

Energy Efficient Coordinated Beamforming for Multicell System: Duality-Based Algorithm Design and Massive MIMO Transition

Shiwen He, *Member, IEEE*, Yongming Huang, *Member, IEEE*, Luxi Yang, *Member, IEEE*, Björn Ottersten, *Fellow, IEEE*, and Wei Hong, *Fellow, IEEE*

Abstract—In this paper, we investigate joint beamforming and power allocation in multicell multiple-input single-output (MISO) downlink networks. Our goal is to maximize the utility function defined as the ratio between the system weighted sum rate and the total power consumption subject to the users' quality of service requirements and per-base-station (BS) power constraints. The considered problem is nonconvex and its objective is in a fractional form. To circumvent this problem, we first resort to an virtual uplink formulations of the the primal problem by introducing an auxiliary variable and applying the uplink-downlink duality theory. By exploiting the analytic structure of the optimal beamformers in the dual uplink problem, an efficient algorithm is then developed to solve the considered problem. Furthermore, to reduce further the exchange overhead between coordinated BSs in a large-scale antenna system, an effective coordinated power allocation solution only based on statistical channel state information is reached by deriving the asymptotic optimization problem, which is used to obtain the power allocation in a long-term timescale. Numerical results validate the effectiveness of our proposed schemes and show that both the spectral efficiency and the energy efficiency can be simultaneously improved over traditional downlink coordinated schemes, especially in the middle-high transmit power region.

Index Terms—Coordinated energy-efficient transmission, uplink-downlink duality, massive multiple-input-single-output system, beamforming and power allocation.

Manuscript received February 26, 2015; revised August 28, 2015 and October 23, 2015; accepted October 23, 2015. Date of publication November 2, 2015; date of current version December 15, 2015. This work was supported in part by the 863 Program of China under Grant 2015AA01A703, in part by the National Basic Research Program of China under Grant 2013CB329002, in part by the National Natural Science Foundation of China under Grant 61471120, Grant 61422105, Grant 61531011, and Grant 61372101, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20130019. The associate editor coordinating the review of this paper and approving it for publication was Z. Zhang. (*Corresponding author: Yongming Huang.*)

S. He is with the School of Information Science and Engineering, Southeast University, Nanjing 210096, China, and also with the State Key Laboratory of Millimeter Waves, Southeast University, Nanjing 210096, China (e-mail: shiwenhe@seu.edu.cn).

Y. Huang is with the School of Information Science and Engineering, Southeast University, Nanjing 210096, China (e-mail: huangym@seu.edu.cn).

L. Yang and W. Hong are with the School of Information Science and Engineering, Southeast University, Nanjing 210096, China (e-mail: lxyang@seu.edu.cn; weihong@seu.edu.cn).

B. Ottersten is with the Interdisciplinary Centre for Security Reliability and Trust (SnT), University of Luxembourg, Luxembourg, and also with the Royal Institute of Technology (KTH), Stockholm, Sweden (e-mail: bjorn.ottersten@uni.lu; bjorn.ottersten@ee.kth.seg).

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

Digital Object Identifier 10.1109/TCOMM.2015.2496948

I. INTRODUCTION

MULTIPLE-INPUT multiple-output (MIMO) technologies have attracted much research interest in the past decade due to the potential of significantly enhancing the link reliability and improving the spectral efficiency (SE) without requiring additional transmit power and bandwidth resources [1]. Among those, multiuser MIMO and coordinated multi-point transmission and reception (CoMP) have been intensively studied for advanced cellular systems [2]. The cooperation between the base stations (BSs) offers great advantages in suppressing the inter-cell interference and enabling efficient resource allocation to significantly increase the SE by exploiting the spatial diversity and multiplexing gain [3]–[5].

Recently, besides the objective of improving the SE, energy efficiency (EE) has become a main focus in mobile communications, due to the explosive growth of energy consumption resulting from the exponential increase in mobile multimedia data traffic and mobile terminals [6], [7]. Motivated by these, advanced technologies aiming at much improved EE have emerged as an important topic both in industry and academia [8].

EE is usually defined as the ratio of the system weighted sum rate (WSR) to the system total power consumption. Recently energy-efficient resource allocation has been widely studied for orthogonal frequency division multiple access (OFDMA) downlink networks with single antenna transceiver or fixed transmit beamformers [9]–[13]. On the other hand, there have been tremendous research efforts to improve the system EE by using multiple antenna technologies [14]–[19]. In [14] the authors studied the power allocation problem in the uplink of multiuser MIMO systems and showed that the EE is maximized by adaptively switching off some of the user antennas. The EE of multiuser MIMO downlink systems was also studied in [15]–[19]. In particular, in [15], the energy-efficient multiuser beamforming optimization problem was solved by approximating the objective with a convex and tight upper bound function. In [16] a zero-gradient-based energy-efficient iterative approach was developed which was guaranteed to converge to a local maximum for MIMO interference channel. The authors in [17] addressed the EE optimization problem by solving three subproblems and choosing the best one. Focusing on the downlink of multicell multiuser systems, we addressed in [18], [19] the energy-efficient beamforming problem with per-BS power constraints by using jointly fractional programming

[20]–[22] and the relationship between the user rate and the minimum mean square error (MMSE). The research results in the above literature have shown that in many cases, the system EE is improved at the cost of decreasing system SE. This means that it is crucial to achieve a reasonable tradeoff between EE and SE in the design of a wireless communications system.

Recently, massive MIMO technology has emerged as a promising solution to improve both SE and EE and is well analyzed especially for the uncorrelated channel model [23]–[25]. Particularly its achievable performance in metrics such as the SINR and sum rate (SR) has been derived with closed-form expressions for some special cases. For instance, the deterministic quantity of the SINR for zero-forcing (ZF) and regularized ZF (RZF) precoding were derived under channel uncertainty for the multiple-input single-output (MISO) broadcast system [26], showing that the performance of massive MIMO systems mainly depends on the channel statistics and the channel small fading has little effect. By exploiting the asymptotic behavior of the achievable performance of massive MIMO system, recently a number of coordinated beamforming algorithms have been developed for multicell systems, which usually have reduced inter-cell communication burden. In [27], [28], the signal to interference plus noise ratio (SINR) balancing problem for the multicell massive MIMO system was solved by deriving the asymptotic expression of the achievable SINR based on large-dimension random matrix theory (RMT) [29]–[32]. Furthermore, the asymptotic properties of the SINR were also used to address the power minimization problem subject to some predefined SINR constraints [33]–[35]. More recently, energy-efficient transmission design for massive MIMO systems has also drawn much attention. In [36], the energy-efficient resource allocation problem was studied in OFDMA downlink network with a large number of transmit antennas and fixed maximum ratio transmission (MRT) beamforming. The tradeoff between SE and EE was investigated in a massive MIMO uplink system with fixed receiver, such as maximum-ratio combining (MRC) or zero-forcing (ZF) [37]. More recently, the authors in [38]–[40] investigated the energy efficiency of massive MIMO systems with non-ideal hardware. The research results illustrated that the maximal EE is achieved at some finite number of transceiver antennas when there is a non-zero circuit power per antenna. Similarly, the studies in [41] also demonstrated that it is optimal from the perspective of EE to have a few transceiver antennas when the transceiver design is power inefficient.

In this paper, we investigate joint beamforming optimization and power allocation (JBOPA) for coordinated multicell downlink systems. First, the problem of interest is formulated as maximizing the network utility function defined as the ratio of system weighted sum rate to total power consumption subject to predefined user rate requirements and per-BS power constraints. Due to the non-convex nature of the user rate and the fractional form in the objective, the problem is hard to solve directly. To circumvent this problem, we first resort to an virtual uplink formulations of the the primal problem by introducing an auxiliary variable and the uplink-downlink duality theory. By exploiting the property that the dual problem has an analytic beamforming solution, we propose to further solve the power allocation using a convex approximation. By

this means, the dual problem can be efficiently solved and its solution is converted into the downlink using the conventional uplink-downlink theory. We show that the developed algorithm has provable convergence. Then, in order to further reduce the exchange of signalling overhead between coordinated BSs, especially in a large-scale system, a new algorithm used to update the downlink power based on the statistical channel information is developed by exploiting the property of a massive MIMO system. Numerical results validate the effectiveness of our proposed schemes and show that both the spectral efficiency and the energy efficiency can be simultaneously improved traditional coordinated schemes, especially in the middle-high transmit power region.

The remainder of this paper is organized as follows. The system model is described in Section II. Section III reveals the dual uplink form of the primal optimization problem via some necessary optimization transformations. An effective optimization algorithm based on instantaneous channel information is proposed in Section IV by exploiting the newly derived uplink-downlink duality. An extended effective optimization algorithm based on statistical channel information is further proposed in Section V to reduce the overhead of information exchange between coordinated BSs. In Section VI, numerical evaluations of these algorithms are carried out and conclusions are finally drawn in Section VII.

The following notations are used throughout this paper. Bold lowercase and uppercase letters represent column vectors and matrices, respectively. The superscripts $(\cdot)^H$, and $(\cdot)^{-1}$ represent the conjugate transpose operator, and the matrix inverse, respectively. $\|A\|$ denotes the Euclidean norm for vectors and the Frobenius norm of matrix A . The probability density function (pdf) of a circular complex Gaussian random vector with mean μ and covariance matrix Σ is denoted as $\mathcal{CN}(\mu, \Sigma)$. The function $\text{ceil}(x)$ rounds the elements of x to the nearest integers greater than or equal to x . The function $\text{floor}(x)$ rounds the elements of x to the nearest integers less than x . $(x, y)_+ = \max(x, y)$. $\mathbf{1}_N$ denotes N dimensions column vector with all elements equal one and $\xrightarrow{a.s.}$ denotes almost sure convergence.

II. SYSTEM MODEL

Consider a K -cell MU-MISO downlink system where the BS in cell j ($j = 1, \dots, K$) is equipped with M_j transmit antennas and serves N_j single-antenna user. We assume $M_j \geq N_j, \forall j$, thus, user scheduling is not taken into account. Let the k -th user in cell j be User- (j, k) and the BS in cell m be BS- m , respectively. Then, the received signal of the User- (j, k) is denoted as

$$y_{j,k} = \sum_{m=1}^K \mathbf{h}_{m,j,k}^H \sum_{n=1}^{N_j} \sqrt{p_{m,n}} \mathbf{w}_{m,n} x_{m,n} + z_{j,k}, \quad (1)$$

where $\mathbf{h}_{m,j,k} \in \mathbb{C}^{M_m \times 1} = \sigma_{m,j,k} \tilde{\mathbf{h}}_{m,j,k}$ denotes the flat fading channel coefficient from BS- m to User- (j, k) ¹, in which $\sigma_{m,j,k}$

¹For time division duplex (TDD) system, the transmitters can estimate the channels from the sounding signals received in the reverse link. For frequency division duplex (FDD) system, the transmitters are provided with quantized CSI via feedback.

represents the large-scale channel information from BS- m to User- (j, k) , and $\bar{\mathbf{h}}_{m,j,k}$ has i.i.d. complex entries of zero mean and variance $\frac{1}{M_m}$ [26], [28], $\mathbf{w}_{j,k} \in \mathbb{C}^{M_m \times 1}$ denotes the unit-norm beamforming vector for User- (j, k) , $p_{j,k}$ and $x_{j,k}$ denote respectively the transmit power and the information signal intended for User- (j, k) , and $z_{j,k}$ is a zero-mean circularly symmetric complex Gaussian random noise with power spectrum density $\omega_{j,k}^2$. Without loss of generality, in the rest of our paper, we assume that each BS serves the same number of users in each cell, i.e., $N_j = N, \forall j^2$.

For notational convenience, define $\bar{\mathbf{h}}_{n,m} = \frac{\mathbf{h}^{\lceil \frac{n}{N} \rceil, \lfloor \frac{m}{N} \rfloor, m-N \lfloor \frac{m}{N} \rfloor}}{\sqrt{W\omega^{\lceil \frac{n}{N} \rceil, m-N \lfloor \frac{m}{N} \rfloor}}}$ as the normalized channel flat fading coefficient from BS- $\lceil \frac{n}{N} \rceil$ to User- $(\lceil \frac{m}{N} \rceil, m - N \lfloor \frac{m}{N} \rfloor)$, where W represents the occupied bandwidth, $\lceil \cdot \rceil$ and $\lfloor \cdot \rfloor$ denote respectively the function *ceil*(\cdot) and the function *floor*(\cdot), $n, m = 1, \dots, \bar{N}$, $\bar{N} = KN$. Then, the instantaneous rate of User- $(\lceil \frac{m}{N} \rceil, m - N \lfloor \frac{m}{N} \rfloor)$ is calculated as

$$\vec{R}_m = W \log(1 + \vec{\gamma}_m), \quad (2)$$

where and $\vec{\gamma}_m$ denotes the SINR of User- $(\lceil \frac{m}{N} \rceil, m - N \lfloor \frac{m}{N} \rfloor)$ and is given as

$$\vec{\gamma}_m = \frac{p_m \|\bar{\mathbf{h}}_{m,m}^H \mathbf{w}_m\|^2}{\sum_{n=1, n \neq m}^{\bar{N}} p_n \|\bar{\mathbf{h}}_{n,m}^H \mathbf{w}_n\|^2 + 1}, \quad (3)$$

with $p_m = P^{\lceil \frac{m}{N} \rceil, m-N \lfloor \frac{m}{N} \rfloor}$ and $\mathbf{w}_m = \mathbf{w}^{\lceil \frac{m}{N} \rceil, m-N \lfloor \frac{m}{N} \rfloor}$ denoting the transmit power and transmit beamforming vector for User- $(\lceil \frac{m}{N} \rceil, m - N \lfloor \frac{m}{N} \rfloor)$, respectively. In this paper, the EE of interest is defined as the ratio of the weighted sum rate to the total power consumption, given by

$$\vec{f}(\mathbf{w}, \mathbf{p}) = \frac{\sum_{m=1}^{\bar{N}} \alpha_m \vec{R}_m}{\xi \sum_{m=1}^{\bar{N}} p_m + \sum_{j=1}^K (M_j P_c + P_0)}, \quad (4)$$

where \mathbf{w} denotes the collection of all beamforming vectors, $\mathbf{p} = [p_1, \dots, p_m, \dots, p_{\bar{N}}]^T$, $\xi \geq 1$ is a constant which accounts for the inefficiency of the power amplifier, P_c is the constant circuit power consumption per antenna which includes the power dissipations in the transmit filter, the mixer, the frequency synthesizer, and the digital-to-analog converter, P_0 is the basic power consumed at the BS which is independent of the number of transmit antennas³ [36], $\alpha_m = \alpha^{\lceil \frac{m}{N} \rceil, m-N \lfloor \frac{m}{N} \rfloor}$ denotes the

²The developed algorithms can be implemented in the case where different BSs serve different number of users in each cell, i.e., $N_i \neq N_j, i \neq j$.

