cause large backhaul

burdens. On the contrary, a distributed implementation of the power allocation scheme requiring only a few parameters is preferable to lighten the burden of the backhaul links. The distributed approach offers a trade-off between performance gains and the amount of overhead in the backhaul and feedback channels. In the following, we first introduce a centralized joint power allocation approach with full CSI sharing. Then, we propose a distributed scheme with limited inter-cell cooperation to solve the sub-optimal problem in (27).

A. Centralized Power Allocation

We first focus on the centralized algorithm for problem $\mathcal{P}_4^{(\ell),[t]}$ in (27). Noting that the objective function of problem $\mathcal{P}_4^{(\ell),[t]}$ is not continuously differentiable due to the minimum operation, we introduce a new variable z to represent the minimum weighted EE and arrive at an equivalent formulation of (27) as follows

$$\mathcal{P}_5^{(\ell),[t]} : \max_{\mathbf{\Lambda}, z} z \quad (32a)$$

$$\text{s.t. } z \leq w_u \left\{ W \bar{R}_u^{(\ell)}(\mathbf{\Lambda}) - \eta^{(\ell),[t]} P_u(\mathbf{\Lambda}) \right\}, \quad \forall u \in \mathcal{U} \quad (32b)$$

$$\sum_{k=1}^{K_u} \text{tr} \{ \mathbf{\Lambda}_{k,u} \} \leq P_{\max,u}, \quad \forall u \in \mathcal{U} \quad (32c)$$

$$\mathbf{\Lambda}_{k,u} \succeq \mathbf{0}, \quad \mathbf{\Lambda}_{k,u} \text{ diagonal}, \quad \forall (k, u) \in \mathcal{K}. \quad (32d)$$

In the following, we apply the dual method [47] to solve problem (32). Specifically, we calculate the dual function, i.e., the maximum of the Lagrangian while fixing the dual variables. Then, we apply the subgradient method to update the dual variables. The diminishing stepsize rule guarantees the convergence of the subgradient method [46]. Regarding the solution to (32), we have the following proposition.

Proposition 2: Denote the optimal power allocation matrices of problem $\mathcal{P}_5^{(\ell),[t]}$ in (32) as $\mathbf{\Lambda}_{k,u}^{(\ell),[t+1]}$, $\forall (k, u) \in \mathcal{K}$. Then, the m th diagonal element of $\mathbf{\Lambda}_{k,u}^{(\ell),[t+1]}$, i.e., $\lambda_{k,u,m}^{(\ell),[t+1]}$, satisfies (33), shown at the top of next page, where $d_{a,u,u,m}^{(\ell)}$ and $\gamma_{k,u,m}$ stand for the m th diagonal element of $\mathbf{D}_{a,u,u}^{(\ell)}$ and $\mathbf{\Gamma}_{k,u}$, respectively. Moreover, the auxiliary variables $\chi_{k,m}^{(\ell),[t+1]}$ and $\mathcal{T}_{k,m,a,u}$ are written as (34) and (35) shown at the top of next page, respectively.

Proof: See Appendix A.

The solutions in (33) suggest that the optimal power allocation matrices follow a water-filling structure. We summarize the centralized iterative generalized water-filling power allocation algorithm in **Algorithm 2**. In the multi-cell multi-user scenario, it is hard to find closed-form solutions to equation (33). In order to solve this problem, we utilize Newton's method [48], where the auxiliary functions $\varrho_{k,u,m}^{(\ell),[t]}(x_{k,u,m})$ and $\varrho_{k,u,m}^{\prime(\ell),[t]}(x_{k,u,m})$ in Step 8 of **Algorithm 2** are defined as (36) and (37) shown at the top of next page, respectively.

B. Distributed Power Allocation

In this subsection, considering the backhaul burden caused by the CSI sharing among the BSs in the centralized ap-

Algorithm 2 Centralized Water-Filling Power Allocation Algorithm

- 1: Initialization: $\mathbf{X}_{k,u}^0 = \mathbf{\Lambda}_{k,u}^{(\ell),[t]}$, $(k, u) \in \mathcal{K}$, and iteration index $q = 0$. $x_{k,u,m}^q$ stands for the m th diagonal entry of $\mathbf{X}_{k,u}^q$. Initialize $v = 0$ and $\beta_u^v = 1/K_u, \mu_u^v = 0, u = 1, \dots, U$. Set thresholds ϵ_3 and ϵ_4 .
 - 2: **repeat**
 - 3: **for** $u = 1$ to U **do**
 - 4: **for** $k = 1$ to K_u **do**
 - 5: **for** $m = 1$ to M **do**
 - 6: Set $q' = q$.
 - 7: **repeat**
 - 8: Calculate $\varrho_{k,u,m}^{(\ell),[t]}(x_{k,u,m}^{q'})$ and $\varrho_{k,u,m}^{\prime(\ell),[t]}(x_{k,u,m}^{q'})$ using (36) and (37), respectively.
 - 9: Update $x_{k,u,m}^{q'+1} = x_{k,u,m}^{q'} - \frac{\varrho_{k,u,m}^{(\ell),[t]}(x_{k,u,m}^{q'})}{\varrho_{k,u,m}^{\prime(\ell),[t]}(x_{k,u,m}^{q'})}$.
 - 10: Set $q' = q' + 1$.
 - 11: **until** $|x_{k,u,m}^{q'} - x_{k,u,m}^{q'-1}| \leq \epsilon_3$.
 - 12: **end for**
 - 13: **end for**
 - 14: **end for**
 - 15: Set $\bar{x}_{k,u,m} = [x_{k,u,m}^{q'}]^+$ and calculate $p_{\text{tot},u} = \sum_{k=1}^{K_u} \sum_{m=1}^M \bar{x}_{k,u,m}$, $u \in \mathcal{U}$.
 - 16: Update β^{v+1} and μ^{v+1} by the subgradient method.
 - 17: Set $v = v + 1$.
 - 18: **until** $\|\beta^v - \beta^{v-1}\| < \epsilon_4$ and $\|\mu^v - \mu^{v-1}\| \leq \epsilon_4$.
-

proach in some cases, we further investigate the distributed approach to find out the power allocation matrices for $\mathcal{P}_4^{(\ell),[t]}$ in (27). For notation convenience, denote the solution of the $(t-1)$ th subproblem, i.e., $\mathcal{P}_4^{(\ell),[t-1]}$, as $\mathbf{\Lambda}^{(\ell),[t]}$, and the set of power allocation matrices without cell u after the $(t-1)$ th iteration of Dinkelbach's transform as $\mathbf{\Lambda}_{-u}^{(\ell),[t]} \triangleq \left\{ \mathbf{\Lambda}_{p,q}^{(\ell),[t]} \mid \forall (p, q) \in \mathcal{K}, q \neq u \right\}$. In the distributed approach, each BS- u optimizes its own power allocation matrices using local CSI, and only a limited number of parameters are shared among cells. Specifically, at the t th iteration, each cell optimizes its own power allocation matrices utilizing the set of matrices $\mathbf{\Lambda}_{-u}^{(\ell),[t]}$.

Note that the objective function of $\mathcal{P}_4^{(\ell),[t]}$ is denoted by the minimum of a set of functions with respect to $\mathbf{\Lambda}$. When the power allocation matrices of other cells are set to be constant, i.e., represented by $\mathbf{\Lambda}_{-u}^{(\ell),[t]}$, the corresponding set of functions becomes

$$F_u \left(\left\{ \mathbf{\Lambda}_{k,u} \right\}_{k=1}^{K_u}, \mathbf{\Lambda}_{-u}^{(\ell),[t]} \right) \\ \triangleq w_u \bar{R}_u^{(\ell)} \left(\left\{ \mathbf{\Lambda}_{k,u} \right\}_{k=1}^{K_u}, \mathbf{\Lambda}_{-u}^{(\ell),[t]} \right) - \eta^{(\ell),[t]} P_u \left(\left\{ \mathbf{\Lambda}_{k,u} \right\}_{k=1}^{K_u} \right)$$

$$\left\{ \begin{array}{l} \frac{\beta_u w_u W \gamma_{k,u,m}}{1 + \gamma_{k,u,m} \lambda_{k,u,m}^{(\ell),[t+1]}} + \sum_{a \neq k} \sum_{n=1}^{N_{a,u}} \frac{\beta_u w_u W [\mathbf{\Omega}_{a,u,u}]_{n,m}}{\tilde{\gamma}_{a,u,n} + [\mathbf{\tilde{K}}_{a,u}(\mathbf{\Lambda})]_{n,n}} + \sum_{j \neq u} \sum_{a=1}^{K_j} \sum_{n=1}^{N_{a,j}} \frac{\beta_j w_j W [\mathbf{\Omega}_{a,j,u}]_{n,m}}{\tilde{\gamma}_{a,j,n} + [\mathbf{\tilde{K}}_{a,j}(\mathbf{\Lambda})]_{n,n}} = \eta^{(\ell),[t]} \xi_u + \mu_u^{[t+1]} \\ + \beta_u w_u W \sum_{a \neq k} d_{a,u,u,m}^{(\ell)} + \sum_{j \neq u} \beta_j w_j W \sum_{a=1}^{K_j} d_{a,j,u,m}^{(\ell)}, \quad \mu_u^{[t+1]} < \chi_{k,m}^{(\ell),[t+1]} \\ \lambda_{k,u,m}^{(\ell),[t+1]} = 0, \quad \mu_u^{[t+1]} \geq \chi_{k,m}^{(\ell),[t+1]} \end{array} \right. \quad (33)$$