³Note that in communication system, a higher R_m implies a higher code rate or higher order modulation being used, thereby may increase the power consumption of the encoder and decoder circuits. This leads to that the power consumption in the baseband is a function of R_m , which is usually nonlinear and makes the EE optimization problem extremely complex. For simplicity, similar to the existing literature [7]–[17], here this fact is ignored and we assume that the power consumption at the baseband is a constant and is also included in P_0 . This is reasonable due to the fact that the variation of power consumption for different levels of code rates or modulations is in general marginal in contrast to the basic power consumption.

weighting factor that is used to represent the priority of User- $(\lceil \frac{m}{N} \rceil, m - N \lfloor \frac{m}{N} \rfloor)$ in the system, which is typically included to achieve a certain fairness index among the users. For example, to obtain the proportional fairness, we can set $\alpha_m = \frac{1}{\bar{R}_m}$, where \bar{R}_m is the average data rate of User- $(\lceil \frac{m}{N} \rceil, m - N \lfloor \frac{m}{N} \rfloor)$ in the previous time slots [42], [43].

In order to balance the SE and the EE, our goal is to maximize the EE of the K -cell MU-MISO downlink transmission system subject to some predefined user rate requirements and per-BS transmit power constrains, i.e.,

$$\begin{aligned} & \max_{\mathbf{w}, \mathbf{p}} \vec{f}(\mathbf{w}, \mathbf{p}) \\ & s.t. \quad \sum_{m=(j-1)N+1}^{jN} p_m \leq P_j, \forall j \in \mathbb{K}, \\ & \quad \vec{\gamma}_m \geq \gamma_m, \|\mathbf{w}_m\| = 1, p_m \geq 0, \forall m \in \mathbb{U}, \end{aligned} \quad (5)$$

where $\mathbb{K} = \{1, \dots, K\}$, $\mathbb{U} = \{1, \dots, \bar{N}\}$, P_j is the transmit power constraint of BS- j , and γ_m represents the target SINR of User- $(\lceil \frac{m}{N} \rceil, m - N \lfloor \frac{m}{N} \rfloor)$. Note that, different from the EE optimization problem considered in the existing literature [14]–[19], the per-user minimum rate requirements are considered in our problem formulation making more challenging. In what follows, we attempt to find its solution from the perspective of uplink-downlink duality and then design an effective algorithm. Also, it is worth mentioning that though there is a feasibility issue in the above optimization problem (5), it is similar to the power minimization problem subject to a given minimum target rate demand set as in [34]. Throughout this paper, we assume that the set of per-user target rate requirements is feasible and only focus on finding the solution to the problem.

III. PROBLEM TRANSFORMATIONS

It is easy to understand that problem (5) is non-convex in the optimization variables and therefore it is difficult to find the globally optimal solution, since the user rate is non-convex in general for the MISO interference channel. Furthermore, the objective function in a fractional form brings more obstacles to solve the problem directly. In what follows, we focus on finding an efficient optimization method to solve problem (5) by extending the conventional uplink-downlink duality with reformulating it into the following form based on the per-BS constraints.

$$\begin{aligned} & \max_{\mathbf{w}, \mathbf{p}, \beta} \vec{g}(\mathbf{w}, \mathbf{p}, \beta) \\ & s.t. \quad \sum_{m=(j-1)N+1}^{jN} p_m \leq P_j, \forall j \in \mathbb{K}, \\ & \quad \vec{\gamma}_m \geq \gamma_m, \|\mathbf{w}_m\| = 1, p_m \geq 0, \forall m \in \mathbb{U}, \\ & \quad \sum_{m=1}^{\bar{N}} p_m \leq \beta \sum_{j=1}^K P_j, \beta^* \leq \beta \leq 1, \end{aligned} \quad (6)$$

where β is a newly introduced auxiliary variable, β^* denotes the ratio of the minimum total transmit power that needs to satisfy the target SINR constraints to the total power constraints⁴, and

$$\vec{g}(\mathbf{w}, \mathbf{p}, \beta) = \frac{\sum_{m=1}^{\bar{N}} \alpha_m \vec{R}_m}{\xi \beta \sum_{j=1}^K P_j + \sum_{j=1}^K (M_j P_c + P_0)}. \quad (7)$$

The equivalent relation between problem (5) and problem (6) is given by the following theorem.

Theorem 1: Problems (5) and (6) have the same optimal objective value and the same optimal solutions.

Proof: Let \mathbf{w}^{opt} and \mathbf{p}^{opt} be the optimal solution to problem (5) and let $\tilde{\mathbf{w}}^{opt}$, $\tilde{\mathbf{p}}^{opt}$ and $\tilde{\beta}^{opt}$ be the optimal solution to problem (6), respectively. One can easily prove by contradiction that $\sum_{m=1}^{\bar{N}} \tilde{p}_m^{opt} = \tilde{\beta}^{opt} \sum_{j=1}^K P_j$. Otherwise, $\tilde{\beta}^{opt}$ can be scaled down to achieve a better objective value. On the one hand, it is easy to know that $\tilde{\mathbf{w}}^{opt}$ and $\tilde{\mathbf{p}}^{opt}$ are also a feasible solution to problem (5) due to the fact that the constraint sets of problem (5) is a subset of the constraint sets of problem (6). On the other hand, let $\check{\beta} = \frac{1}{\sum_{j=1}^K P_j} \sum_{m=1}^{\bar{N}} p_m^{opt}$, recalling

the per-BS transmit power constraints, we have $0 \leq \check{\beta} \leq 1$ and easily know that \mathbf{w}^{opt} and \mathbf{p}^{opt} are also a feasible solution to problem (6) in conjunction with the rate constraints. With these two observations, we can prove by contradiction that $\vec{f}(\mathbf{w}^{opt}, \mathbf{p}^{opt}) = \vec{g}(\tilde{\mathbf{w}}^{opt}, \tilde{\mathbf{p}}^{opt}, \tilde{\beta}^{opt})$ must hold and the same objective is achieved with the same beamforming and power solution. Otherwise, the objective with a lower value can be improved by replacing the beamforming matrix and power vector with the solution of the other problem, which contradicts the assumption that these two objective values are both optimal ones. ■

It is easy to verify that the achievable objective value of (6) monotonically increases with the auxiliary variable β in the range $\beta^* \leq \beta \leq \tilde{\beta}^{opt}$, and monotonically decreases with the auxiliary variable β in the range $\tilde{\beta}^{opt} \leq \beta \leq 1$. Thus the optimal value of β can be efficiently obtained through a one dimension search method such as the simulated annealing algorithm, or the hill-climbing method in the range $\beta^* \leq \beta \leq 1$ [44], [45]. Based on these analysis, in what follows, the original EE optimization problem is first transformed into a sum-rate maximization problem subject to per-BS transmit power constraints, per-user target rate requirements and additional sum power consumption, i.e., (8), where the value of the auxiliary variable β is fixed. Then, we solve iteratively (8) by updating β via a one dimension search method to obtain the solution to the energy efficiency optimization problem (6).

⁴The value of β^* can be determined by solving a power minimization problem (PMP) subject to the same per-BS transmit power constraints and the same per-user minimum rate demand constraints [34]. Note that the solution to the PMP is not necessarily the solution to problem (5) [17], [18].

$$\begin{aligned} & \max_{\mathbf{w}, \mathbf{p}} \sum_{m=1}^{\bar{N}} \alpha_m \vec{R}_m \\ & s.t. \quad \sum_{m=(j-1)N+1}^{jN} p_m \leq P_j, \forall j \in \mathbb{K}, \\ & \quad \vec{\gamma}_m \geq \gamma_m, \|\mathbf{w}_m\| = 1, p_m \geq 0, \forall m \in \mathbb{U}, \\ & \quad \sum_{m=1}^{\bar{N}} p_m \leq \beta \sum_{j=1}^K P_j. \end{aligned} \quad (8)$$

In contrast to the conventional WSR maximization (WSRMax) problem, problem (8) has additional total power constraint and minimum SINR requirements, which is named as extended WSRMax problem, making the problem more difficult to solve due to the coupling between the optimization variables. In order to design an effective algorithm, in the sequel, the uplink-downlink duality is applied to solve problem (8) which can be rewritten as the following problem.

$$\begin{aligned} & \psi(\boldsymbol{\lambda}) = \max_{\mathbf{w}, \mathbf{p}} \sum_{m=1}^{\bar{N}} \alpha_m \vec{R}_m \\ & s.t. \quad \vec{\gamma}_m \geq \gamma_m, \|\mathbf{w}_m\| = 1, p_m \geq 0, \forall m \in \mathbb{U}, \\ & \quad \sum_{m=1}^{\bar{N}} \tilde{\lambda}_m p_m \leq \sum_{j=1}^K (\lambda_j + \lambda_{K+1} \beta) P_j, \end{aligned} \quad (9)$$

where $\boldsymbol{\lambda} = \{\lambda_j \geq 0, \sum_{j=1}^{K+1} \lambda_j \neq 0\}$ are auxiliary variables, and $\tilde{\lambda}_m = \lambda_{\lceil \frac{m}{N} \rceil} + \lambda_{K+1}, \forall m$. Note that a feasible solution to problem (8) is also feasible for problem (9). It also means that the optimal objective value of problem (9) for any given $\boldsymbol{\lambda} \geq \mathbf{0}$, is an upper bound on the optimal objective value of problem (8). Furthermore, it has been shown in [46], [47] that the solution to problem (8) can be equivalently achieved by solving the problem $\min_{\boldsymbol{\lambda}} \psi(\boldsymbol{\lambda})$. One can easily verify that the function $\psi(\boldsymbol{\lambda})$ is a convex function but is not necessarily differentiable⁵, so we can solve the problem $\min_{\boldsymbol{\lambda}} \psi(\boldsymbol{\lambda})$ via the sub-gradient algorithm or ellipsoid algorithm. Furthermore, it is shown in [48] that such an algorithm is guaranteed to converge to its optimal solution. The sub-gradient of the function $\psi(\boldsymbol{\lambda})$ is given by the following proposition, that can be proved with a similar method as used in [46].

Proposition 1: The sub-gradient \mathbf{g} of the function $\psi(\boldsymbol{\lambda})$ with fixed parameters $\bar{\boldsymbol{\lambda}}$

$$\mathbf{g}_j = \begin{cases} P_j - \sum_{m=(j-1)N+1}^{jN} p_m^{(*)}, & j = 1, \dots, K, \\ \beta \sum_{j=1}^K P_j - \sum_{m=1}^{\bar{N}} p_m^{(*)} & j = K + 1, \end{cases} \quad (10)$$

⁵Note that the interior point method requires the twice continuously differentiable property, however, the function $\psi(\boldsymbol{\lambda})$ is a convex function but is not necessarily differentiable, so the interior point method cannot be used to solve the problem of interest.

where $\mathbf{p}^{(*)}$ is the optimal solution of problem (9) with fixed parameters $\lambda = \bar{\lambda}$ and β .

The conventional uplink-downlink duality theory in [49], [50] shows that problem (9) is dual to the following virtual uplink problem:

$$\begin{aligned} \max_{\mathbf{w}, \mathbf{q}} \quad & \sum_{m=1}^{\bar{N}} \alpha_m \overleftarrow{R}_m \\ \text{s.t.} \quad & \overleftarrow{\gamma}_m \geq \gamma_m, \|\mathbf{w}_m\| = 1, q_m \geq 0, \forall m \in \mathbb{U}, \\ & \sum_{m=1}^{\bar{N}} q_m \leq \sum_{j=1}^K (\lambda_j + \beta \lambda_{K+1}) P_j, \end{aligned} \quad (11)$$

in which $\mathbf{q} \in \mathbb{R}_+^{\bar{N} \times 1}$ denotes the virtual uplink transmit power vector with $q_m = q_{\lceil \frac{m}{N} \rceil, m-N \lfloor \frac{m}{N} \rfloor}$, and $\overleftarrow{R}_m = \overleftarrow{R}_{\lceil \frac{m}{N} \rceil, m-N \lfloor \frac{m}{N} \rfloor}$ denotes the virtual uplink rate of User- $(\lceil \frac{m}{N} \rceil, m-N \lfloor \frac{m}{N} \rfloor)$ and is given by

$$\overleftarrow{R}_m = \mathcal{W} \log(1 + \overleftarrow{\gamma}_m), \quad (12)$$

where $\overleftarrow{\gamma}_m = \overleftarrow{\gamma}_{\lceil \frac{m}{N} \rceil, m-N \lfloor \frac{m}{N} \rfloor}$ denotes the virtual uplink SINR of User- $(\lceil \frac{m}{N} \rceil, m-N \lfloor \frac{m}{N} \rfloor)$ and is calculated as

$$\overleftarrow{\gamma}_m = \frac{q_m \|\bar{\mathbf{h}}_{m,m}^H \mathbf{w}_m\|^2}{\sum_{n=1, n \neq m}^{\bar{N}} q_n \|\bar{\mathbf{h}}_{m,n}^H \mathbf{w}_m\|^2 + \tilde{\lambda}_m}. \quad (13)$$

According to the uplink-downlink duality theory [46], [47], the uplink and the downlink have the same achievable SINR region with the same beamforming vectors and the same total transmit power constraint. This means that once the uplink problem is solved, the beamforming solution to the downlink problem can be obtained at the same time, while the downlink power solution remains to be calculated. Using the fact that the uplink and the downlink achieve the same SINR, we have

$$\begin{aligned} & \frac{p_m \|\bar{\mathbf{h}}_{m,m}^H \mathbf{w}_m\|^2}{\sum_{n=1, n \neq m}^{\bar{N}} p_n \|\bar{\mathbf{h}}_{n,m}^H \mathbf{w}_n\|^2 + 1} \\ &= \frac{q_m \|\bar{\mathbf{h}}_{m,m}^H \mathbf{w}_m\|^2}{\sum_{n=1, n \neq m}^{\bar{N}} q_n \|\bar{\mathbf{h}}_{m,n}^H \mathbf{w}_m\|^2 + \tilde{\lambda}_m}. \end{aligned} \quad (14)$$

By rearranging the terms in (14), we have

$$\begin{aligned} & q_m \sum_{n=1, n \neq m}^{\bar{N}} p_n \|\bar{\mathbf{h}}_{n,m}^H \mathbf{w}_n\|^2 + q_m \\ &= p_m \sum_{n=1, n \neq m}^{\bar{N}} q_n \|\bar{\mathbf{h}}_{m,n}^H \mathbf{w}_m\|^2 + p_m \tilde{\lambda}_m. \end{aligned} \quad (15)$$

Adding the term $q_m p_m \|\bar{\mathbf{h}}_{m,m}^H \mathbf{w}_m\|^2$ on both sides of (15) and summing with respect to the subscript m , we have

$$\sum_{m=1}^{\bar{N}} \tilde{\lambda}_m p_m = \sum_{m=1}^{\bar{N}} q_m. \quad (16)$$

The above equalities can be used to calculate the downlink power solution from the corresponding uplink solution.

IV. JBOPA ALGORITHM WITH INSTANTANEOUS CHANNEL INFORMATION

In this section, we focus on finding the solution to the rate constrained EE maximization problem (5). Section III has revealed that solving (5) is equivalent to iteratively solving the extended WSRMax problem (8) for fixed β and searching for the optimal value of β , which can be further recast to solve the virtual uplink problem (11).

A. Algorithm Design

It is well known that the optimal receiver beamforming vector that maximizes the uplink SINR expressed in (13) is the MMSE filter, which is in general calculated as⁶ [50]

$$\mathbf{w}_m = \frac{(\boldsymbol{\Sigma}_m + \tilde{\lambda}_m \mathbf{I})^{-1} \bar{\mathbf{h}}_{m,m}}{\|(\boldsymbol{\Sigma}_m + \tilde{\lambda}_m \mathbf{I})^{-1} \bar{\mathbf{h}}_{m,m}\|}, \forall m \in \mathbb{U}, \quad (17)$$

where $\boldsymbol{\Sigma}_m = \sum_{n=1, n \neq m}^{\bar{N}} q_n \bar{\mathbf{h}}_{m,n} \bar{\mathbf{h}}_{m,n}^H$. In what follows, we focus on the maximization over \mathbf{q} of problem (11) for fixed \mathbf{w} and β that can be formulated as follows:

$$\begin{aligned} \min_{\mathbf{q}, \tilde{\gamma}} \quad & \prod_{m=1}^{\bar{N}} (1 + \tilde{\gamma}_m)^{-\alpha_m} \\ \text{s.t.} \quad & \overleftarrow{\gamma}_m \geq \tilde{\gamma}_m, \overleftarrow{\gamma}_m \geq \gamma_m, q_m \geq 0, \forall m \in \mathbb{U}, \\ & \sum_{m=1}^{\bar{N}} q_m \leq \sum_{j=1}^K (\lambda_j + \beta \lambda_{K+1}) P_j, \end{aligned} \quad (18)$$

⁶Note that though the constants P_0 , P_c and ζ which model the power consumption in transceiver circuits and the power amplifier efficiency affect the energy efficiency through the denominator of the energy efficiency defined in (4), P_0 and P_c are independent of the expression of the normalized beamformer \mathbf{w} , as shown in (17). In other words, they are only related with the search over β and thus has an impact on the value of energy efficiency. It also means that a possible mismatch between the true value of P_c , P_0 and the value that is used to design the proposed beamformer only impacts the value of energy efficiency but does not impact the expression of the beamformer. Furthermore, it is worth mentioning that the considered problem (5) is more complex than the energy efficient design in [15]–[19] where the user rate requirements is not taking into account. Different from them, we solve the complex energy efficiency optimization problem from the perspective of uplink-downlink duality by iteratively updating an auxiliary variable. In addition, we would like to point out that the solution of the beamforming vector \mathbf{w} given by (17) is the optimal receiver to the dual problem (11) of the extended WSRMax problem with a fixed β , as a result ζ is not included in the structure of the normalized beamformer \mathbf{w} .

where $\tilde{\boldsymbol{\gamma}}$ denotes the collection of the auxiliary variables $\{\tilde{\gamma}_m\}$. The optimization problem (18) is a signomial optimization problem and is non-convex in general. Fortunately, note that, all constraints are monomial function in problem (18). If therefore the objective function can reformulated as a monomial function, problem (18) become a GP problem in standard form. Motivated by this observation, a successive convex approximation approach is first presented in the following proposition, for the proof please see [52].

Lemma 1: Let $\phi(\gamma) = \kappa \gamma^\vartheta$ be a function used to approximate function $\varphi(\gamma) = 1 + \gamma$, near the point $\hat{\gamma}$. The parameters κ and ϑ of the best monomial local approximation are given by

$$\vartheta = \frac{\hat{\gamma}}{1 + \hat{\gamma}}, \kappa = \frac{1 + \hat{\gamma}}{\hat{\gamma}^\vartheta}, \quad (19)$$

and $\phi(\gamma) < \varphi(\gamma), \forall \gamma > 0$.

To overcome the non-convex nature of problem (18), we start with reformulating problem (18) by applying the local approximation of Lemma 1 in its objective function, yielding

$$\begin{aligned} & \min_{\mathbf{q}, \tilde{\boldsymbol{\gamma}}} \prod_{m=1}^{\bar{N}} \tilde{\gamma}_m^{-\alpha_m} \frac{\tilde{\gamma}_m}{1 + \tilde{\gamma}_m} \\ & \text{s.t. } \tilde{\gamma}_m \geq \tilde{\gamma}_m, \tilde{\gamma}_m \geq \gamma_m, q_m \geq 0, \forall m \in \mathbb{U}, \\ & \sum_{m=1}^{\bar{N}} q_m \leq \sum_{j=1}^K (\lambda_j + \beta \lambda_{K+1}) P_j, \\ & (1 - \delta) \hat{\gamma}_m \leq \tilde{\gamma}_m \leq (1 + \delta) \hat{\gamma}_m, \forall m \in \mathbb{U}, \end{aligned} \quad (20)$$

where δ is a small constant used to control the desired approximation accuracy, typically set to be $\delta = 0.1$. Notice that problem (20) is lower boundary problem of the original problem (18) that is difficult to solve directly. It has been proven in [53] that problem (20) can be reformulated as a convex GP problem in standard form by using jointly the logarithmic change of the variable and a logarithmic transformation of the objective function and constraints, respectively. In other words, problem (20) can be easily solved by using the powerful GP optimization tool packages [54]. Furthermore, the results in [52] have shown that by starting from an initial point, we can search for a close local optimum by solving a sequence of GPs that locally approximate the original problem (18).

Based on the above analysis, an effective JBOPA optimization algorithm is developed to solve the virtual uplink problem (8) with fixed β and is summarized as Algorithm 1⁷. According to the uplink-downlink duality, it is known that the optimal uplink beamforming vectors \mathbf{w} also serve as the optimal downlink ones, while the optimal downlink power allocation \mathbf{p} can be calculated from the uplink solution. The details are discussed in the following subsection.

Remark 1: Note that when the value of β equals to one, Algorithm 1 is equivalent to solving the conventional WSRMax problem. In order to solve the original EE maximization

⁷The optimal value of β can be efficiently obtained through a search method such as the simulated annealing algorithm or hill-climbing researching in the range of $\beta^* \leq \beta \leq 1$ [44], [45].

Algorithm 1. JBOPA Algorithm With Instantaneous Channel Information

- 1: Initialize the uplink transmit power $\mathbf{q}^{(0)}$ and the beamformers $\mathbf{w}^{(0)}$ such that the SINR constraints and the power constraint are satisfied, and calculate the approximation point $\tilde{\boldsymbol{\gamma}}^{(0)}$ with $\mathbf{q}^{(0)}$ and $\mathbf{w}^{(0)}$.
 - 2: Initialize the auxiliary variable $\boldsymbol{\lambda}^{(*)} \succeq \mathbf{0}$, $\mathbf{q}^{(*)} = \mathbf{0}$, $R^{(*)} = 0$, $\tilde{\boldsymbol{\gamma}}^{(*)} = \tilde{\boldsymbol{\gamma}}^{(0)}$, and $\mathbf{w}^{(*)} = \mathbf{w}^{(0)}$.
 - 3: Calculate the approximation constants with (19) and $\tilde{\boldsymbol{\gamma}}^{(*)}$. Solve problem (20) with $\tilde{\boldsymbol{\gamma}}^{(*)}$, $\mathbf{w}^{(*)}$, $\boldsymbol{\lambda}^{(*)}$, and β , then obtain $\mathbf{q}^{(**)}$ and $\tilde{\boldsymbol{\gamma}}^{(**)}$.
 - 4: If $\max_m \left| \tilde{\gamma}_m^{(**)} - \tilde{\gamma}_m^{(*)} \right| \leq \zeta$, then let $\tilde{\boldsymbol{\gamma}}^{(*)} = \tilde{\boldsymbol{\gamma}}^{(**)}$, $\mathbf{q}^{(*)} = \mathbf{q}^{(**)}$, and go to step 6, otherwise, let $\tilde{\boldsymbol{\gamma}}^{(*)} = \tilde{\boldsymbol{\gamma}}^{(**)}$, and go to step 3.
 - 5: Update the beamformers \mathbf{w} with $\mathbf{q}^{(*)}$, $\boldsymbol{\lambda}^{(*)}$ and (17), then obtain $\mathbf{w}^{(**)}$. Calculate the objective value of problem (18) with $\mathbf{w}^{(**)}$ and $\mathbf{q}^{(*)}$, then obtain $R^{(**)}$. If $|R^{(**)} - R^{(*)}| \leq \zeta$, let $\mathbf{w}^{(*)} = \mathbf{w}^{(**)}$, and go to step 6, otherwise let $\mathbf{w}^{(*)} = \mathbf{w}^{(**)}$, and go to step 3.
 - 6: Judge whether the stop criterion of the ellipsoid method is satisfied or not. If yes, then stop the loop, otherwise update the auxiliary variables $\boldsymbol{\lambda}$ with $\mathbf{w}^{(*)}$, $\mathbf{q}^{(*)}$ and $\tilde{\boldsymbol{\gamma}}^{(*)}$ by using the ellipsoid method, then obtain the updated $\boldsymbol{\lambda}^{(*)}$ and go to step 3.
-

problem, we need to iteratively run Algorithm 1 to obtain the solution of β by using a one dimension search method that usually only requires a limited number of iterations. Let \mathcal{C} be the computational complexity of Algorithm 1, the computational complexity of our proposed EE maximization algorithm is $O(N_\beta \mathcal{C})$ where N_β is the required number of iterations in searching for an optimal β [44], [45].

B. Downlink Transmit Power Computation

It is necessary to point out that the sub-gradient given in (10) is calculated based on a given downlink transmit power \mathbf{p} . Therefore, it is important to compute the downlink transmit power from the obtained uplink transmit power \mathbf{q} and transmit beamforming vector \mathbf{w} . Combining (14) with (16), we have

$$\begin{cases} \vec{\gamma}_m = \overleftarrow{\gamma}_m, \forall m \in \mathbb{U}, & (21a) \\ \sum_{m=1}^{\bar{N}} \tilde{\lambda}_m p_m = P_{max}, & (21b) \end{cases}$$

where $P_{max} = \sum_{m=1}^{\bar{N}} q_m$. After some basic manipulations, we have

$$\begin{cases} \mathbf{p} = \mathbf{D}\mathbf{G}\mathbf{p} + \mathbf{D}\mathbf{1}_{\bar{N}}, & (22a) \\ \sum_{m=1}^{\bar{N}} \tilde{\lambda}_m p_m = P_{max}, & (22b) \end{cases}$$

where matrices \mathbf{G} and \mathbf{D} are given by, respectively

$$[\mathbf{G}]_{m,n} = \begin{cases} 0, & m = n, \\ \|\bar{\mathbf{h}}_{n,m}^H \mathbf{w}_n\|^2 & m \neq n. \end{cases} \quad (23)$$

$$[\mathbf{D}]_{m,n} = \begin{cases} \frac{\bar{\gamma}_m}{\|\bar{\mathbf{h}}_{m,m}^H \mathbf{w}_m\|^2} & m = n, \\ 0, & m \neq n. \end{cases} \quad (24)$$

Multiplying both sides of (22a) by $\tilde{\boldsymbol{\lambda}}^T = [\tilde{\lambda}_1, \dots, \tilde{\lambda}_m, \dots, \tilde{\lambda}_N]^T$ and using (22), we have

$$1 = \frac{1}{P_{max}} \tilde{\boldsymbol{\lambda}}^T \mathbf{D} \mathbf{G} \mathbf{p} + \frac{1}{P_{max}} \tilde{\boldsymbol{\lambda}}^T \mathbf{D} \mathbf{1}_{\bar{N}}. \quad (25)$$

Defining an extended power vector $\tilde{\mathbf{p}} = \begin{bmatrix} \mathbf{p} \\ 1 \end{bmatrix}$ and an extended coupling matrix

$$\mathbf{Q}(\mathbf{w}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\gamma}}, P_{max}) = \begin{bmatrix} \mathbf{D} \mathbf{G} & \mathbf{D} \mathbf{1}_{\bar{N}} \\ \frac{1}{P_{max}} \tilde{\boldsymbol{\lambda}}^T \mathbf{D} \mathbf{G} & \frac{1}{P_{max}} \tilde{\boldsymbol{\lambda}}^T \mathbf{D} \mathbf{1}_{\bar{N}} \end{bmatrix}. \quad (26)$$

Thus, we have the following eigensystem by combining (22a) and (26):

$$\tilde{\mathbf{p}} = \mathbf{Q}(\mathbf{w}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\gamma}}, P_{max}) \tilde{\mathbf{p}}, \text{ with } \tilde{p}_{\bar{N}+1} = 1. \quad (27)$$

According to the conclusions in [51], we can easily obtain the optimal power vector $\tilde{\mathbf{p}}$ as the first components of the dominant eigenvector of $\mathbf{Q}(\mathbf{w}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\gamma}}, P_{max})$, which can be scaled so that its last component equals one.