$$\chi_{k,m}^{(\ell),[t+1]} = \gamma_{k,u,m} - \eta^{(\ell),[t]} \xi_u - \beta_u w_u W \sum_{a \neq k} d_{a,u,u,m}^{(\ell)} + \sum_{a \neq k} \sum_{n=1}^{N_{a,u}} \frac{\beta_u w_u W [\mathbf{\Omega}_{a,u,u}]_{n,m}}{\tilde{\gamma}_{a,u,n} + \sigma^2 + \sum_{(a',q,m') \in \mathcal{T}_{k,m,a,u}} \lambda_{a',q,m'}^{(\ell),[t+1]} [\mathbf{\Omega}_{a,u,u}]_{n,m'}} \\ - \sum_{j \neq u} \beta_j w_j W \sum_{a=1}^{K_j} d_{a,j,u,m}^{(\ell)} + \sum_{j \neq u} \sum_{a=1}^{K_j} \sum_{n=1}^{N_{a,j}} \frac{\beta_j w_j W [\mathbf{\Omega}_{a,j,u}]_{n,m}}{\tilde{\gamma}_{a,j,n} + \sigma^2 + \sum_{(a',q,m') \in \mathcal{T}_{k,m,a,j}} \lambda_{a',q,m'}^{(\ell),[t+1]} [\mathbf{\Omega}_{a,j,u}]_{n,m'}} \quad (34)$$

$$\mathcal{T}_{k,m,a,u} = \{(a',q,m') \mid (a',q) \neq (a,u), (a',q,m') \neq (k,u,m), (a',q,m') \in \mathcal{K}, m' \in \{1, \dots, M\}\} \quad (35)$$

$$\varrho_{k,u,m}^{(\ell),[t]}(x_{k,u,m}) = \frac{\beta_u^{[t]} w_u W \gamma_{k,u,m}}{1 + \gamma_{k,u,m} x_{k,u,m}} - \beta_u^{[t]} w_u W \sum_{a \neq k} d_{a,u,u,m}^{(\ell)} - \eta^{(\ell),[t]} \xi_u - \mu_u^{[t]} - \sum_{j \neq u} \beta_j^{[t]} w_j W \sum_{a=1}^{K_j} d_{a,j,u,m}^{(\ell)} \\ + \sum_{a \neq k} \sum_{n=1}^{N_{a,u}} \frac{\beta_u^{[t]} w_u W [\mathbf{\Omega}_{a,u,u}]_{n,m}}{\tilde{\gamma}_{a,u,n} + \sigma^2 + x_{k,u,m} [\mathbf{\Omega}_{a,u,u}]_{n,m} + \sum_{(a',q,m') \in \mathcal{T}_{k,m,a,u}} x_{a',q,m'} [\mathbf{\Omega}_{a,u,u}]_{n,m'}} \\ + \sum_{j \neq u} \sum_{a=1}^{K_j} \sum_{n=1}^{N_{a,j}} \frac{\beta_j^{[t]} w_j W [\mathbf{\Omega}_{a,j,u}]_{n,m}}{\tilde{\gamma}_{a,j,n} + \sigma^2 + x_{k,u,m} [\mathbf{\Omega}_{a,j,u}]_{n,m} + \sum_{(a',q,m') \in \mathcal{T}_{k,m,a,j}} x_{a',q,m'} [\mathbf{\Omega}_{a,j,u}]_{n,m'}} \quad (36)$$

$$\varrho_{k,u,m}'^{(\ell),[t]}(x_{k,u,m}) = - \frac{\beta_u^{[t]} w_u W (\gamma_{k,u,m})^2}{(1 + \gamma_{k,u,m} x_{k,u,m})^2} - \sum_{a \neq k} \sum_{n=1}^{N_{a,u}} \frac{\beta_u^{[t]} w_u W [\mathbf{\Omega}_{a,u,u}]_{n,m}^2}{\left(\tilde{\gamma}_{a,u,n} + \sigma^2 + x_{k,u,m} [\mathbf{\Omega}_{a,u,u}]_{n,m} + \sum_{(a',q,m') \in \mathcal{T}_{k,m,a,u}} x_{a',q,m'} [\mathbf{\Omega}_{a,u,u}]_{n,m'} \right)^2} \\ + \sum_{j \neq u} \sum_{a=1}^{K_j} \sum_{n=1}^{N_{a,j}} \frac{\beta_j^{[t]} w_j W [\mathbf{\Omega}_{a,j,u}]_{n,m}^2}{\left(\tilde{\gamma}_{a,j,n} + \sigma^2 + x_{k,u,m} [\mathbf{\Omega}_{a,j,u}]_{n,m} + \sum_{(a',q,m') \in \mathcal{T}_{k,m,a,j}} x_{a',q,m'} [\mathbf{\Omega}_{a,j,u}]_{n,m'} \right)^2} \quad (37)$$

$$= w_u \sum_{k=1}^{K_u} \left(\log \det (\mathbf{I}_M + \mathbf{\Gamma}_{k,u} \mathbf{\Lambda}_{k,u}) - \text{tr} \left\{ \mathbf{I}_{N_{k,u}} - \tilde{\mathbf{\Phi}}_{k,u}^{-1} \right\} \right) \\ + w_u \sum_{k=1}^{K_u} \log \det \left(\tilde{\mathbf{\Gamma}}_{k,u} + \tilde{\mathbf{K}}_{k,u} \left(\{\mathbf{\Lambda}_{k,u}\}_{k=1}^{K_u}, \mathbf{\Lambda}_{-u}^{(\ell),[t]} \right) \right) \\ - \eta^{(\ell),[t]} \left(\xi_u \sum_{k=1}^{K_u} \text{tr} \{ \mathbf{\Lambda}_{k,u} \} + MP_{c,u} + P_{s,u} \right) \\ - w_u \sum_{k=1}^{K_u} \Delta B_{k,u}^{(\ell)} \left(\{\mathbf{\Lambda}_{k,u}\}_{k=1}^{K_u}, \mathbf{\Lambda}_{-u}^{(\ell),[t]} \right), \quad \forall u \in U, \quad (38)$$

where $F_u \left(\{\mathbf{\Lambda}_{k,u}\}_{k=1}^{K_u}, \mathbf{\Lambda}_{-u}^{(\ell),[t]} \right)$ is defined for notation convenience. Note that $F_u \left(\{\mathbf{\Lambda}_{k,u}\}_{k=1}^{K_u}, \mathbf{\Lambda}_{-u}^{(\ell),[t]} \right)$ is a concave

function over each $\mathbf{\Lambda}_{k,u}$, $k = 1, \dots, K_u$ for given $\mathbf{\Lambda}_{-u}^{(\ell),[t]}$. In addition, denote the optimal solution of problem $\mathcal{P}_4^{(\ell),[t]}$ in (27) as $\mathbf{\Lambda}^{(\ell),[t+1]}$, we can show that the objective functions of problem $\mathcal{P}_4^{(\ell),[t]}$ tend to be identical [22], i.e.,

$$w_1 W \bar{R}_1^{(\ell)} \left(\mathbf{\Lambda}^{(\ell),[t+1]} \right) - \eta^{(\ell),[t]} P_1 \left(\mathbf{\Lambda}^{(\ell),[t+1]} \right) \\ = w_u W \bar{R}_u^{(\ell)} \left(\mathbf{\Lambda}^{(\ell),[t+1]} \right) - \eta^{(\ell),[t]} P_u \left(\mathbf{\Lambda}^{(\ell),[t+1]} \right), \quad u = 2, 3, \dots, U. \quad (39)$$

Based on this property, we can develop a distributed algorithm where each BS designs its own power allocation matrices individually while sharing limited inter-cell information. For

cell u , the power allocation problem is cast as

$$\mathcal{P}_{6,u}^{(\ell),[t]} : \max_{\{\mathbf{\Lambda}_{k,u}\}_{k=1}^{K_u}} F_u \left(\{\mathbf{\Lambda}_{k,u}\}_{k=1}^{K_u}, \mathbf{\Lambda}_{-u}^{(\ell),[t]} \right) \quad (40a)$$

$$\text{s.t.} \quad \sum_{k=1}^{K_u} \text{tr} \{ \mathbf{\Lambda}_{k,u} \} \leq P_{\max,u} \quad (40b)$$

$$\mathbf{\Lambda}_{k,u} \succeq \mathbf{0}, \mathbf{\Lambda}_{k,u} \text{ diagonal}, \forall k \in \mathcal{K}_u, \quad (40c)$$

where set $\mathcal{K}_u \triangleq \{1, \dots, K_u\}$ denotes all the UTs in cell u . We present the following proposition for the solution to problem $\mathcal{P}_{6,u}^{(\ell),[t]}$.