C. Algorithm Convergence and Complexity Analysis

Notice that at each iteration the achieved virtual uplink energy efficiency is monotonically non-decreasing due to the fact that steps 3 and 5 in Algorithm 1 all aim to maximize the objective function in problem (7), i.e., $\bar{g}^{(0)} \leq \bar{g}^{(1)} \leq \dots \leq \bar{g}^{(n)} \leq \dots$. Since the achievable rate region under the sum power constraint is bounded, the convergence of Algorithm 1 is guaranteed by the monotonic convergence theorem [56]. Meanwhile, the studies in [46] have revealed that the convergence of the proposed update method based on the sub-gradient ellipsoid method for the auxiliary variables $\boldsymbol{\lambda}$ is guaranteed to minimize $\psi(\boldsymbol{\lambda})$. Thus, the convergence of the whole algorithm is guaranteed and the convergence speed of our proposed algorithm is similar with that of the proposed algorithm developed in [46]. It is easy to see that the major computational load involves steps 3 and 5 in Algorithm 1. The GP operation at step 3 fortunately has very fast convergence speed and low computational complexity [53]. The matrix inversion at step 5 can be computed efficiently with $\sum_{m=1}^K \mathcal{N}O(M_m^{2.736})$ [55].

V. JBOPA ALGORITHM WITH STATISTICAL CHANNEL INFORMATION

In this section, the joint beamforming optimization and power allocation are studied based on the statistical channel

information to reduce the frequency of power updates and the overhead of signal exchange between coordinated BSs⁸. The results in [26]–[28] have shown that the SINR of each user converges to different deterministic values when BS is equipped with a large number of transmit antennas. In what follows, these observations can be used to obtain an approximated power allocation problem related only to the statistical channel information which is used to optimize the transmit power. By plugging (17) in the expression of the virtual uplink SINR of User- $(\lceil \frac{m}{N} \rceil, m - N \lfloor \frac{m}{N} \rfloor)$, the asymptotic approximation for $\bar{\gamma}_m$ is given in the following Theorem 2.

Theorem 2: The instantaneous random variable $\bar{\gamma}_m$ can be approximated by a deterministic quantity $\bar{\Upsilon}_m$ such that $\bar{\Upsilon}_m - \bar{\gamma}_m \xrightarrow{a.s.} 0$ as the system dimension $M_{\lceil \frac{m}{N} \rceil} \rightarrow \infty$. $\bar{\Upsilon}_m$ is given by:

$$\bar{\Upsilon}_m = \frac{q_m \sigma_{m,m}^2}{\omega_m^2} M_{\boldsymbol{\Sigma}_m}(-\tilde{\lambda}_m), \quad (28)$$

where $\boldsymbol{\Sigma}_m = \sum_{n=1, n \neq m}^{\bar{N}} q_n \bar{\mathbf{h}}_{m,n} \bar{\mathbf{h}}_{m,n}^H$, and $M_{\boldsymbol{\Sigma}_m}(z)$ denotes the Stieltjes transform of the matrix $\boldsymbol{\Sigma}_m$ evaluated at the point z . Also, $\bar{\Upsilon}_m$ can be obtained by solving the following fixed-point equation

$$\bar{\Upsilon}_m = \frac{1}{\omega_m^2} \frac{q_m \sigma_{m,m}^2}{\tilde{\lambda}_m + \frac{1}{M_{\lceil \frac{m}{N} \rceil}} \sum_{\substack{n=1 \\ n \neq m}}^{\bar{N}} \frac{q_n \sigma_{m,n}^2 q_n \sigma_{m,n}^2}{q_m \sigma_{m,m}^2 \omega_n^2 + q_n \sigma_{m,n}^2 \omega_m^2 \bar{\Upsilon}_m}}. \quad (29)$$

Proof: Please refer to Appendix A. ■

In order to reduce the overhead of signal exchange between coordinated BSs and apply the conclusion obtained in Theorem 2, in the sequel, we resort to achieve the power allocation solution by solving an approximated new power allocation problem directly defined as (30) which only relies on the large scale fading coefficients in place of solving problem (11) which is related with the instantaneous CSI⁹.

$$\begin{aligned} & \max_q \sum_{m=1}^{\bar{N}} \alpha_m \log_2(1 + \bar{\Upsilon}_m) \\ & \text{s.t. } \bar{\Upsilon}_m \geq \gamma_m, q_m \geq 0, \forall m \in \mathbb{U}, \\ & \sum_{m=1}^{\bar{N}} q_m \leq \sum_{j=1}^K (\lambda_j + \beta \lambda_{K+1}) P_j. \end{aligned} \quad (30)$$

⁸In this subsection, both the number of transmit antennas M_j and the number of serving users per cell N go to infinity while the ratio (load factor) $\lim \frac{N}{M_j}$ remains bounded, i.e., the notation $M_j \rightarrow \infty$ denotes that both M_j and N become large, while $\liminf \frac{N}{M_j} > 0$ and $\limsup \frac{N}{M_j} < \infty, \forall j$ [26]–[28].

⁹In what follows, we assume that the target SINR $\gamma_m, \forall m$ are achievable in the communication system with statistical CSI.

Defining $\varphi_m = \frac{\sigma_{m,m}^2}{\omega_m^2} M_{\Sigma_m}(-\tilde{\lambda}_m)$, $\forall m$, problem (30) is then reformulated as follows:

$$\begin{aligned} \min_{\mathbf{q}} \quad & - \sum_{m=1}^{\bar{N}} \alpha_m \log_2(1 + q_m \varphi_m) \\ \text{s.t.} \quad & q_m \geq \frac{\gamma_m}{\varphi_m}, q_m \geq 0, \forall m \in \mathbb{U}, \\ & \sum_{n=1}^{\bar{N}} q_m \leq \sum_{j=1}^K (\lambda_j + \beta \lambda_{K+1}) P_j. \end{aligned} \quad (31)$$

It is easy to verify that problem (31) is convex and thus can be tackled using standard convex approaches [45]. Particularly, the power constraint meets with equality in the optimal solution. The corresponding Lagrange function is given by

$$\begin{aligned} \mathfrak{L}(\mathbf{q}, \boldsymbol{\xi}, \varsigma) = & - \sum_{m=1}^{\bar{N}} \alpha_m \log_2(1 + q_m \varphi_m) - \sum_{m=1}^{\bar{N}} \xi_m \left(q_m - \frac{\gamma_m}{\varphi_m} \right) \\ & + \varsigma \left(\sum_{n=1}^{\bar{N}} q_m - \sum_{j=1}^K (\lambda_j + \beta \lambda_{K+1}) P_j \right), \end{aligned} \quad (32)$$

where $\boldsymbol{\xi}$ and ς are respectively the Lagrange multipliers associated with the SINR constraints and the power constraint. The Karush-Kuhn-Tucker (KKT) conditions are then given by

$$\begin{cases} q_m \geq \frac{\gamma_m}{\varphi_m}, q_m \geq 0, \xi_m \geq 0, \varsigma \geq 0, \end{cases} \quad (33a)$$

$$\begin{cases} - \frac{\alpha_m \varphi_m}{(1 + q_m \varphi_m) \ln(2)} - \xi_m + \varsigma = 0, \end{cases} \quad (33b)$$

$$\begin{cases} \xi_m \left(q_m - \frac{\gamma_m}{\varphi_m} \right) = 0, \end{cases} \quad (33c)$$

$$\begin{cases} \sum_{n=1}^{\bar{N}} q_m = \sum_{j=1}^K (\lambda_j + \beta \lambda_{K+1}) P_j. \end{cases} \quad (33d)$$

Note that $\boldsymbol{\xi} \geq 0$ acts as a slack variable in the last equation and therefore can be removed. The above equations are rewritten as

$$\begin{cases} q_m \geq \frac{\gamma_m}{\varphi_m}, \varsigma \geq 0, \end{cases} \quad (34a)$$

$$\begin{cases} \frac{\alpha_m \varphi_m}{(1 + q_m \varphi_m) \ln(2)} \leq \varsigma, \end{cases} \quad (34b)$$

$$\begin{cases} \left(q_m - \frac{\gamma_m}{\varphi_m} \right) \left(\varsigma - \frac{\alpha_m \varphi_m}{(1 + q_m \varphi_m) \ln(2)} \right) = 0, \end{cases} \quad (34c)$$

$$\begin{cases} \sum_{n=1}^{\bar{N}} q_m = \sum_{j=1}^K (\lambda_j + \beta \lambda_{K+1}) P_j. \end{cases} \quad (34d)$$

Using a similar approach as in [45, pp. 245–246], we have

$$q_m = \begin{cases} \frac{\alpha_m}{\ln(2)\varsigma} - \frac{1}{\varphi_m}, & \varsigma < \frac{\alpha_m \varphi_m}{(1 + \gamma_m) \ln(2)}, \\ \frac{\gamma_m}{\varphi_m}, & \varsigma \geq \frac{\alpha_m \varphi_m}{(1 + \gamma_m) \ln(2)}, \end{cases} \quad (35)$$

or, simply expressed as $q_m = \left(\frac{\alpha_m}{\ln(2)\varsigma} - \frac{1}{\varphi_m}, \frac{\gamma_m}{\varphi_m} \right)_+$. Substituting this expression into the condition (34d), it yields

$$\sum_{m=1}^{\bar{N}} \left(\frac{\alpha_m}{\ln(2)\varsigma} - \frac{1}{\varphi_m}, \frac{\gamma_m}{\varphi_m} \right)_+ = \sum_{j=1}^K (\lambda_j + \beta \lambda_{K+1}) P_j. \quad (36)$$

It is seen that the power solution in the energy-efficient large-scale system is water-filling-like power allocation algorithm. As noted in Proposition 1 and Algorithm 1 the update of auxiliary variables $\boldsymbol{\lambda}$ needs to know the downlink transmit power \mathbf{p} obtained with $\tilde{\mathbf{p}} = \mathbf{Q}(\mathbf{w}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\gamma}}, P_{max}) \tilde{\mathbf{p}}$. A similar procedure for analyzing the instantaneous random variable $\tilde{\boldsymbol{\gamma}}_m$ can be applied to the matrices \mathbf{G} and \mathbf{D} such that the asymptotic approximation that only depends on the statistical CSI of the matrix $\mathbf{Q}(\mathbf{w}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\gamma}}, P_{max})$ can be obtained. The asymptotic approximation of \mathbf{G} and \mathbf{D} are described as the following Theorem.

Theorem 3: The entries of the instantaneous uplink-downlink power transformation matrices \mathbf{G} and \mathbf{D} can be approximated respectively by the entries of a deterministic uplink-downlink power transformation matrices \mathbb{G} and \mathbb{D} such that $[\mathbb{G}]_{m,n} - [\mathbf{G}]_{m,n} \xrightarrow{a.s.} 0$ and $[\mathbb{D}]_{m,n} - [\mathbf{D}]_{m,n} \xrightarrow{a.s.} 0$, as the system dimension $M_{\lceil \frac{\mu}{N} \rceil} \rightarrow \infty, \forall m, n$, the element $[\mathbb{G}]_{m,n}$ is given by

$$[\mathbb{G}]_{m,n} = \begin{cases} 0, & m = n, \\ \frac{\sigma_{n,m}^2}{\sigma_m^2 M_{\lceil \frac{\mu}{N} \rceil} \left(1 + \frac{q_m \sigma_{n,m}^2}{\omega_m^2} \phi_n(\mathbf{q}) \right)^2}, & m \neq n, \end{cases} \quad (37)$$

and the element $[\mathbb{D}]_{m,n}$ is given by (38), shown at the bottom of the page, where $\phi_m(\mathbf{q}) = \frac{\tilde{\gamma}_m \omega_m^2}{q_m \sigma_{m,m}^2} = M_{\Sigma_m}(-\tilde{\lambda}_m)$, $\forall m$.

Proof: Please refer to Appendix A. \blacksquare

In what follows, in order to realize the ideas of power allocation based on statistical channel information, the extended coupling matrix $\mathbf{Q}(\mathbf{w}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\gamma}}, P_{max})$ is approximated by the following matrix.

$$\mathbb{Q}(\mathbf{w}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\gamma}}, P_{max}) = \begin{bmatrix} \mathbb{D}\mathbb{G} & & & \\ \frac{1}{P_{max}} \tilde{\boldsymbol{\lambda}}^T \mathbb{D}\mathbb{G} & & & \\ & & \mathbb{D}\mathbb{I}_{\bar{N}} & \\ & & \frac{1}{P_{max}} \tilde{\boldsymbol{\lambda}}^T \mathbb{D}\mathbb{I}_{\bar{N}} & \end{bmatrix}. \quad (39)$$

$$[\mathbb{D}]_{m,n} = \begin{cases} \frac{\tilde{\gamma}_m}{\sigma_{m,m}^2 \phi_m(\mathbf{q})} \left(\tilde{\lambda}_m + \frac{1}{M_{\lceil \frac{\mu}{N} \rceil}} \sum_{l=1, l \neq m}^{\bar{N}} \frac{q_l \sigma_{m,l}^2}{\omega_l^2} \right), & m = n, \\ 0, & m \neq n, \end{cases} \quad (38)$$

Algorithm 2. JBOPA Algorithm With Statistical Channel Information

- 1: Initialize the uplink transmit power $\mathbf{q}^{(0)}$ by solving the PMP such that the SINR constraints and the power constraint are satisfied.
 - 2: Initialize the auxiliary variable $\lambda^{(*)}$, let $\hat{\Upsilon}^{(*)} = 0$ and $\mathbf{q}^{(*)} = \mathbf{q}^{(0)}$, and compute the objective values of problem (31) with $\mathbf{q}^{(*)}$ and $\lambda^{(*)}$, then obtain $\varrho^{(*)}$.
 - 3: Solve problem (36) with $\mathbf{q}^{(*)}$, $\lambda^{(*)}$ and β , and then obtain $\mathbf{q}^{(**)}$, and compute the objective values of problem (31) with $\mathbf{q}^{(**)}$ and $\lambda^{(*)}$, then obtain $\varrho^{(**)}$ and $\hat{\Upsilon}^{(**)}$. If $|\varrho^{(**)} - \varrho^{(*)}| \leq \eta$, let $\mathbf{q}^{(*)} = \mathbf{q}^{(**)}$, $\hat{\Upsilon}^{(*)} = \hat{\Upsilon}^{(**)}$ and then go to step 4, otherwise let $\mathbf{q}^{(*)} = \mathbf{q}^{(**)}$, $\hat{\Upsilon}^{(*)} = \hat{\Upsilon}^{(**)}$ and go to step 3.
 - 4: Judge whether satisfying the stop criterion of the ellipsoid method or not. If yes, the stop the loop and go to step 5, otherwise update the auxiliary variables λ with $\mathbf{q}^{(*)}$ and $\hat{\Upsilon}^{(*)}$ by using the ellipsoid method, then obtain the updated $\lambda^{(*)}$ and go to step 3.
 - 5: Calculate the beamformers \mathbf{w} with (17).
-

TABLE I

ACHIEVED SR COMPARISON, $P_j = 46$ dBm, $N_j = 2$, $\forall j$, $\mathcal{W} = 10$ MHz, $P_0 = 40$ dBm, $P_c = 30$ dBm, $\eta = 10^{-3}$, $\beta = 1$

$M_j, \forall j$	8	16	32	64	128
Algorithm 1	11.667	12.315	13.036	13.209	13.226
Algorithm 2	7.340	8.109	9.061	9.443	9.527
percentage(%)	62.91	65.85	69.51	71.49	72.03

Based on the above analysis, the power allocation problem (30) can be solved via the alternating optimization approach. To be more specific, we solve the power allocation problem (30) by sequentially fixing two of the three variables β , λ , \mathbf{q} and updating the third. The details of the alternating optimization algorithm based on duality theorem that is used to solve the power allocation problem (30) with fixed β based on statistical channel information is summarized as Algorithm 2, where ρ and ϱ denote the large-scale system EE and the objective value of problem (31), respectively.

Remark 2: In Algorithm 2, the power allocation is optimized based on statistical channel information and thereby is carried out in a long-term timescale, in which only the large-scale channel information and a few real numbers are exchanged between the BSs. While the beamforming vector \mathbf{w}_m can be updated based on local instantaneous channel state information, i.e., $\hat{\mathbf{h}}_{m,n}$, $\forall n$ and thereby on a short-term basis. Compared with Algorithm 1, the computational complexity can be significantly reduce due the long-term timescale power allocation update. In addition, Algorithm 2 only needs one time beamforming vector computation for each change of the instantaneous CSI and only needs the local CSI to compute the beamforming vector. Table I lists the achieved SR in terms of 10^7 bps of Algorithm 1 and Algorithm 2 for several antenna configurations over 1000 random channel realizations. It is easy to see that the

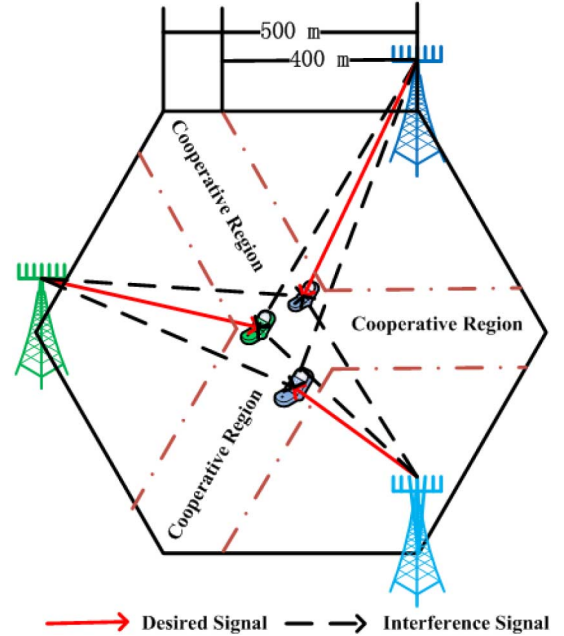


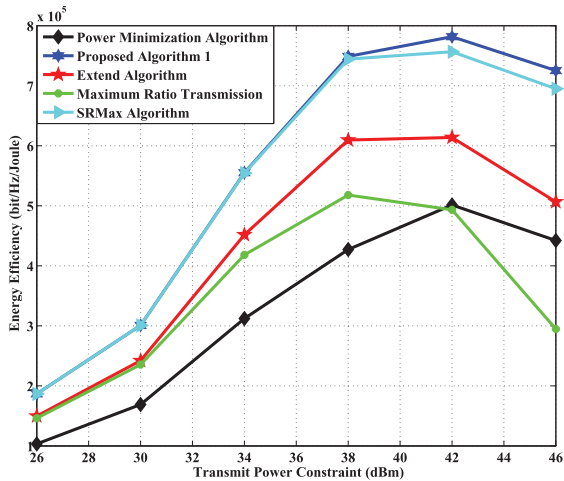
Fig. 1. Simulation Model.

difference between the achieved SR of Algorithm 1 and that of Algorithm 2 decreases with an increasing number of transmit antennas indicated by the percentage between the achieved SR of Algorithm 2 and that of Algorithm 1.

VI. NUMERICAL RESULTS

In this section, we investigate the performance of the proposed JBOPA algorithms via numerical simulations. We consider a cooperative cluster of $K = 3$ hexagonal adjacent cells each consisting of one BS and multiple users. In cell j , BS- j is equipped with M transmit antennas and serves N single antenna users. The cell radius is set to 500 m and each user randomly locates in the cooperation region and has a least 400 m distance from its serving BS as shown in Fig. 1. The channel vector $\mathbf{h}_{m,j,k}$ from BS- m to User- (j, k) is generated based on the formulation $\mathbf{h}_{m,j,k} \in \mathbb{C}^{M \times 1} = \sigma_{m,j,k} \mathbf{h}_{m,j,k}^w = \sqrt{\beta_{m,j,k}} \mathbf{h}_{m,j,k}^w$, where $\mathbf{h}_{m,j,k}^w$ denotes the small scale fading part and is assumed to be Gaussian distributed with zero mean and variance $\frac{1}{M}$, and $\beta_{m,j,k}$ denotes the large scale fading factor which in decibels is given as $10 \log_{10}(\beta_{m,j,k}) = -38 \log_{10}(d_{m,j,k}) - 34.5 + \eta_{j,k}$, where $\eta_{j,k}$ represents the log-normal shadow fading with zero mean and standard deviation 8 dB, $d_{m,j,k}$ denotes the distance between the BS- m and User- (j, k) [57]. The circuit power per antenna is $P_c = 30$ dBm, and the basic power consumed at the BS is $P_0 = 40$ dBm [36]. The noise power spectrum density $\omega_{j,k}^2 = -174$ dBm/Hz, $\forall j, k$ and the occupied bandwidth $\mathcal{W} = 10$ MHz. For simply, we assume that each BS has the same transmit power budget P , i.e., $P_j = P, \forall j$. The noise figure is 9 dB [4]. The weighting factor α and the inefficiency factor of power amplifier ξ are all set to unit.

For comparison, several related existing algorithms are also evaluated in our numerical simulations, including the

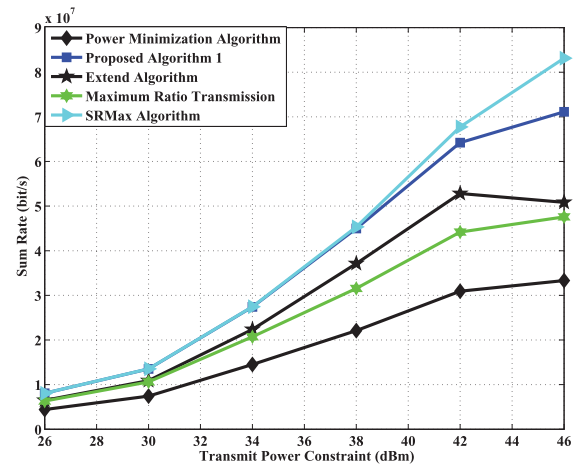
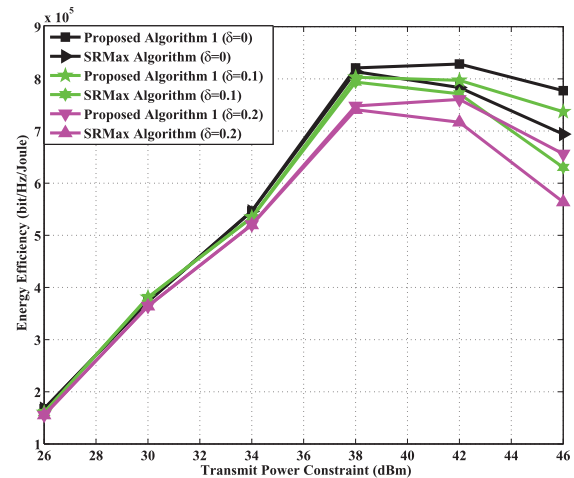
Fig. 2. Average EE Comparison, $M = 4$, $N = 2$, $\eta = 10^{-3}$.

power minimization algorithm that aims to minimize the transmit power subject to the same constraints, the conventional WSRMax algorithm that aims to maximize the SE with the same constraints, the performance of an extension of the algorithm developed in [15] that is described in Appendix D, and the normalized MRT beamforming transmission with equal power allocation at each BS and subject to no rate constraints. In our simulations, the target rate of each user is set respectively to be 70% of the user rate achieved by the normalized MRT beamforming with equal power allocation.

Fig. 2 illustrates the average EE of the above coordinated beamforming algorithms for finite-size system over 1000 random channel realizations. It can be seen that our proposed Algorithm 1 achieves the best EE performance among all algorithms, with a considerable gain at the high transmit power region. It is interesting to see that the Algorithm 1 and the SRMax algorithm achieve the same EE performance in the lower transmit power region such as 26–38 dBm. This implies that in this region it is possible to simultaneously achieve the optimal EE and SE, while in other regions a reasonable tradeoff between SE and EE is desired, which is the goal of our algorithm design. The results also show that our proposed Algorithm 1 outperforms the power minimization algorithm in all regions. This suggests that the power minimization subject to rate constraints in general cannot achieve the optimal EE, in which all the achieved user rates exactly meet with the constraints. In other words, it is very possible that achieving rates greater than the constraints also bring improved SE.

Fig. 3 illustrates the average WSR performances of some coordinated beamforming algorithms for finite-size system over 1000 random channel realizations. It can be seen that Algorithm 1 and the SRMax algorithm achieve the same SE performance in the lower transmit power region. In other words, the optimal EE algorithm also achieves the optimal system SE in the lower transmit power region, and vice versa. However, in terms of the SR, in the middle-high transmit region, the SRMax algorithm achieves the best WSR performance at the cost of a certain EE loss.