Proposition 3: Denote the optimal power allocation matrices of problem $\mathcal{P}_{6,u}^{(\ell),[t]}$ in (40) as $\mathbf{\Lambda}_{k,u}^{(\ell),[t+1]}$, $\forall (k, u) \in \mathcal{K}$. Then, the m th diagonal element of $\mathbf{\Lambda}_{k,u}^{(\ell),[t+1]}$, i.e., $\lambda_{k,u,m}^{(\ell),[t+1]}$, satisfies (41), shown at the top of next page, where $\lambda_{k,u,m}^{(\ell),[t+1]}$ is the m th diagonal elements of $\mathbf{\Lambda}_{k,u}^{(\ell),[t+1]}$. The Lagrange multiplier $\theta_u^{[t+1]}$ is chosen to satisfy

$$\theta_u^{[t+1]} \left\{ \sum_{k=1}^{K_u} \text{tr} \left\{ \mathbf{\Lambda}_{k,u}^{(\ell),[t+1]} \right\} - P_{\max,u} \right\} = 0, \theta_u^{[t+1]} \geq 0, \quad (42)$$

and $\nu_{k,m}^{(\ell),[t+1]}$ is the auxiliary variable written as

$$\begin{aligned} \nu_{k,m}^{(\ell),[t+1]} &= w_u \gamma_{k,u,m} - w_u \sum_{a \neq k} d_{a,u,u,m}^{(\ell)} - \eta^{(\ell),[t]} \xi_u \\ &+ \sum_{a \neq k} \sum_{n=1}^{N_{a,u}} \frac{w_u [\mathbf{\Omega}_{a,u,u}]_{n,m}}{\tilde{\gamma}_{a,u,n} + \sigma^2 + C_{k,m,a,u}(\mathbf{\Lambda}^{(\ell),[t]}, \mathbf{\Lambda}^{(\ell),[t+1]})}, \end{aligned} \quad (43)$$

where the function $C_{k,m,a,u}(\mathbf{\Lambda}^{(\ell),[t]}, \mathbf{\Lambda}^{(\ell),[t+1]})$ and $\mathcal{S}_{k,m,a}$ are defined by (44) and (45) shown at the top of next page, respectively.

Proof: See Appendix B.

As we can see from *Proposition 3*, the solutions of (41) resemble a classical water-filling solutions. Similar to the centralized approach, we apply Newton's method to find the approximate roots of Eq. (41). The auxiliary functions $\rho_{k,m}^{(\ell),[t]}(x_{k,m})$ and $\rho_{k,m}'^{(\ell),[t]}(x_{k,m})$ for Newton's method are defined by (46) and (47) shown at the top of next page, respectively. Moreover, the Lagrange multiplier $\theta_u^{[t+1]}$ is updated by the subgradient method.

From (46) and (44), we can see that for calculating $\rho_{k,m}^{(\ell),[t]}(x_{k,m})$, calculations of variables $\eta^{(\ell),[t]} = \min_u \left\{ \eta_u^{(\ell),[t]} \right\}$ and $\sum_{q \neq u} \sum_{p=1}^{K_q} \text{tr} \left\{ \mathbf{\Lambda}_{p,q}^{(\ell),[t]} \widehat{\mathbf{R}}_{a,u,q,n} \right\}$ require information exchange across BSs, while $\gamma_{k,u,m}$, w_u , $d_{a,u,u,m}^{(\ell)}$, ξ_u , $\theta_u^{[t]}$ and $\mathbf{\Omega}_{a,u,u}$ are all local variables. In addition, the required information exchange for calculating $\rho_{k,m}'^{(\ell),[t]}(x_{k,m})$ is the same as that for $\rho_{k,m}^{(\ell),[t]}(x_{k,m})$. Then, we write $\sum_{q \neq u} \sum_{p=1}^{K_q} \text{tr} \left\{ \mathbf{\Lambda}_{p,q}^{(\ell),[t]} \widehat{\mathbf{R}}_{a,u,q,n} \right\} = \sum_{q \neq u} c_{a,n,q}$ where

$$c_{a,n,q} = \sum_{p=1}^{K_q} \text{tr} \left\{ \mathbf{\Lambda}_{p,q}^{(\ell),[t]} \widehat{\mathbf{R}}_{a,u,q,n} \right\},$$

$$a = 1 \dots K_u, a \neq k; n = 1 \dots N_{a,u}. \quad (48)$$

Note that auxiliary variable $c_{a,n,q}$ is calculated at cell $q \neq u$ and then broadcast to BS- u .

Based on the above discussions, we present the distributed iterative water-filling algorithm in **Algorithm 3**. In the following, we quantify the exchanged information of the proposed distributed algorithm. The backhaul signaling mainly happens when calculating the Newton's method auxiliary functions, which corresponds to Step 8 of **Algorithm 3**. In each iteration of the distributed algorithm, the required information to be sent from each BS- u for exchange includes $1 + (U - 1)(K_u - 1)N_{a,u}$ real-valued scalar variables. So the total number of exchanged variables for all the BSs is $\sum_{u=1}^U (1 + (U - 1)(K_u - 1)N_{a,u})$ in each iteration. Note that the proposed distributed algorithm converges within few iterations. In contrast, for the centralized algorithm, all the statistical CSI, i.e., $\mathbf{\Omega}_{k,u,v} \in \mathbb{R}^{N_{k,u} \times M}$, is required at the central unit for optimization, and then the power allocation results are sent back to all BSs. Then, the required information for exchange includes $MN_{k,u}U \sum_{u=1}^U K_u + M \sum_{u=1}^U K_u$ real-valued scalar variables. Note that the amount of backhaul signaling of the centralized algorithm scales with the number of antennas M , while that of the distributed algorithm does not rely on it. Thus, the proposed distributed algorithm can significantly reduce the backhaul overhead when compared with the centralized one, especially in massive MIMO systems where M is large.

C. Complexity Analysis

We analyze the complexity of the proposed algorithms in this subsection. For **Algorithm 1**, the main complexity of each iteration lies in the complexity of tackling problem $\mathcal{P}_4^{(\ell),[t]}$ in (27), which is solved by the centralized **Algorithm 2** or by the distributed **Algorithm 3**. In the following, we analyze the complexity for the proposed centralized and distributed approaches, respectively.

For the centralized approach, the major complexity of **Algorithm 2** lies in two aspects, i.e., solving Eq. (33) via Newton's method and updating the Lagrange multipliers through the subgradient method. Specifically, with the precision set to be d digits [46], the required number of iterations for Newton's method is $\log d$ [49]. Then, the computational complexity of **Algorithm 2** is $\mathcal{O}(L_S U K M (\log d + 1))$ [39], where L_S is the number of iterations in the subgradient method, and $K = \sum_{u=1}^U K_u$. Therefore, the computational complexity of the centralized method is approximately $\mathcal{O}(L_M L_D L_S U K M (\log d + 1))$, where L_M and L_D are the required numbers of iterations for the MM method, Dinkelbach's transform, respectively.

For the distributed approach where the problem $\mathcal{P}_4^{(\ell),[t]}$ is solved by **Algorithm 3**, the complexity analysis of **Algorithm 3** is similar to that of **Algorithm 2**, which is $\mathcal{O}(L_S K_u M (\log d + 1))$. Then we can formulate the complexity of the distributed algorithm as $\mathcal{O}(L_M L_D L_S K M (\log d + 1))$, which is reduced compared with the centralized algorithm, especially when U is large.

$$\left\{ \begin{array}{l} \frac{w_u \gamma_{k,u,m}}{1 + \gamma_{k,u,m} \lambda_{k,u,m}^{(\ell),[t+1]}} + \sum_{a \neq k} \sum_{n=1}^{K_u} \frac{w_u [\mathbf{\Omega}_{a,u,u}]_{n,m}}{\left[\tilde{\mathbf{r}}_{a,u} + \tilde{\mathbf{K}}_{a,u} \left(\left\{ \mathbf{\Lambda}_{k,u}^{(\ell),[t+1]} \right\}_{k=1}^{K_u}, \mathbf{\Lambda}_{-u}^{(\ell),[t]} \right) \right]_{n,n}} = w_u \sum_{a \neq k} d_{a,u,u,m}^{(\ell)} + \eta^{(\ell),[t]} \xi_u + \theta_u^{[t+1]}, \\ \lambda_{k,u,m}^{(\ell),[t+1]} = 0, \end{array} \right. \quad \begin{array}{l} \theta_u^{[t+1]} < \nu_{k,m}^{(\ell),[t+1]} \\ \theta_u^{[t+1]} \geq \nu_{k,m}^{(\ell),[t+1]} \end{array} \quad (41)$$

$$C_{k,m,a,u} \left(\mathbf{\Lambda}^{(\ell),[t]}, \mathbf{\Lambda}^{(\ell),[t+1]} \right) = \sum_{q \neq u} \sum_{p=1}^{K_q} \text{tr} \left\{ \mathbf{\Lambda}_{p,q}^{(\ell),[t]} \hat{\mathbf{R}}_{a,u,q,n} \right\} + \sum_{(a',m') \in \mathcal{S}_{k,m,a}} \lambda_{a',m',u}^{(\ell),[t+1]} [\mathbf{\Omega}_{a,u,u}]_{n,m'} \quad (44)$$

$$\mathcal{S}_{k,m,a} = \{(a', m') \mid a' \neq a, (a', m') \neq (k, m), a' \in \{1, \dots, K_u\}, m' \in \{1, \dots, M\}\} \quad (45)$$

$$\begin{aligned} \rho_{k,m}^{(\ell),[t]}(x_{k,m}) &= \frac{w_u \gamma_{k,u,m}}{1 + \gamma_{k,u,m} x_{k,m}} - w_u \sum_{a \neq k} d_{a,u,u,m}^{(\ell)} - \eta^{(\ell),[t]} \xi_u - \theta_u^{[t]} \\ &+ \sum_{a \neq k} \sum_{n=1}^{K_u} \frac{w_u [\mathbf{\Omega}_{a,u,u}]_{n,m}}{\tilde{\gamma}_{a,u,n} + \sigma^2 + x_{k,m} [\mathbf{\Omega}_{a,u,u}]_{n,m} + C_{k,m,a,u} \left(\mathbf{\Lambda}^{(\ell),[t]}, \mathbf{X}^s \right)} \end{aligned} \quad (46)$$

$$\rho_{k,m}'^{(\ell),[t]}(x_{k,m}) = - \frac{w_u (\gamma_{k,u,m})^2}{(1 + \gamma_{k,u,m} x_{k,m})^2} - \sum_{a \neq k} \sum_{n=1}^{K_u} \frac{w_u [\mathbf{\Omega}_{a,u,u}]_{n,m}^2}{\left(\tilde{\gamma}_{a,u,n} + \sigma^2 + x_{k,m} [\mathbf{\Omega}_{a,u,u}]_{n,m} + C_{k,m,a,u} \left(\mathbf{\Lambda}^{(\ell),[t]}, \mathbf{X}^s \right) \right)^2} \quad (47)$$

Algorithm 3 Distributed Water-Filling Power Allocation Algorithm

- 1: Initialization: $\mathbf{X}_k^0 = \mathbf{\Lambda}_{k,u}^{(\ell),[t]}$, $k = 1, \dots, K_u$, and iteration index $s = 0$. $x_{k,m}^s$ stands for the m th diagonal entry of \mathbf{X}_k^s . Initialize $w = 0$ and $\theta_u^w = 1/K_u$. Set thresholds ϵ_5 and ϵ_6 .
- 2: **repeat**
- 3: **for** $k = 1$ to K_u **do**
- 4: **for** $m = 1$ to M **do**
- 5: Set $s' = s$.
- 6: **repeat**
- 7: BS- u receives $c_{a,n,q}$, $q \neq u; a = 1 \dots K_u, a \neq k; n = 1 \dots N_{a,u}$.
- 8: Calculate $\rho_{k,m}^{(\ell),[t]}(x_{k,m}^{s'})$ and $\rho_{k,m}'^{(\ell),[t]}(x_{k,m}^{s'})$ with (46) and (47), respectively.
- 9: Update $x_{k,m}^{s'+1} = x_{k,m}^{s'} - \frac{\rho_{k,m}^{(\ell),[t]}(x_{k,m}^{s'})}{\rho_{k,m}'^{(\ell),[t]}(x_{k,m}^{s'})}$.
- 10: $s' = s' + 1$.
- 11: **until** $|x_{k,m}^{s'} - x_{k,m}^{s'-1}| \leq \epsilon_5$.
- 12: **end for**
- 13: **end for**
- 14: Set $\bar{x}_{k,m} = \left[x_{k,m}^{s'} \right]^+$ and calculate $p_{\text{tot},u} = \sum_{k,m} \bar{x}_{k,m}$, $k = 1, \dots, K_u; m = 1, \dots, M$.
- 15: Update θ_u^{w+1} by the subgradient method.
- 16: **until** $|\theta_u^w - \theta_u^{w-1}| \leq \epsilon_6$.

Moreover, the distributed approach can perform power allocation with limited intercell coordination at an acceptable sacrifice on the EE performance compared with the centralized one, which will be shown in Section V by numerical results. Note that if problem $\mathcal{P}_4^{(\ell),[t]}$ in (27) of **Algorithm 1** is solved via the interior point method, the complexity can reach $\mathcal{O}(L_M L_D K^3 M^3)$ [46]. So our proposed algorithm can significantly reduce the computational complexity.

V. NUMERICAL RESULTS

We provide numerical analysis to illustrate the performance of the proposed max-min EE iterative algorithms in this section. The WINNER II channel model is employed to conduct the simulations [50]. Note that the suburban scenario with non-line-of-sight (NLOS) propagation is considered. In our simulations, the number of cells is $U = 3$, each cell contains one BS with $M = 128$ antennas and $K_u = 4$ UTs with $N_{k,u} = 4$ antennas. Both the BS and the UTs are equipped with ULAs where the spacing between antennas is half-wavelength. We set the noise power σ^2 as -105 dBm, the weighting factor as $w_u = 1, \forall u$, $P_{c,u}$ and $P_{s,u}$ are set as 30 dBm and 40 dBm, respectively. The amplifier inefficiency factor is set as $\xi_u = 5, \forall u$. The large scale fading factor is given as $10 \log_{10}(\theta_{k,u,v}) = -38 \log_{10}(d_{k,u,v} - 34.5 + c_{k,u,v})$, where $d_{k,u,v}$ represents the distance between the BS- v and UT- (k, u) and $c_{k,u,v}$ stands for the log-normal shadow fading with zero mean and 8 dB deviation [11]. The power budgets are equal for all cells, i.e., $P_{\max,u} = P_{\max}, \forall u$.

Firstly, we evaluate the convergence performance of **Algorithm 1** in Fig. 1. The results indicate that the proposed

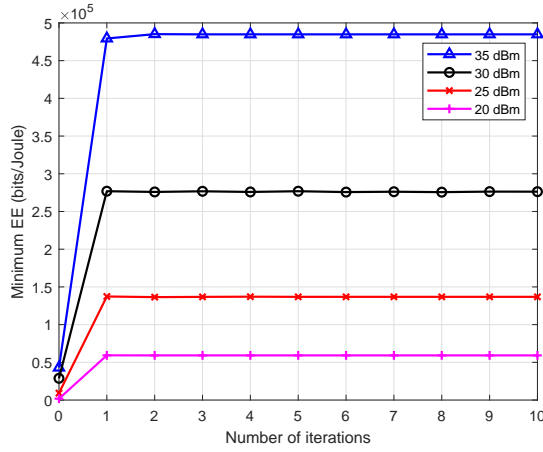


Fig. 1. Convergence behavior of **Algorithm 1** versus the numbers of iterations for different values of power budgets P_{\max} .

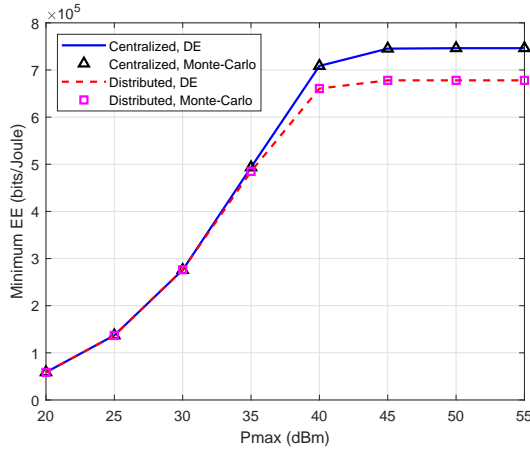
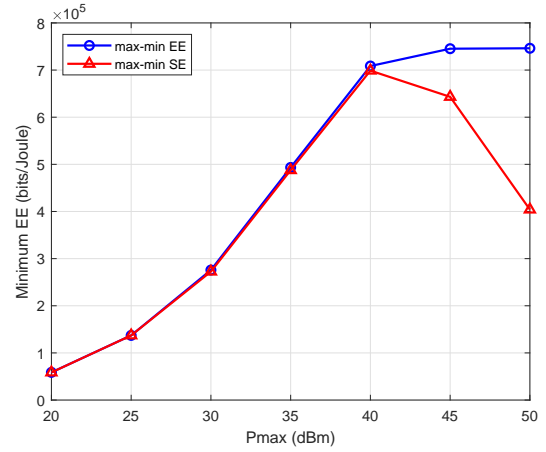


Fig. 2. The minimum EE performance of the distributed and centralized algorithms.

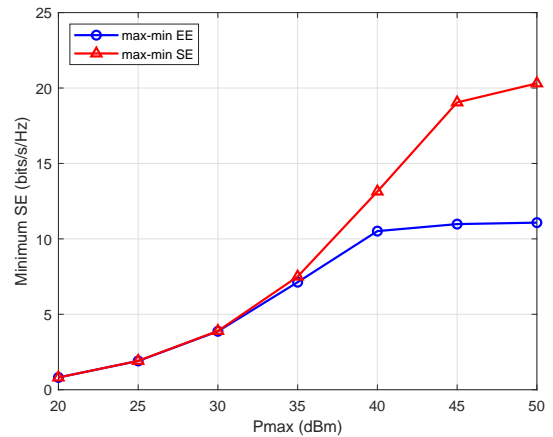
Algorithm 1 converges quickly after about two or three iterations under different power budgets.