Fig. 4 illustrates the average EE of the proposed algorithm and the WSRMax algorithm achieved with imperfect CSI,

Fig. 3. Achieved WSR Comparison, $M = 4$, $N = 2$, $\eta = 10^{-3}$.Fig. 4. Average EE Versus Number of Transmit Antennas, $M = 4$, $N = 2$, $\eta = 10^{-3}$.

under configuration $\{M, N\} = \{4, 2\}$. In our numerical simulation, only an imperfect estimate $\hat{\mathbf{h}}_{m,j,k}$ of the true channel $\mathbf{h}_{m,j,k}$ is available at the transmitter which is modeled as [26]

$$\begin{aligned} \mathbf{h}_{m,j,k} \in \mathbb{C}^{M \times 1} &= \hat{\mathbf{h}}_{m,j,k} + \tilde{\mathbf{h}}_{m,j,k}^e \\ &= \sqrt{\beta_{m,j,k}} (1 - \delta^2) \hat{\mathbf{h}}_{m,j,k}^w + \delta \sqrt{\beta_{m,j,k}} \tilde{\mathbf{h}}_{m,j,k}^w \end{aligned} \quad (40)$$

where $\tilde{\mathbf{h}}_{m,j,k}^e$ denotes the estimation error, $\hat{\mathbf{h}}_{m,j,k}$ and $\tilde{\mathbf{h}}_{m,j,k}^e$ have i.i.d entries of zero mean and variance $\frac{1}{M}$ and independent of each other and independent of $z_{j,k}$, $\beta_{m,j,k}$ denotes the large scale fading factor which in decibels is given as $10 \log_{10}(\beta_{m,j,k}) = -38 \log_{10}(d_{m,j,k}) - 34.5 + \eta_{j,k}$, where $\eta_{j,k}$ represents the log-normal shadow fading with zero mean and standard deviation 8 dB, $d_{m,j,k}$ denotes the distance between the BS- m and User- (j, k) [57]. The parameter $\delta \in [0, 1]$ reflects the accuracy or quality of the channel estimate $\hat{\mathbf{h}}_{m,j,k}$, i.e., $\delta = 0$ corresponds to perfect channel state information, whereas for $\delta = 1$ the imperfect estimate $\hat{\mathbf{h}}_{m,j,k}$ is completely uncorrelated to the true channel. Numerical results show that the EE performance achieved by the related algorithms become worse with an increasing value of δ . It is see

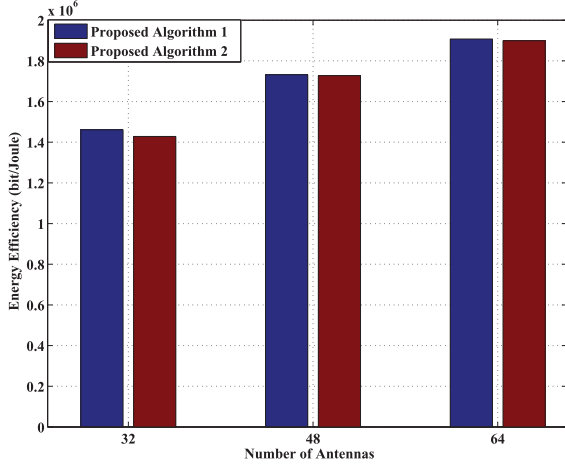


Fig. 5. Average EE Versus Number of Transmit Antennas, $M/N = 16$, $\eta = 10^{-3}$, $P_c = 10$ dBm, $P_0 = 20$ dBm.

that under different levels of imperfect CSI, the EE performance achieved by Algorithm 1 is always better than that of the WSRMax Algorithm at the high transmit power region.

Fig. 5 illustrates the performance comparison of the two developed coordinated beamforming algorithms versus the number of transmit antennas over 1000 random channel realizations, where the ratio between the number of transmit antennas and the number of served users is set to be 16. Numerical results show that the performance achieved by the proposed Algorithm 2 is very closed to that of the proposed Algorithm 1 with lower computational complexity and limited scale information between the coordinated BSs. This suggests that the proposed Algorithm 2 based on statistical state channel information is feasible for using the determination approximate expression of SINR to optimize the transmit power while the beamformers are calculated based on the instantaneous channel information.

VII. CONCLUSIONS

In this paper we have studied joint beamforming and power allocation in multicell downlink systems to maximize the system energy efficiency subject to per-BS transmit power constraints and per-user target SINR requirements. The considered objective function in fractional form is non-convex and it is hard to find the globally optimal solution. In order to obtain a tractable form, the primal problem was investigated from the viewpoint of the uplink-downlink duality by introducing some auxiliary variables. Furthermore, the uplink power allocation problem with fixed beamformers can be efficiently solved by using standard GP solvers. The convergence of the proposed JBOPA Algorithm 1 has been proven with the monotonic boundary theorem and the property of sub-gradient of convex problem. To reduce the overhead of the CSI collection, a new power allocation algorithm based on statistical channel information has been further developed. Numerical results have validated the effectiveness of our proposed schemes and shown that both SE and EE performance can be simultaneously improved

over traditional downlink coordinated schemes, especially in the middle-high transmit power region.

APPENDIX

A. Related Proof

Useful Results from Random Matrix Theory: We reproduce the following Lemmas [29]–[32].

Lemma 2: Let $\mathbf{x} \in \mathbb{C}^N$, independent and identically distributed (i.i.d.), with zero mean and variance $\frac{1}{N}$, $\mathbf{A} \in \mathbb{C}^{N \times N}$ Hermitian with bounded spectral norm whose elements are independent of \mathbf{x} , then

$$\mathbf{x}^H \mathbf{A} \mathbf{x} - \frac{1}{N} \text{tr}(\mathbf{A}) \xrightarrow[N \rightarrow \infty]{a.s.} 0. \quad (41)$$

Lemma 3: Consider an $N \times M$ random matrix $\mathbf{A} = (A_{i,j})_{i=1, j=1}^{N, M}$ whose the entries are given by: $A_{i,j} = \frac{\sigma_{i,j}}{\sqrt{M}} B_{i,j}$, where $B_{i,j}$ is i.i.d., with the following assumptions hold.

S1: The complex random variables $B_{i,j}$ are i.i.d. with $\mathbb{E}[B_{i,j}] = 0$, $\mathbb{E}[B_{i,j}^2] = 0$, $\mathbb{E}[|B_{i,j}|^2] = 1$ and $\mathbb{E}[|B_{i,j}|^8] < \infty$.

S2: There exists a real number $\sigma_{max} < \infty$ such that: $\sup_{M \geq 1} \max_{1 \leq i \leq N, 1 \leq j \leq M} |\sigma_{i,j}| \leq \sigma_{max}$

There exists a deterministic $N \times N$ matrix-value function $\Psi(z) = \text{diag}(\Psi_1(z), \dots, \Psi_N(z))$ analytic in $\mathbb{C} - \mathbb{R}_+$ such that:

$$\frac{1}{N} \text{tr}(\mathbf{A} \mathbf{A}^H - z \mathbf{I})^{-1} - \frac{1}{N} \text{tr} \Psi(z) \xrightarrow{a.s.} 0, \quad (42)$$

whose elements are the unique solutions of the deterministic system of $N + M$ equations

$$\Psi_i(z) = \frac{-1}{z \left(1 + \frac{1}{M} \sum_{j=1}^M \sigma_{i,j}^2 \Phi_j(z) \right)}, \quad 1 \leq i \leq N, \quad (43)$$

$$\Phi_j(z) = \frac{-1}{z \left(1 + \frac{1}{M} \sum_{i=1}^N \sigma_{i,j}^2 \Psi_i(z) \right)}, \quad 1 \leq j \leq M, \quad (44)$$

where $\frac{1}{N} \text{tr}(\Psi(z))$ is the Stieltjes transform [31] of the probability measure.

The values of $\Psi_i(z)$ and $\Phi_j(z)$, $\forall i, j$ can be obtained by initializing them to known values and iterating over equations (43) and (44) until their values converge. Furthermore, the differential of the Stieltjes transform of the matrix $\mathbf{A} \mathbf{A}^H$ can be calculated with the method used in [33].

Proof of Theorem 2: Plugging (13) in the expression of the virtual uplink SINR of User- $(\lceil \frac{m}{N} \rceil, m - N \lfloor \frac{m}{N} \rfloor)$, and employing Lemma 2 and Lemma 3, we have

$$\begin{aligned} \overleftarrow{\gamma}_m &= q_m \bar{\mathbf{h}}_{m,m}^H (\boldsymbol{\Sigma}_m + \tilde{\lambda}_m \mathbf{I})^{-1} \bar{\mathbf{h}}_{m,m} \\ &\xrightarrow[M \lceil \frac{m}{N} \rceil \rightarrow \infty]{a.s.} \frac{q_m \sigma_{m,m}^2}{M \lceil \frac{m}{N} \rceil \omega_m^2} \text{Tr} \left((\boldsymbol{\Sigma}_m + \tilde{\lambda}_m \mathbf{I})^{-1} \right) \\ &= \frac{q_m \sigma_{m,m}^2}{\omega_m^2} M_{\boldsymbol{\Sigma}_m}(-\tilde{\lambda}_m) \triangleq \overleftarrow{\Upsilon}_m, \end{aligned} \quad (45)$$

where $M_{\Sigma_m}(z)$ denotes the Stieltjes transform of the matrix Σ_m evaluated at the point z which can be computed using Lemma 3. Let $\phi_m(\mathbf{q}) = \frac{\tilde{\gamma}_m \omega_m^2}{q_m \sigma_{m,m}^2}$, we can obtain the following fixed-point equation by employing Lemma 3:

$$\phi_m(\mathbf{q}) = \frac{1}{\tilde{\lambda}_m + \frac{1}{M_{\lceil \frac{m}{N} \rceil}} \sum_{n=1, n \neq m}^{\bar{N}} \frac{q_n \sigma_{m,n}^2}{\omega_n^2 + q_n \sigma_{m,n}^2 \phi_m(\mathbf{q})}}. \quad (46)$$

This completes the proof of Theorem 2.

Proof of Theorem 3: To mitigate the impact of the small scale fading coefficients on the matrix \mathbf{G} , we firstly obtain the asymptotic determine expression of $\|\bar{\mathbf{h}}_{m,m}^H \mathbf{w}_m\|^2$ and $\|\bar{\mathbf{h}}_{n,m}^H \mathbf{w}_n\|^2$ based on the large dimension random theory. Then the asymptotic approximation of the uplink-downlink transformation matrix \mathbf{G} is obtained and is denoted as \mathbb{G} . Similar to the derivations of (45), we have the following two asymptotic determine expressions of $\|\bar{\mathbf{h}}_{m,m}^H \mathbf{w}_m\|^2$ and $\|\bar{\mathbf{h}}_{n,m}^H \mathbf{w}_n\|^2$ [28]:

$$\begin{aligned} \|\bar{\mathbf{h}}_{m,m}^H \mathbf{w}_m\|^2 &= \frac{(\bar{\mathbf{h}}_{m,m}^H (\Sigma_m + \tilde{\lambda}_m \mathbf{I})^{-1} \bar{\mathbf{h}}_{m,m})^2}{\bar{\mathbf{h}}_{m,m}^H (\Sigma_m + \tilde{\lambda}_m \mathbf{I})^{-2} \bar{\mathbf{h}}_{m,m}} \\ &\xrightarrow[M_{\lceil \frac{m}{N} \rceil} \rightarrow \infty]{a.s.} \frac{\sigma_{m,m}^2 (M_{\Sigma_m}(-\tilde{\lambda}_m))^2}{\omega_m^2 M'_{\Sigma_m}(-\tilde{\lambda}_m)} = \frac{\sigma_{m,m}^2 \phi_m^2(\mathbf{q})}{-\omega_m^2 \phi_m'(\mathbf{q})}, \end{aligned} \quad (47)$$

where $\Sigma_m = \sum_{n=1, n \neq m}^{\bar{N}} q_n \bar{\mathbf{h}}_{m,n} \bar{\mathbf{h}}_{m,n}^H$, $M_{\Sigma_m}(z)$ denotes the Stieltjes transform of the matrix Σ_m evaluated at the point z , and $M'_{\Sigma_m}(z)$ denotes the differential of $M_{\Sigma_m}(z)$ with respect to z evaluated at the point z and $\phi_m'(\mathbf{q})$ is given by

$$\phi_m'(\mathbf{q}) = \frac{-\phi_m(\mathbf{q})}{\tilde{\lambda}_m + \frac{1}{M_{\lceil \frac{m}{N} \rceil}} \sum_{n=1, n \neq m}^{\bar{N}} \frac{q_n \sigma_{m,n}^2}{\omega_n^2 (1 + \frac{q_n \sigma_{m,n}^2}{\omega_n^2} \phi_m(\mathbf{q}))^2}}. \quad (48)$$

By using jointly the Rank-1 perturbation Lemma [29] and Lemma 3, we have

$$\begin{aligned} \|\bar{\mathbf{h}}_{n,m}^H \mathbf{w}_n\|^2 &= \frac{\bar{\mathbf{h}}_{n,m}^H \Xi_n^{-1} \bar{\mathbf{h}}_{n,n} \bar{\mathbf{h}}_{n,n}^H \Xi_n^{-1} \bar{\mathbf{h}}_{n,m}}{\bar{\mathbf{h}}_{n,n}^H \Xi_n^{-2} \bar{\mathbf{h}}_{n,n}} \\ &= \frac{\bar{\mathbf{h}}_{n,m}^H \Theta_n^{-1} \bar{\mathbf{h}}_{n,n} \bar{\mathbf{h}}_{n,n}^H \Theta_n^{-1} \bar{\mathbf{h}}_{n,m}}{\bar{\mathbf{h}}_{n,n}^H \Xi_n^{-2} \bar{\mathbf{h}}_{n,n} (1 + q_m \bar{\mathbf{h}}_{n,m}^H \Theta_n^{-1} \bar{\mathbf{h}}_{n,m})^2} \\ &\xrightarrow[M_{\lceil \frac{n}{N} \rceil} \rightarrow \infty]{a.s.} \frac{\sigma_{n,m}^2}{\omega_m^2 M_{\lceil \frac{n}{N} \rceil} (1 + \frac{q_m \sigma_{n,m}^2}{\omega_m^2} M_{\Sigma_n}(-\tilde{\lambda}_n))^2} \\ &= \frac{\sigma_{n,m}^2}{\omega_m^2 M_{\lceil \frac{n}{N} \rceil} (1 + \frac{q_m \sigma_{n,m}^2}{\omega_m^2} M_{\Sigma_n}(-\tilde{\lambda}_n))^2} \\ &= \frac{\sigma_{n,m}^2}{\omega_m^2 M_{\lceil \frac{n}{N} \rceil} (1 + \frac{q_m \sigma_{n,m}^2}{\omega_m^2} \phi_n(\mathbf{q}))^2}, \end{aligned} \quad (49)$$

where $\Theta_n = \sum_{k=1, k \neq n, m}^{\bar{N}} q_k \bar{\mathbf{h}}_{n,k} \bar{\mathbf{h}}_{n,k}^H + \tilde{\lambda}_n \mathbf{I}$. Thus, the elements of the matrix \mathbb{G} can be calculated as follows:

$$[\mathbb{G}]_{m,n} = \begin{cases} 0, & m = n, \\ \frac{\sigma_{n,m}^2}{\omega_m^2 M_{\lceil \frac{n}{N} \rceil} (1 + \frac{q_m \sigma_{n,m}^2}{\omega_m^2} \phi_n(\mathbf{q}))^2}, & m \neq n. \end{cases} \quad (50)$$

$$[\mathbb{D}]_{m,n} = \begin{cases} \frac{\tilde{\gamma}_m}{\sigma_{m,m}^2 \phi_m^2(\mathbf{q})} & m = n, \\ -\omega_m^2 \phi_m'(\mathbf{q}) & m \neq n. \end{cases} \quad (51)$$

Plugging (48) in (51) yields the deterministic uplink-downlink power transformation matrix \mathbb{G} in Theorem 3. This completes the proof of Theorem 3.