Next, the EE performance of the proposed distributed and centralized power allocation algorithms are compared in Fig. 2. We can observe that as the power budget gets higher, both the minimum EE of the distributed and centralized algorithms become higher, and eventually reach a stationary point. In addition, in the low power budget regime, the noise dominates the interference. Then, the EE performance of centralized and distributed approaches are almost identical. While in the higher power budget regime, the centralized algorithm outperforms the distributed one in terms of the minimum EE performance as the noise becomes negligible compared to the interference. We can also find that the DE results match Monte-Carlo results accurately.

In Fig. 3, we compare our proposed max-min EE approach to the max-min SE approach, which is carried out by letting $\xi_u = 0, \forall u$ in the proposed max-min EE method. In particular, we show the minimum EE and minimum SE performance versus the transmit power constraint of the two methods in Fig. 3(a) and Fig. 3(b), respectively. We can observe that the max-



(a) Minimum EE



(b) Minimum SE

Fig. 3. Comparison between the proposed max-min EE and max-min SE approaches.

min EE and max-min SE approaches achieve almost identical performance when the power budget is low, which indicates that the minimum SE optimal is nearly minimum EE optimal when transmitting with a full power budget. Nevertheless, in the higher power budget regime, the EE value of the max-min EE method tends to saturate, while that of the max-min SE approach starts to decline rapidly, as depicted in Fig. 3(a). Moreover, in Fig. 3(b), the minimum SE value of the max-min SE method continues to grow, and that of max-min EE method tends to saturate. This is due to the existence of a threshold for the transmit power to maximize the minimum EE. When the transmission power budget exceeds that threshold, the system EE will not increase. On the contrary, the max-min SE method tends to optimize the minimum SE regardless of the power used, which might lead to a decline of the system EE.

Fig. 4 compares the minimum EE performance of our proposed centralized and distributed approaches with the conventional non-cooperative baseline, where each cell conducts its own precoding design under the criterion of maximizing the minimum EE of the UTs in its cell, regardless of the interference signal from other cells. We can observe that our proposed algorithms achieve better performance than

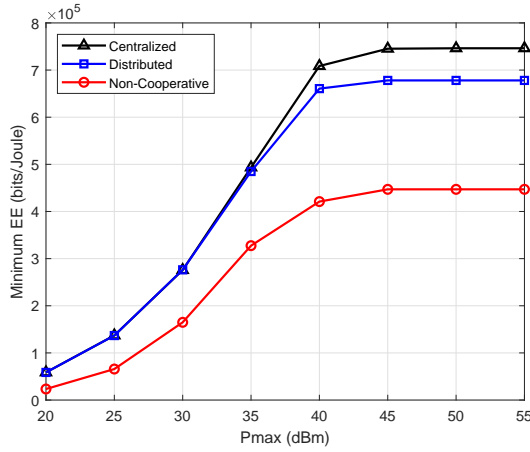


Fig. 4. Comparison between the proposed centralized and distributed approaches and the non-cooperative baseline.

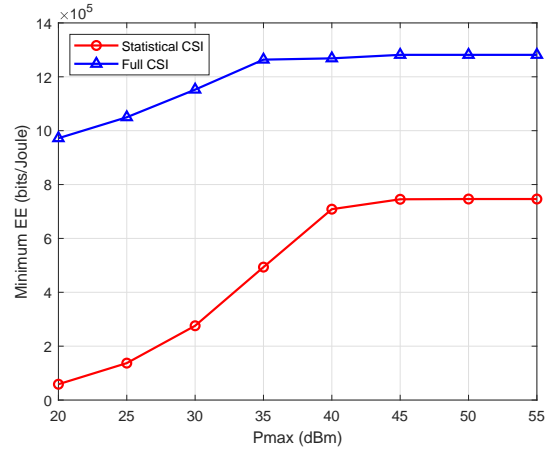


Fig. 6. The comparison between the statistical CSI and full CSI approaches on minimum EE performance.

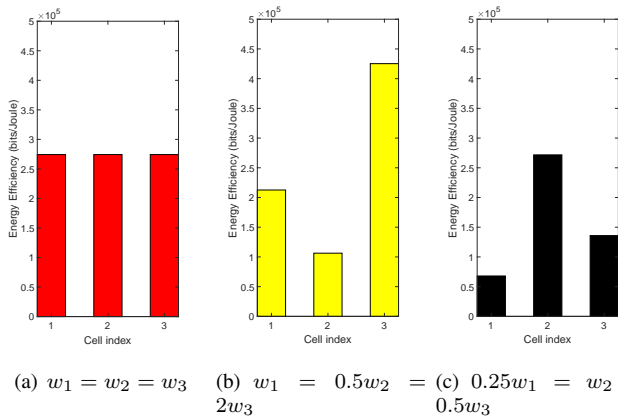


Fig. 5. The impacts of weighting factor on EE distribution under the power constraint $P_{\max,u} = 30$ dBm, $\forall u \in \mathcal{U}$.

the non-cooperative baseline. This is because during the cell-cooperation, precoding procedures are jointly processed among cells on the basis of information sharing and then the ICI is greatly mitigated.

The impact of the weighting factors on the distribution of the EE performance is shown in Fig. 5. For the equal weighting factor case where $w_u = 1, \forall u \in \mathcal{U}$, as shown in Fig. 5(a), the EE is almost identical among the cells. We further evaluate the cases when the weighting factors are unequal. In particular, we consider the cases with $w_1 = 0.5w_2 = 2w_3$ and $0.25w_1 = w_2 = 0.5w_3$ in Fig. 5(b) and Fig. 5(c), respectively. We can observe that the EE performance differs from each other according to their weights, which offers us a way to control the EE distribution between the cells when specific EE demands are required for some cells.

Finally, the EE performance of the proposed power allocation approach and that of the full CSI approach which assumes instantaneous CSI is available at each BS are compared in Fig. 6. Note that although the performance of the instantaneous CSI approach is better than our statistical CSI based one, more signalling overhead and higher complexity is incurred.

VI. CONCLUSION

We studied the precoding design in multi-cell multi-user massive MIMO downlink systems utilizing statistical CSI. Our objective was to maximize the minimum weighted EE by designing the transmit covariance matrices under the per BS power constraint. We first exploited the massive MIMO channel properties to derive the closed-form optimal transmit directions at each BS, thus simplifying the precoding design problem into a power allocation one. We adopted an asymptotically accurate surrogate expression of the ergodic user rate, i.e., its DE, which significantly reduced the optimization complexity. We further applied the MM technique and Dinkelbach's transform to deal with this max-min fractional optimization problem. Furthermore, we proposed two methods to address the obtained subproblems, i.e., the centralized and the distributed ones. For each method, we proposed generalized water-filling schemes to obtain the solution. Numerical results demonstrated the performance of the proposed algorithms.

APPENDIX A PROOF OF PROPOSITION 2

The Lagrangian function of problem $\mathcal{P}_5^{(\ell),[t]}$ is

$$\begin{aligned} \mathcal{L}(z, \mathbf{\Lambda}, \boldsymbol{\beta}, \boldsymbol{\mu}) &= z - \sum_{u=1}^U \mu_u \left\{ \sum_{k=1}^{K_u} \text{tr} \{ \mathbf{\Lambda}_{k,u} \} - P_{\max,u} \right\} \\ &\quad - \sum_{u=1}^U \beta_u \left\{ z - w_u \left\{ W \overline{R}_u^{(\ell)}(\mathbf{\Lambda}) - \eta^{(\ell),[t]} P_u(\mathbf{\Lambda}) \right\} \right\} \\ &= \left(1 - \sum_{u=1}^U \beta_u \right) z + \sum_{u=1}^U \mu_u \left(P_{\max,u} - \sum_{k=1}^{K_u} \text{tr} \{ \mathbf{\Lambda}_{k,u} \} \right) \\ &\quad + \sum_{u=1}^U \beta_u w_u \left(W \overline{R}_u^{(\ell)}(\mathbf{\Lambda}) - \eta^{(\ell),[t]} P_u(\mathbf{\Lambda}) \right), \quad (49) \end{aligned}$$

where the Lagrange multipliers $\boldsymbol{\beta} = \{\beta_1, \dots, \beta_U\}, \beta_u \geq 0$ ($\forall u$) and $\boldsymbol{\mu} = \{\mu_1, \dots, \mu_U\}, \mu_u \geq 0$ ($\forall u$) are associated with the dual variables corresponding to the constraints in (32b) and (32c), respectively.

To maximize \mathcal{L} in (49) with fixed β and μ , the following conditions have to be satisfied

$$\frac{\partial \mathcal{L}}{\partial z} = 1 - \sum_{u=1}^U \beta_u = 0, \quad (50a)$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \Lambda_{k,u}} &= \sum_{j=1}^U \beta_j w_j \left(W \frac{\partial \bar{R}_j^{(\ell)}(\Lambda)}{\partial \Lambda_{k,u}} - \eta^{(\ell),[t]} \xi_j \mathbf{I}_M \right) \\ &\quad - \sum_{j=1}^U \mu_j \mathbf{I}_M = \mathbf{0}, \quad \forall (k, u) \in \mathcal{K}. \end{aligned} \quad (50b)$$

It can be observed from (50a) that to obtain the maximal value of \mathcal{L} , the summation of β_u has to be equal to one. By plugging (50a) into (49), we can find that z will not influence the value of \mathcal{L} , which further leads to the simplification of $\mathcal{L}(z, \Lambda, \beta, \mu)$ to $\mathcal{L}(\Lambda, \beta, \mu)$.

To derive the gradient of $\bar{R}_j^{(\ell)}(\Lambda)$ over $\Lambda_{k,u}$, we rewrite the derivative $\frac{\partial \bar{R}_j^{(\ell)}(\Lambda)}{\partial \Lambda_{k,u}}$ in (50b) as

$$\frac{\partial \bar{R}_j^{(\ell)}(\Lambda)}{\partial \Lambda_{k,u}} = \frac{\partial}{\partial \Lambda_{k,u}} \sum_{a=1}^{K_j} \bar{A}_{a,j}(\Lambda) - \frac{\partial}{\partial \Lambda_{k,u}} \sum_{a=1}^{K_j} \Delta B_{a,j}^{(\ell)}(\Lambda), \quad (51)$$

where the gradient of $\bar{A}_{a,j}(\Lambda)$ and $\Delta B_{a,j}^{(\ell)}(\Lambda)$ over $\Lambda_{k,u}$ are described by (52) and (53), respectively, shown at the top of next page, and $\tilde{\gamma}_{a,u,n}$ represents the n th diagonal element of $\tilde{\Gamma}_{a,u}$.

With the aid of (51), we can rewrite the KKT condition (50b) as

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \Lambda_{k,u}} &= \beta_u w_u W (\mathbf{I}_M + \Gamma_{k,u} \Lambda_{k,u})^{-1} \Gamma_{k,u} \\ &\quad + \sum_{a \neq k} \sum_{n=1}^{N_{a,u}} \frac{\beta_u w_u W \hat{\mathbf{R}}_{a,u,u,n}}{\tilde{\gamma}_{a,u,n} + \sigma^2 + \sum_{(p,q) \neq (a,u)} \text{tr} \{ \Lambda_{p,q} \hat{\mathbf{R}}_{a,u,q,n} \}} \\ &\quad + \sum_{j \neq u} \sum_{a=1}^{K_j} \sum_{n=1}^{N_{a,j}} \frac{\beta_j w_j W \hat{\mathbf{R}}_{a,j,u,n}}{\tilde{\gamma}_{a,j,n} + \sigma^2 + \sum_{(p,q) \neq (a,j)} \text{tr} \{ \Lambda_{p,q} \hat{\mathbf{R}}_{a,j,q,n} \}} \\ &\quad - \sum_{j \neq u} \beta_j w_j W \sum_{a=1}^{K_j} \mathbf{D}_{a,j,u}^{(\ell)} - \beta_u w_u W \sum_{a \neq k} \mathbf{D}_{a,u,u}^{(\ell)} \\ &\quad - \sum_{j=1}^U \left(\beta_j w_j \eta^{(\ell),[t]} \xi_j + \mu_j \right) \mathbf{I}_M = \mathbf{0}, \quad \forall (k, u) \in \mathcal{K}. \end{aligned} \quad (54)$$

Note that $\frac{\partial \mathcal{L}}{\partial \Lambda_{k,u}}$ is a diagonal matrix. Then, the KKT condition in (54) can be further reduced to

$$\begin{aligned} \left[\frac{\partial \mathcal{L}}{\partial \Lambda_{k,u}} \right]_{m,m} &= \frac{\beta_u w_u W \gamma_{k,u,m}}{1 + \gamma_{k,u,m} \lambda_{k,u,m}} - \beta_u w_u W \sum_{a \neq k} d_{a,u,u,m}^{(\ell)} \\ &\quad + \sum_{a \neq k} \sum_{n=1}^{N_{a,u}} \frac{\beta_u w_u W [\Omega_{a,u,u}]_{n,m}}{\tilde{\gamma}_{a,u,n} + \sigma^2 + \sum_{(p,q) \neq (a,u)} \text{tr} \{ \Lambda_{p,q} \hat{\mathbf{R}}_{a,u,q,n} \}} \\ &\quad + \sum_{j \neq u} \sum_{a=1}^{K_j} \sum_{n=1}^{N_{a,j}} \frac{\beta_j w_j W [\Omega_{a,j,u}]_{n,m}}{\tilde{\gamma}_{a,j,n} + \sigma^2 + \sum_{(p,q) \neq (a,j)} \text{tr} \{ \Lambda_{p,q} \hat{\mathbf{R}}_{a,j,q,n} \}} \end{aligned}$$

$$- \sum_{j=1}^U \left(\beta_j w_j \eta^{(\ell),[t]} \xi_j + \mu_j \right) - \sum_{j \neq u} \beta_j w_j W \sum_{a=1}^{K_j} d_{a,j,u,m}^{(\ell)} = 0, \quad m = 1, \dots, M. \quad (55)$$

Solving the KKT conditions in (55), we can find that the m th diagonal element of $\Lambda_{k,u}^{(\ell),[t+1]}$, $\forall (k, u) \in \mathcal{K}$ satisfies (56) shown at the top of next page, where the auxiliary variables $\chi_{k,m}^{(\ell),[t+1]}$ and $\mathcal{T}_{k,m,a,u}$ are written as (57) and (58) shown at the top of next page, respectively. This concludes the proof.

APPENDIX B

PROOF OF PROPOSITION 3

The Lagrangian function of problem $\mathcal{P}_{6,u}^{(\ell),[t]}$ is

$$\begin{aligned} \mathcal{L} \left(\{ \Lambda_{k,u} \}_{k=1}^{K_u}, \theta_u \right) &= F_u \left(\{ \Lambda_{k,u} \}_{k=1}^{K_u}, \Lambda_{-u}^{(\ell),[t]} \right) \\ &\quad + \sum_{k=1}^{K_u} \text{tr} \{ \Lambda_{k,u} \Lambda_{k,u} \} \theta_u \left\{ \sum_{k=1}^{K_u} \text{tr} \{ \Lambda_{k,u} \} - P_{\max,u} \right\}, \end{aligned} \quad (59)$$

where the Lagrange multiplier $\{ \Lambda_{k,u} \succeq \mathbf{0} \}$ and $\theta_u \geq 0$ are associated with the problem constraints in (40b) and (40c), respectively.

The KKT conditions of (40) are

$$\frac{\partial \mathcal{L}}{\partial \Lambda_{k,u}} = \frac{\partial F_u}{\partial \Lambda_{k,u}} - \theta_u \mathbf{I}_M = \mathbf{0}, \quad \forall k \in \mathcal{K}_u, \quad (60)$$

$$\theta_u \left\{ \sum_{k=1}^{K_u} \text{tr} \{ \Lambda_{k,u} \} - P_{\max,u} \right\} = 0, \quad \theta_u \geq 0, \quad (61)$$

$$\text{tr} \{ \Lambda_{k,u} \Lambda_{k,u} \} = 0, \quad \Lambda_{k,u} \succeq \mathbf{0}, \quad \Lambda_{k,u} \succeq \mathbf{0}. \quad (62)$$

Since (40) is a convex optimization problem, we can solve the KKT conditions to find the optimal solution. Calculate the derivative of $F_u \left(\{ \Lambda_{k,u} \}_{k=1}^{K_u}, \Lambda_{-u}^{(\ell),[t]} \right)$ over the transmit power matrices of cell u , the first KKT condition in (60) can be rewritten as

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \Lambda_{k,u}} &= \frac{\partial F_u}{\partial \Lambda_{k,u}} - \theta_u \mathbf{I}_M = w_u (\mathbf{I}_M + \Gamma_{k,u} \Lambda_{k,u})^{-1} \Gamma_{k,u} \\ &\quad + \sum_{a \neq k} \sum_{n=1}^{N_{a,u}} \frac{w_u \hat{\mathbf{R}}_{a,u,u,n}}{\left[\tilde{\Gamma}_{a,u} + \tilde{\mathbf{K}}_{a,u} \left(\{ \Lambda_{a,u} \}_{a=1}^{K_u}, \Lambda_{-u}^{(\ell),[t]} \right) \right]_{n,n}} \\ &\quad - w_u \sum_{a \neq k} \mathbf{D}_{a,u,u}^{(\ell)} - \left(\eta^{(\ell),[t]} \xi_u + \theta_u \right) \mathbf{I}_M = \mathbf{0}, \quad k \in \mathcal{K}_u. \end{aligned} \quad (63)$$