B. JBOPA Initialization Algorithm With Instantaneous Channel Information

Before proceeding to propose an effective method to solve problem (11), we would like to discuss the feasibility of the target SINRs in the considered systems [51]. In general, the PMP subject to SINR constraints can be adopted to validate the feasibility of some given SINR subject to transmit power constraint in wireless communications. Herein, we propose to solve the PMP that is formulated as (17) for fixed feasible target SINRs, given by

$$\begin{aligned} \min_{\tilde{\mathbf{w}}, \varpi} \quad & \sum_{j=1}^K P_j \\ \text{s.t.} \quad & \frac{\|\bar{\mathbf{h}}_{m,m}^H \tilde{\mathbf{w}}_m\|^2}{\sum_{n=1, n \neq m}^{\bar{N}} \|\bar{\mathbf{h}}_{n,m}^H \tilde{\mathbf{w}}_n\|^2 + 1} \geq \gamma_m, \forall m \in \mathbb{U}, \\ & \sum_{m=(j-1)N+1}^{jN} \|\tilde{\mathbf{w}}_m\|^2 \leq \varpi P_j, \forall j. \end{aligned} \quad (52)$$

By using the similar derivation method used in [27], [50], we can obtain the dual problem of problem (14) that is given by the following optimization problem:

$$\begin{aligned} \max_v \min_{\mathbf{w}, \mathbf{q}} \quad & \sum_{m=1}^{\bar{N}} q_m \\ \text{s.t.} \quad & \bar{\gamma}_m \geq \gamma_m, q_m \geq 0, \forall m \in \mathbb{U}, \\ & \sum_{j=1}^K (1 - v_j) P_j = 0, \end{aligned} \quad (53)$$

where $\bar{\gamma}_m$ denotes the SINR of User- $(\lceil \frac{m}{N} \rceil, m - N \lfloor \frac{m}{N} \rfloor)$ and is calculated as

$$\bar{\gamma}_m = \frac{q_m \|\bar{\mathbf{h}}_{m,m}^H \mathbf{w}_m\|^2}{\sum_{n=1, n \neq m}^{\bar{N}} q_n \|\bar{\mathbf{h}}_{m,n}^H \mathbf{w}_n\|^2 + v_{\lceil \frac{m}{N} \rceil}}, \quad (54)$$

where q_m represents the dual uplink power of User- $(\lceil \frac{m}{N} \rceil, m - N \lfloor \frac{m}{N} \rfloor)$, $v_{\lceil \frac{m}{N} \rceil}$ denotes the virtual uplink noise variance of BS- $\lceil \frac{m}{N} \rceil$. Similarly, it is easily known that the maximization over \mathbf{w} in (53) can be solved explicitly for fixed power \mathbf{q} . In particular, the optimal receiver beamforming vector \mathbf{w}_k that maximizes the SINR is the MMSE filter. According to the Sherman-Morrison identity, the MMSE receiver filter can be equivalent to the following form:

$$\mathbf{w}_m = \frac{(\boldsymbol{\Sigma}_m + v_{\lceil \frac{m}{N} \rceil} \mathbf{I})^{-1} \bar{\mathbf{h}}_{m,m}}{\|(\boldsymbol{\Sigma}_m + v_{\lceil \frac{m}{N} \rceil} \mathbf{I})^{-1} \bar{\mathbf{h}}_{m,m}\|}, \forall m \in \mathbb{U}. \quad (55)$$

Bases on this, problem (53) can be equivalent to the following optimization problem by substituting the expression of beamformers \mathbf{w} given by (55) into the target SINR constraints:

$$\begin{aligned} & \max_{\mathbf{v}} \min_{\mathbf{q}} \sum_{m=1}^{\bar{N}} q_m \\ & \text{s.t. } q_m \geq \frac{\gamma_m}{\bar{\mathbf{h}}_{m,m}^H (\bar{\boldsymbol{\Sigma}}_m + v_{\lceil \frac{m}{N} \rceil} \mathbf{I})^{-1} \bar{\mathbf{h}}_{m,m}}, \forall m \in \mathbb{U}, \\ & \sum_{j=1}^K (1 - v_j) P_j = 0, q_m \geq 0, \forall m \in \mathbb{U}. \end{aligned} \quad (56)$$

It is easy to see that the inner optimization of problem (53) over variable \mathbf{q} can be realized by using the fixed-point equation. The optimal beamforming vectors on the downlink are of the form $\tilde{\mathbf{w}}_m = \sqrt{p_m} \mathbf{w}_m$, where $\sqrt{p_m}$ is the power allocated to the beamforming vector $\tilde{\mathbf{w}}_m$ [27]. From the downlink-uplink SINR relations for all users, we have

$$\left\{ \begin{aligned} p_m \|\bar{\mathbf{h}}_{m,m}^H \mathbf{w}_m\|^2 &= \bar{\gamma}_m \sum_{\substack{n=1, \\ n \neq m}}^{\bar{N}} p_n \|\bar{\mathbf{h}}_{n,m}^H \mathbf{w}_n\|^2 + \bar{\gamma}_m, \end{aligned} \right. \quad (57a)$$

$$\left\{ \begin{aligned} \sum_{m=1}^{\bar{N}} v_{\lceil \frac{m}{N} \rceil} p_m &= \sum_{m=1}^{\bar{N}} q_m. \end{aligned} \right. \quad (57b)$$

Similarly, we have the following eigensystem

$$\tilde{\mathbf{p}} = \mathbf{Q}(\mathbf{w}, \tilde{\mathbf{v}}, \bar{\boldsymbol{\gamma}}, \bar{P}_{max}) \tilde{\mathbf{p}}, \text{ with } \tilde{p}_{\bar{N}+1} = 1, \quad (58)$$

where $\bar{P}_{max} = \sum_{m=1}^{\bar{N}} q_m$, $\tilde{\mathbf{v}}$ is a \bar{N} dimension column vector and is defined as $\tilde{v}_m = v_{\lceil \frac{m}{N} \rceil}$, $\forall m$. The details of the optimal uplink power allocation algorithm for finite-size system are summarized as Algorithm 3 where $\bar{\mathbf{g}}$ denotes the subgradient of variables \mathbf{v} given by $\bar{\mathbf{g}} = [\bar{g}_1, \dots, \bar{g}_j, \dots, \bar{g}_K]$ with $\bar{g}_j = \sum_{m=(j-1)N+1}^{jN} p_m$ and $\mathbb{S}_{\mathbf{v}} = \left\{ \mathbf{v} : \mathbf{v} \geq 0, \sum_{j=1}^K (1 - v_j) P_j = 0 \right\}$ [50]. Furthermore, the convergence of Algorithm 3 can be guaranteed by the fixed function theorem [49].

Algorithm 3. JBOPA Initialization Algorithm With Instantaneous Channel Information

- 1: Initialize the auxiliary variable $\mathbf{v}^{(*)} > 0$, $\mathbf{q}^{(*)} = \mathbf{0}$ and $\mathbf{q}^{(0)} = \mathbf{0}$.
 - 2: Update the virtual uplink transmit power \mathbf{q} , i.e., $q_m^{(**)} = \frac{\gamma_m}{\bar{\mathbf{h}}_{m,m}^H (\bar{\boldsymbol{\Sigma}}_m + v_{\lceil \frac{m}{N} \rceil}^{(*)} \mathbf{I})^{-1} \bar{\mathbf{h}}_{m,m}}$ with $\mathbf{q}^{(*)}$, then obtain $\mathbf{q}^{(**)}$.
 - 3: If $\|\mathbf{q}^{(**)} - \mathbf{q}^{(*)}\| \leq \zeta$, then let $\mathbf{q}^{(*)} = \mathbf{q}^{(**)}$ and go to step 4, otherwise let $\mathbf{q}^{(*)} = \mathbf{q}^{(**)}$ and go to step 2.
 - 4: If $\|\mathbf{q}^{(*)} - \mathbf{q}^{(0)}\| \leq \zeta$, then stop, otherwise, calculate the downlink power \mathbf{p} , i.e., $\tilde{\mathbf{p}} = \mathbf{Q}(\mathbf{w}, \tilde{\mathbf{v}}^{(t)}, \bar{\boldsymbol{\gamma}}, \bar{P}_{max}) \tilde{\mathbf{p}}$ with $\mathbf{q}^{(*)}$ and $\bar{P}_{max} = \sum_{m=1}^{\bar{N}} q_m^{(*)}$, then obtain $\mathbf{p}^{(*)}$. Update \mathbf{v} with sub-gradient method i.e., $v_k^{(**)} = \mathbb{P}_{\mathbb{S}_{\mathbf{v}}} \left\{ v_k^{(*)} + \theta \bar{g}_k \right\}$ with $\mathbf{p}^{(*)}$, and obtain $\mathbf{v}^{(**)}$, let $\mathbf{q}^{(0)} = \mathbf{q}^{(*)}$, and $\mathbf{v}^{(*)} = \mathbf{v}^{(**)}$, then go to step 2.
-

C. JBOPA Initialization Algorithm With Statistical Channel Information

Similar to the JBOPA Algorithm 1, we also need to solve the initialization problem of the virtual uplink transmit power before applying the developed power allocation optimization Algorithm 2 to solve effectively problem (6). Similar procedure for obtaining the deterministic quantity $\bar{\gamma}_m$ of $\tilde{\gamma}_m$ can be applied also to $\bar{\gamma}_m$ in order to examine the asymptotically optimal solution of the PMP (53) based on statistical CSI.

Proposition 2: The instantaneous random variable $\tilde{\gamma}_m$ can be approximated by a deterministic quantity $\bar{\Gamma}_m$ such that $\bar{\Gamma}_m - \tilde{\gamma}_m \xrightarrow{a.s.} 0$ as the system dimension $M_{\lceil \frac{m}{N} \rceil} \rightarrow \infty$. Also, $\bar{\Gamma}_m$ is given by:

$$\bar{\Gamma}_m = \frac{q_m \sigma_{m,m}^2}{\omega_m^2} M_{\Lambda_m} \left(-v_{\lceil \frac{m}{N} \rceil} \right), \quad (59)$$

where $\Lambda_m = \sum_{n=1, n \neq m}^{\bar{N}} q_n \bar{\mathbf{h}}_{m,n} \bar{\mathbf{h}}_{m,n}^H$, and $M_{\Lambda_m}(z)$ denotes the Stieltjes transform of the matrix Λ_m evaluated at point z .

Now, instead of solving directly problem (53) to obtain initial solution, we resort to solve the following problem.

$$\begin{aligned} & \max_{\mathbf{v}} \min_{\mathbf{q}} \sum_{m=1}^{\bar{N}} q_m \\ & \text{s.t. } \bar{\Gamma}_m \geq \gamma_m, q_m \geq 0, \forall m \in \mathbb{U}, \sum_{j=1}^K (1 - v_j) P_j = 0. \end{aligned} \quad (60)$$

For fixed \mathbf{v} , the optimal solution of problem (60) with respect to variables \mathbf{q} can be easily obtained by using fixed equation theory. In addition, the optimization of problem (60) with respect to variables \mathbf{v} can also be solved by using the sub-gradient method. The details of the algorithm that is used to solve problem (60) in large-scale system is summarized as Algorithm 4.

Algorithm 4. JBOPA Initialization Algorithm With Statistical Channel Information

- 1: Initialize the auxiliary variable $\mathbf{v}^{(*)} > 0$, $\mathbf{q}^{(0)} = \mathbf{0}$, and $\mathbf{q}^{(*)} = \mathbf{0}$.
- 2: Update the virtual uplink transmit power μ , i.e., $\mu_m^* = \frac{\gamma_m}{\frac{\sigma_{m,m}^2}{\omega_m^2} M_{\Lambda_m}(-v \lceil \frac{m}{N} \rceil)}$ with $\mathbf{v}^{(t)}$ and $\mathbf{q}^{(*)}$, then obtain $\mathbf{q}^{(**)}$.
- 3: If $\|\mathbf{q}^{(*)} - \mathbf{q}^{(**)}\| \leq \zeta$, then let $\mathbf{q}^{(*)} = \mathbf{q}^{(**)}$ and go to step 4, otherwise let $\mathbf{q}^{(*)} = \mathbf{q}^{(**)}$ and go to step 2.
- 4: If $\|\mathbf{q}^{(*)} - \mathbf{q}^{(0)}\| \leq \zeta$, then stop, otherwise, calculate the downlink power \mathbf{p} , i.e., $\tilde{\mathbf{p}} = \mathbb{Q}(\mathbf{w}, \tilde{\mathbf{v}}^{(t)}, \bar{\Gamma}, \bar{P}_{max}) \tilde{\mathbf{p}}$ with $\mathbf{q}^{(*)}$ and $\bar{P}_{max} = \sum_{m=1}^{\bar{N}} q_m^{(*)}$, then obtain $\mathbf{p}^{(*)}$. Update \mathbf{v} with sub-gradient method and $\mathbf{p}^{(*)}$ i.e., $v_k^{(**)} = \mathbb{P}_{S_v} \left\{ v_k^{(*)} + \theta \bar{v}_k \right\}$, then obtain $\mathbf{v}^{(**)}$, let $\mathbf{q}^{(*)} = \mathbf{q}^{(0)}$, $\mathbf{q}^{(0)} = \mathbf{q}^{(**)}$, $\mathbf{v}^{(*)} = \mathbf{v}^{(**)}$ and go to step 2.