Using the KKT conditions in (60)–(62), the solution of $\mathcal{P}_{6,u}^{(\ell),[t]}$, which is $\{ \Lambda_{k,u}^{(\ell),[t+1]} \}_{k=1}^{K_u}$, satisfies (64) shown at the top of next page, where $\theta_u^{[t+1]}$ is chosen to satisfy the KKT condition in (61), and the auxiliary variable $\nu_{k,m}^{(\ell),[t+1]}$ is written as

$$\begin{aligned} \nu_{k,m}^{(\ell),[t+1]} &= w_u \gamma_{k,u,m} - w_u \sum_{a \neq k} d_{a,u,u,m}^{(\ell)} - \eta^{(\ell),[t]} \xi_u \\ &\quad + \sum_{a \neq k} \sum_{n=1}^{N_{a,u}} \frac{w_u [\Omega_{a,u,u}]_{n,m}}{\tilde{\gamma}_{a,u,n} + \sigma^2 + C_{k,m,a,u} \left(\Lambda^{(\ell),[t]}, \Lambda^{(\ell),[t+1]} \right)}, \end{aligned} \quad (65)$$

$$\frac{\partial}{\partial \Lambda_{k,u}} \sum_{a=1}^{K_j} \bar{A}_{a,j}(\Lambda) = \begin{cases} (\mathbf{I}_M + \Gamma_{k,u} \Lambda_{k,u})^{-1} \Gamma_{k,u} + \sum_{a \neq k} \sum_{n=1}^{N_{a,u}} \frac{\hat{\mathbf{R}}_{a,u,u,n}}{\tilde{\gamma}_{a,u,n} + \sigma^2 + \sum_{(p,q) \neq (a,u)} \text{tr}\{\Lambda_{p,q} \hat{\mathbf{R}}_{a,u,q,n}\}}, & j = u \\ \sum_{a=1}^{K_j} \sum_{n=1}^{N_{a,j}} \frac{\hat{\mathbf{R}}_{a,j,u,n}}{\tilde{\gamma}_{a,j,n} + \sigma^2 + \sum_{(p,q) \neq (a,j)} \text{tr}\{\Lambda_{p,q} \hat{\mathbf{R}}_{a,j,q,n}\}}, & j \neq u \end{cases} \quad (52)$$

$$\frac{\partial}{\partial \Lambda_{k,u}} \sum_{a=1}^{K_j} \Delta B_{a,j}^{(\ell)}(\Lambda) = \begin{cases} \frac{\partial}{\partial \Lambda_{k,u}} \sum_{a \neq k} B_{a,u}(\Lambda^{(\ell)}) = \sum_{a \neq k} \mathbf{D}_{a,u}^{(\ell)}, & j = u \\ \frac{\partial}{\partial \Lambda_{k,u}} \sum_{a=1}^{K_j} B_{a,j}(\Lambda^{(\ell)}) = \sum_{a=1}^{K_j} \mathbf{D}_{a,j,u}^{(\ell)}, & j \neq u \end{cases} \quad (53)$$

$$\begin{cases} \frac{\beta_u w_u W \gamma_{k,u,m} \lambda_{k,u,m}^{(\ell),[t+1]} + \sum_{a \neq k} \sum_{n=1}^{N_{a,u}} \frac{\beta_u w_u W [\Omega_{a,u,u}]_{n,m}}{\tilde{\gamma}_{a,u,n} + \sigma^2 + [\tilde{\mathbf{K}}_{a,u}(\Lambda)]_{n,n}} + \sum_{j \neq u} \sum_{a=1}^{K_j} \sum_{n=1}^{N_{a,j}} \frac{\beta_j w_j W [\Omega_{a,j,u}]_{n,m}}{\tilde{\gamma}_{a,j,n} + \sigma^2 + [\tilde{\mathbf{K}}_{a,j}(\Lambda)]_{n,n}} = \eta^{(\ell),[t]} \xi_u + \mu_u^{[t+1]} \\ + \beta_u w_u W \sum_{a \neq k} d_{a,u,u,m}^{(\ell)} + \sum_{j \neq u} \beta_j w_j W \sum_{a=1}^{K_j} d_{a,j,u,m}^{(\ell)}, & \mu_u^{[t+1]} < \chi_{k,m}^{(\ell),[t+1]} \\ \lambda_{k,u,m}^{(\ell),[t+1]} = 0, & \mu_u^{[t+1]} \geq \chi_{k,m}^{(\ell),[t+1]} \end{cases} \quad (56)$$

$$\begin{aligned} \chi_{k,m}^{(\ell),[t+1]} &= \gamma_{k,u,m} - \beta_u w_u W \sum_{a \neq k} d_{a,u,u,m}^{(\ell)} + \sum_{a \neq k} \sum_{n=1}^{N_{a,u}} \frac{\beta_u w_u W [\Omega_{a,u,u}]_{n,m}}{\tilde{\gamma}_{a,u,n} + \sigma^2 + \sum_{(a',q,m') \in \mathcal{T}_{k,m,a,u}} \lambda_{a',q,m'}^{(\ell),[t+1]} [\Omega_{a,u,u}]_{n,m'}} \\ &\quad - \eta^{(\ell),[t]} \xi_u - \sum_{j \neq u} \beta_j w_j W \sum_{a=1}^{K_j} d_{a,j,u,m}^{(\ell)} + \sum_{j \neq u} \sum_{a=1}^{K_j} \sum_{n=1}^{N_{a,j}} \frac{\beta_j w_j W [\Omega_{a,j,u}]_{n,m}}{\tilde{\gamma}_{a,j,n} + \sigma^2 + \sum_{(a',q,m') \in \mathcal{T}_{k,m,a,j}} \lambda_{a',q,m'}^{(\ell),[t+1]} [\Omega_{a,j,u}]_{n,m'}} \end{aligned} \quad (57)$$

$$\mathcal{T}_{k,m,a,u} = \{(a', q, m') \mid (a', q) \neq (a, u), (a', q, m') \neq (k, u, m), (a', q, m') \in \mathcal{K}, m' \in \{1, \dots, M\}\} \quad (58)$$

$$\begin{cases} \frac{w_u \gamma_{k,u,m}}{1 + \gamma_{k,u,m} \lambda_{k,u,m}^{(\ell),[t+1]}} + \sum_{a \neq k} \sum_{n=1}^{N_{a,u}} \frac{w_u [\Omega_{a,u,u}]_{n,m}}{[\tilde{\Gamma}_{a,u} + \tilde{\mathbf{K}}_{a,u}(\{\Lambda_{k,u}^{(\ell),[t+1]}\}_{k=1}^{K_u}, \Lambda_{-u}^{(\ell),[t]})]_{n,n}} \\ = w_u \sum_{a \neq k} d_{a,u,u,m}^{(\ell)} + \eta^{(\ell),[t]} \xi_u + \theta_u^{[t+1]}, & \theta_u^{[t+1]} < \nu_{k,m}^{(\ell),[t+1]} \\ \lambda_{k,u,m}^{(\ell),[t+1]} = 0, & \theta_u^{[t+1]} \geq \nu_{k,m}^{(\ell),[t+1]} \end{cases} \quad (64)$$

where

$$\begin{aligned} C_{k,m,a,u}(\Lambda^{(\ell),[t]}, \Lambda^{(\ell),[t+1]}) &= \sum_{q \neq u} \sum_{p=1}^{K_q} \text{tr} \left\{ \Lambda_{p,q}^{(\ell),[t]} \hat{\mathbf{R}}_{a,u,q,n} \right\} \\ &\quad + \sum_{(a',m') \in \mathcal{S}_{k,m,a}} \lambda_{a',m',u}^{(\ell),[t+1]} [\Omega_{a,u,u}]_{n,m'}, \quad (66) \\ \mathcal{S}_{k,m,a} &= \{(a', m') \mid a' \neq a, (a', m') \neq (k, m), \\ &\quad a' \in \{1, \dots, K_u\}, m' \in \{1, \dots, M\}\}. \quad (67) \end{aligned}$$

This concludes the proof.

REFERENCES

- [1] Y. Huang, L. You, J. Xiong, W. Wang, and X. Gao, "Max-min energy-efficient multi-cell massive MIMO transmission exploiting statistical CSI," in *Proc. IEEE ICC, Chongqing, China*, Aug. 2020, pp. 1–6.
- [2] A. Gupta and R. K. Jha, "A survey of 5G network: Architecture and emerging technologies," *IEEE Access*, vol. 3, pp. 1206–1232, Jul. 2015.
- [3] J. Zhang, E. Björnson, M. Matthaiou, D. W. K. Ng, H. Yang, and D. J. Love, "Prospective multiple antenna technologies for beyond 5G," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1637–1660, Jun. 2020.
- [4] M. E. Morocho-Cayamcela, H. Lee, and W. Lim, "Machine learning for 5G/B5G mobile and wireless communications: Potential, limitations, and future directions," *IEEE Access*, vol. 7, pp. 137 184–137 206, Sep. 2019.
- [5] 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.
- [6] G. Interdonato, M. Karlsson, E. Björnson, and E. G. Larsson, "Local partial zero-forcing precoding for cell-free massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 19, no. 7, pp. 4758–4774, Jul. 2020.
- [7] Y. Huang, G. Zheng, M. Bengtsson, K.-K. Wong, L. Yang, and B. Ottersten, "Distributed multicell beamforming design approaching Pareto boundary with max-min fairness," *IEEE Trans. Wireless Commun.*, vol. 11, no. 8, pp. 2921–2933, Aug. 2012.
- [8] D. Gesbert, S. Hanly, H. Huang, S. Shamai Shitz, O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, Dec. 2010.
- [9] D. Ponukumati, F. Gao, M. Bode, and X. Liao, "Multicell downlink beamforming with imperfect channel knowledge at both transceiver sides," *IEEE Commun. Lett.*, vol. 15, no. 10, pp. 1075–1077, Aug. 2011.
- [10] C. Sun, X. Q. Gao, and Z. Ding, "BDMA in multicell massive MIMO communications: Power allocation algorithms," *IEEE Trans. Signal Process.*, vol. 65, no. 11, pp. 2962–2974, Jun. 2017.
- [11] S. He, Y. Huang, S. Jin, and L. Yang, "Coordinated beamforming for energy efficient transmission in multicell multiuser systems," *IEEE Trans. Commun.*, vol. 61, no. 12, pp. 4961–4971, Dec. 2013.
- [12] Y. Huang, G. Zheng, M. Bengtsson, K.-K. Wong, L. Yang, and B. Ottersten, "Distributed multicell beamforming with limited intercell coordination," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 728–738, Feb. 2011.
- [13] O. Somekh, B. M. Zaidel, and S. Shamai, "Sum rate characterization

- of joint multiple cell-site processing," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4473–4497, Dec. 2007.
- [14] B. L. Ng, J. S. Evans, S. V. Hanly, and D. Aktas, "Distributed downlink beamforming with cooperative base stations," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5491–5499, Dec. 2008.
- [15] C. Pan, W. Zhang, B. Du, H. Ren, and M. Chen, "Totally distributed energy-efficient transmission design in MIMO interference channels," in *Proc. IEEE GLOBECOM*, Austin, TX, USA, Dec. 2014, pp. 3964–3969.
- [16] A. Zappone and E. Jorswieck, "Energy efficiency in wireless networks via fractional programming theory," *Found. Trends Commun. Inf. Theory*, vol. 11, no. 3–4, pp. 185–396, Jun. 2015.
- [17] L. You, J. Xiong, X. Yi, J. Wang, W. Wang, and X. Gao, "Energy efficiency optimization for downlink massive MIMO with statistical CSIT," *IEEE Trans. Wireless Commun.*, vol. 19, no. 4, pp. 2684–2698, Apr. 2020.
- [18] O. Tervo, L.-N. Tran, H. Pannanen, S. Chatzinotas, B. Ottersten, and M. Juntti, "Energy-efficient multicell multigroup multicasting with joint beamforming and antenna selection," *IEEE Trans. Signal Process.*, vol. 66, no. 18, pp. 4904–4919, Sep. 2018.
- [19] O. Tervo, A. Tölli, M. Juntti, and L. Tran, "Energy-efficient beam coordination strategies with rate-dependent processing power," *IEEE Trans. Signal Process.*, vol. 65, no. 22, pp. 6097–6112, Jul. 2017.
- [20] S. He, Y. Huang, L. Yang, and B. Ottersten, "Coordinated multicell multiuser precoding for maximizing weighted sum energy efficiency," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 741–751, Feb. 2014.
- [21] Y. Li, P. Fan, and N. C. Beaulieu, "Cooperative downlink max-min energy-efficient precoding for multicell MIMO networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 11, pp. 9425–9430, Nov. 2016.
- [22] B. Du, C. Pan, W. Zhang, and M. Chen, "Distributed energy-efficient power optimization for CoMP systems with max-min fairness," *IEEE Commun. Lett.*, vol. 18, no. 6, pp. 999–1002, Jun. 2014.
- [23] K. Nguyen, Q. Vu, M. Juntti, and L. Tran, "Distributed solutions for energy efficiency fairness in multicell MISO downlink," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 6232–6247, Sep. 2017.
- [24] L. You, X. Q. Gao, X.-G. Xia, N. Ma, and Y. Peng, "Pilot reuse for massive MIMO transmission over spatially correlated Rayleigh fading channels," *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3352–3366, Jun. 2015.
- [25] S. Gong, S. Wang, S. Chen, C. Xing, and L. Hanzo, "Robust energy efficiency optimization for amplify-and-forward MIMO relaying systems," *IEEE Trans. Wireless Commun.*, vol. 18, no. 9, pp. 4326–4343, Sep. 2019.
- [26] J. Tan, Q. Zhang, T. Q. S. Quek, and H. Shin, "Robust energy efficiency maximization in multicast downlink C-RAN," *IEEE Trans. Veh. Technol.*, vol. 68, no. 9, pp. 8951–8965, Sep. 2019.
- [27] L. You, W. Wang, and X. Gao, "Energy-efficient multicast precoding for massive MIMO transmission with statistical CSI," *Energies*, vol. 11, no. 11, p. 3175, Nov. 2018.
- [28] W. Wang, Y. Huang, L. You, J. Xiong, J. Li, and X. Gao, "Energy efficiency optimization for massive MIMO non-orthogonal unicast and multicast transmission with statistical CSI," *Electronics*, vol. 11, no. 11, p. 857, Aug. 2019.
- [29] L. You, J. Xiong, A. Zappone, W. Wang, and X. Gao, "Spectral efficiency and energy efficiency tradeoff in massive MIMO downlink transmission with statistical CSIT," *IEEE Trans. Signal Process.*, vol. 68, pp. 2645–2659, Apr. 2020.
- [30] L. You, J. Xiong, D. W. K. Ng, C. Yuen, W. Wang, and X. Gao, "Energy efficiency and spectral efficiency tradeoff in RIS-aided multiuser MIMO uplink transmission," *IEEE Trans. Signal Process.*, vol. 69, pp. 1407–1421, Mar. 2021.
- [31] W. Weichselberger, M. Herdin, H. Özcelik, and E. Bonek, "A stochastic MIMO channel model with joint correlation of both link ends," *IEEE Trans. Wireless Commun.*, vol. 5, no. 1, pp. 90–100, Jan. 2006.
- [32] X. Q. Gao, B. Jiang, X. Li, A. B. Gershman, and M. R. McKay, "Statistical eigenmode transmission over jointly correlated MIMO channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3735–3750, Aug. 2009.
- [33] L. You, X. Q. Gao, G. Y. Li, X.-G. Xia, and N. Ma, "BDMA for millimeter-wave/Terahertz massive MIMO transmission with per-beam synchronization," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1550–1563, Jul. 2017.
- [34] L. You, X. Q. Gao, A. L. Swindlehurst, and W. Zhong, "Channel acquisition for massive MIMO-OFDM with adjustable phase shift pilots," *IEEE Trans. Signal Process.*, vol. 64, no. 6, pp. 1461–1476, Mar. 2016.
- [35] 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.
- [36] C. Sun, X. Q. Gao, S. Jin, M. Matthaiou, Z. Ding, and C. Xiao, "Beam division multiple access transmission for massive MIMO communications," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2170–2184, Jun. 2015.
- [37] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.
- [38] A.-A. Lu, X. Gao, W. Zhong, C. Xiao, and X. Meng, "Robust transmission for massive MIMO downlink with imperfect CSI," *IEEE Trans. Commun.*, vol. 67, no. 8, pp. 5362–5376, Aug. 2019.
- [39] W. Wu, X. Q. Gao, Y. Wu, and C. Xiao, "Beam domain secure transmission for massive MIMO communications," *IEEE Trans. Veh. Technol.*, vol. 67, no. 8, pp. 7113–7127, Aug. 2018.
- [40] D. W. K. Ng, E. S. Lo, and R. Schober, "Energy-efficient resource allocation in OFDMA systems with large numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 11, no. 9, pp. 3292–3304, Sep. 2012.
- [41] A.-A. Lu, X. Q. Gao, and C. Xiao, "Free deterministic equivalents for the analysis of MIMO multiple access channel," *IEEE Trans. Inf. Theory*, vol. 62, no. 8, pp. 4604–4629, Aug. 2016.
- [42] R. Couillet and M. Debbah, *Random Matrix Methods for Wireless Communications*. New York, NY, USA: Cambridge Univ. Press, 2011.
- [43] C.-K. Wen, S. Jin, and K.-K. Wong, "On the sum-rate of multiuser MIMO uplink channels with jointly-correlated Rician fading," *IEEE Trans. Commun.*, vol. 59, no. 10, pp. 2883–2895, Oct. 2011.
- [44] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 794–816, Feb. 2017.
- [45] K. Shen and W. Yu, "Fractional programming for communication systems—Part I: Power control and beamforming," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2616–2630, May 2018.
- [46] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge Univ. Press, 2004.
- [47] K.-X. Li, L. You, J. Wang, and X. Gao, "Physical layer multicasting in massive MIMO systems with statistical CSIT," *IEEE Trans. Veh. Technol.*, vol. 69, no. 2, pp. 1651–1665, Feb. 2020.
- [48] R. L. Burden, J. D. Faires, and A. M. Burden, *Numerical Analysis*, 10th ed. Boston, MA, USA: Cengage Learning, 2014.
- [49] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, 3rd ed. Cambridge, MA, USA: MIT Press, 2009.
- [50] P. Kyösti, J. Meinilä, L. Henttilä, X. Zhao, T. Jämsä, C. Schneider, M. Narandzic, M. Milojević, A. Hong, J. Ylitalo, V.-M. Holappa, M. Alatossava, R. Bultitude, Y. Jong, and T. Rautiainen, *WINNER II Channel Models*. IST-4-027756 WINNER II D1.1.2 V1.2, Feb. 2008.