D. An Extended JBOPA Optimization Algorithm

It is interesting to find that the authors in [15] also adopted the convex approximation method to investigate the energy efficiency optimization problem for single cell multiuser downlink broadcast channel. It is easy to see that the developed algorithm in [15] can be easily extended to solve the considered energy efficiency optimization problem. To optimize the transmit power, the following lemma is applied to approximate the objective function of problem (5).

Lemma 4: Let $\phi(\tilde{\gamma}) = \frac{2\kappa\tilde{\gamma}}{2+\vartheta\tilde{\gamma}}$ be a function used to approximate function $\varphi(\tilde{\gamma}) = \alpha \log_2(1 + \tilde{\gamma})$, near the point $\hat{\gamma}$. The parameters κ and ϑ of the best monomial local approximation are given by

$$\kappa = \frac{\alpha(2 + \vartheta\hat{\gamma})^2}{4(1 + \hat{\gamma})\ln 2}, \quad \vartheta = \frac{2(1 + \hat{\gamma})\ln(1 + \hat{\gamma}) - 2\hat{\gamma}}{\hat{\gamma}^2} \quad (61)$$

and $\phi(\tilde{\gamma}) < \varphi(\tilde{\gamma})$, $\forall \tilde{\gamma} > 0$.

By using the local approximation given by Lemma 4 and the arithmetic-geometric mean inequality in the objective function of problem (5), the energy efficiency optimization can be reformulated into the following classical GP problem for fixed transmit beamformers.

$$\begin{aligned} \min_{\mathbf{p}, \tilde{\gamma}} \mu & \frac{\prod_{m=1}^{\bar{N}} (2 + \vartheta_m \tilde{\gamma}_m)^{\frac{1}{\bar{N}}}}{\prod_{m=1}^{\bar{N}} (\kappa_m \tilde{\gamma}_m)^{\frac{1}{\bar{N}}}}, \\ \text{s.t. } \tilde{\gamma}_m & \geq \tilde{\gamma}_m, \quad \vec{\gamma}_m \geq \gamma_m, \quad p_m \geq 0, \quad \forall m \in \mathbb{U}, \\ & \sum_{m=(j-1)N+1}^{jN} p_m \leq P_j, \quad \forall j \in \mathbb{K}, \end{aligned} \quad (62)$$

where $\mu = \left(\xi \sum_{j=1}^{\bar{N}} p_m + \sum_{j=1}^K (M_j P_c + P_0) \right)$. Similar to problem (23), problem (62) can also be solved by using the powerful

Algorithm 5. An Extended JBOPA Algorithm

- 1: Let $l = 0$, initial the downlink transmit power $\mathbf{p}^{(l)}$ and the beamformers $\mathbf{w}^{(l)}$ such that the SINR constraints and the power constraint are satisfied, $\vec{\mathbf{g}}^{(l)} = \mathbf{0}$, calculate the approximation point $\tilde{\mathbf{y}}^{(l)}$ with $\mathbf{p}^{(l)}$ and $\mathbf{w}^{(l)}$, and calculate the approximation constants $\vartheta_m^{(l)}$ and $\kappa_m^{(l)}$ with $\tilde{\mathbf{y}}^{(l)}$, $\forall m$.
- 2: Let $l = l + 1$, solve problem (62) with approximation point $\tilde{\mathbf{y}}^{(l)}$ for fixed $\mathbf{w}^{(l)}$, and obtain the optimal transmit power $\mathbf{p}^{(l)}$ and the optimal SINR $\tilde{\mathbf{y}}^{(*)}$.
- 3: Solve problem (63) with the target user SINRs are set to be $\tilde{\mathbf{y}}^{(*)}$, and obtain the update beamformers $\mathbf{w}_m^{(l)}$.
- 4: Compute $\vec{\mathbf{g}}^{(l)}$ with $\mathbf{p}^{(l)}$, $\mathbf{w}^{(l)}$, and (7), and check the stopping criterion, i.e., $|\vec{\mathbf{g}}^{(l)} - \vec{\mathbf{g}}^{(l-1)}| \leq \zeta$. If it is not satisfied, let $\tilde{\mathbf{y}}^{(l)} = \tilde{\mathbf{y}}^{(*)}$, update the approximation constants $\vartheta_m^{(l)}$ and $\kappa_m^{(l)}$ with $\tilde{\mathbf{y}}^{(l)}$, $\forall m$, and go to step 2, otherwise stop.

GP optimization tool packets. In order to solve the energy efficiency optimization problem with respect to the beamformers \mathbf{w} , we resort to solve a power minimization problem (PMP) that is a classical SOCP problem which preserves the SINR of each user and given by

$$\begin{aligned} \min_{\tilde{\mathbf{w}}} & \sum_{m=1}^{\bar{N}} \|\tilde{\mathbf{w}}_m\|^2 \\ \text{s.t. } & \frac{\|\bar{\mathbf{h}}_{m,m}^H \tilde{\mathbf{w}}_m\|^2}{\sum_{n=1, n \neq m}^{\bar{N}} \|\bar{\mathbf{h}}_{n,m}^H \tilde{\mathbf{w}}_n\|^2 + 1} \geq \gamma_m^{(*)}, \quad \forall m \in \mathbb{U}, \\ & p_m \geq 0, \quad \forall m \in \mathbb{U}, \\ & \sum_{m=(j-1)N+1}^{jN} \|\tilde{\mathbf{w}}_m\|^2 \leq P_j, \quad \forall j. \end{aligned} \quad (63)$$

where $\tilde{\mathbf{w}}$ denotes the collection of the beamformers $\{\tilde{\mathbf{w}}_k\}$. Assume that $\tilde{\mathbf{w}}^{(*)}$ is the optimal solution of problem (63), then let $\mathbf{w}_m^{(l)} = \frac{\tilde{\mathbf{w}}_m^{(*)}}{\|\tilde{\mathbf{w}}_m^{(*)}\|}$ and $p_m^{(*)} = \|\tilde{\mathbf{w}}_m^{(*)}\|^2$, $\forall m \in \mathbb{U}$, respectively¹⁰.

Based on these, the details of the extended energy efficiency optimization algorithm is summarized as Algorithm 2.

Note that at step 2, the objective value of problem (62) is further minimized at current approximation point, i.e., obtaining an improvement of the energy efficiency. The improvement at step 3 is that less or at most equal power is used for each beam as that at step 2, which in turn satisfies power constraints and all user rate that achieved in step 2 constraints. In other words, an increasing or equal energy efficiency can be obtained after the update of step 3 of Algorithm 5. Therefore, the sequence generated by Algorithm 5 is a monotonic sequence. Since the achievable SINR region under the transmit power constraint is bounded, the sum rate is also bounded. In other word, the achieved system energy efficiency is also bounded. Based on

¹⁰We would like to point out that this method can also be used to initialize the initial beamformers and power allocation.

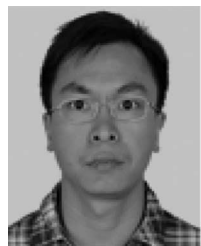
this, the convergence of Algorithm 5 is also guaranteed by the monotonic boundary theorem [56].

In addition, problem (60) has $\tilde{M} = 2N \sum_{j=1}^K M_j$ real optimization variables, $\bar{N} + 1$ second-order-cone (SOC) constraints where each of them consists of \tilde{M} real dimensions. It is known from [48] that the computational complexity of the power minimization problem in terms of number of iterations is upper bounded by $O(\sqrt{\bar{N} + 1})$, and the complexity of each iteration is within the order of $O(2\tilde{M}^3\bar{N})$. Thus, the total worst-case computational complexity of PMP is given by $O(2\sqrt{\bar{N} + 1}\tilde{M}^3\bar{N})$. Also, since both the Algorithm 1 and the extended Algorithm 5 include a GP step, we ignore this GP step in the complexity analysis.

REFERENCES

- [1] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, 1st ed. Cambridge, U.K.: Cambridge Univ. Press, 2005.
- [2] D. Gesbert, M. Kountouris, R. Heath Jr., C. Chae, T. Sälzer, "Shifting the MIMO paradigm," *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 36–46, Sep. 2007.
- [3] D. Gesbert, S. Hanly, H. Huang, S. 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.
- [4] Y. Huang, G. Zheng, M. Bengtsson, K. Wong, L. Yang, and B. Ottersten, "Distributed multi-cell beamforming design with limited intercell coordination," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 728–738, Feb. 2011.
- [5] G. Zheng, Y. Huang, and K.-K. Wong, "Network MIMO techniques," in *Heterogeneous Cellular Networks: Theory, Simulation and Deployment*, X. Chu, D. López-Pérez, Y. Yang, and F. Gunnarsson, Eds., Cambridge, U.K.: Cambridge Univ. Press, 2013, pp. 312–348.
- [6] S. Tombaz, A. Västberg, and J. Zander, "Energy- and cost-efficient ultra high-capacity wireless access," *IEEE Wireless Commun. Mag.*, vol. 18, no. 5, pp. 18–24, Oct. 2011.
- [7] Y. Chen, S. Zhang, S. Xu, and G. Li, "Fundamental trade-offs on green wireless networks," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 30–37, Jun. 2011.
- [8] D. Feng, C. Jiang, G. Lim, L. Cimini Jr., G. Feng, G. Li, "A survey of energy-efficient wireless communications," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 1, pp. 167–178, Feb. 2013.
- [9] G. Miao, N. Himayat, and G. Li, "Energy-efficient link adaptation in frequency-selective channels," *IEEE Trans. Commun.*, vol. 58, no. 2, pp. 545–554, Feb. 2010.
- [10] G. Miao, N. Himayat, G. Li, and S. Talwar, "Distributed interference-aware energy-efficient power optimization," *IEEE Trans. Wireless Commun.*, vol. 10, no. 4, pp. 1323–1333, Apr. 2011.
- [11] D. Ng, E. Lo, and R. Schober, "Energy-efficient resource allocation in multi-cell OFDMA systems with limited Backhaul capacity," *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 3618–3631, Oct. 2012.
- [12] D. Ng, E. Lo, and R. Schober, "Energy-efficient resource allocation in OFDMA systems with hybrid energy harvesting base station," *IEEE Trans. Wireless Commun.*, vol. 12, no. 7, pp. 3412–3427, Jul. 2013.
- [13] C. Xiong, G. Li, Y. Liu, Y. Chen, and S. Xu, "Energy-efficient design for downlink OFDMA with delay-sensitive traffic," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 3085–3095, Jun. 2013.
- [14] G. Miao, "Energy-efficient uplink multi-user MIMO," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 2302–2313, May 2013.
- [15] K. Manosha, S. Joshi, N. Rajatheva, and M. Latva-Aho, "Energy efficient power control and beamforming in multi-antenna enabled femtocells," in *Proc. IEEE Globecom*, Dec. 2012, pp. 3472–3477.
- [16] C. Jiang and L. Cimini Jr., "Energy-efficient transmission for MIMO interference channels," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2988–2999, Jun. 2013.
- [17] J. Xu and L. Qiu, "Energy efficiency optimization for MIMO broadcast channels," *IEEE Trans. Wireless Commun.*, vol. 12, no. 2, pp. 690–701, Feb. 2013.
- [18] S. He, Y. Huang, S. Jin, F. Yu, and L. Yang, "Max-min energy efficient beamforming for multicell multiuser joint transmission systems," *IEEE Commun. Lett.*, vol. 17, no. 10, pp. 1956–1959, Oct. 2013.
- [19] 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.
- [20] R. Jagannathan, "On some properties of programming problems in parametric form pertaining to fractional programming," *Manage. Sci.*, vol. 12, no. 7, pp. 609–615, Mar. 1966.
- [21] W. Dinkelbach, "On nonlinear fractional programming," *Bull. Aust. Math. Soc.*, vol. 13, pp. 492–498, Mar. 1967.
- [22] S. Schaible and J. Shi, "Fractional programming: The sum-of-ratios case," *Opt. Meth. Softw.*, vol. 18, pp. 219–229, 2003.
- [23] T. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.
- [24] J. Jose, A. Ashikhmin, T. 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.
- [25] F. Rusek et al., "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.
- [26] S. Wagner, R. Couillet, M. Debbah, and D. Slock, "Large system analysis of linear precoding in correlated MISO broadcast channels under limited feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4509–4537, Jul. 2012.
- [27] R. Zakhour and S. V. Hanly, "Base station cooperation on the downlink: Large system analysis," *IEEE Trans. Inf. Theory*, vol. 58, no. 4, pp. 2079–2106, Apr. 2012.
- [28] Y. Huang, C. Tan, and B. Rao, "Joint beamforming and power control in coordinated multicell: Max-min duality, effective network and large system transition," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2730–2742, Jun. 2013.
- [29] J. Silverstein and Z. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," *J. Multivariate Anal.*, vol. 54, no. 2, pp. 175–192, 1995.
- [30] Z. Bai and J. Silverstein, "No eigenvalue outside the support of the limiting spectral distribution of large dimensional sample covariance matrices," *Ann. Probab.*, vol. 26, no. 1, pp. 316–345, 1998.
- [31] A. Tulino and S. Verdú, "Random Matrix Theory and Wireless Communications (Found. Commun. Inf. Theory). Delft, The Netherlands: now Publishers, 2004, vol. 1, no. 1, pp. 38–40.
- [32] W. Hachem, P. Loubaton, and J. Najim, "Deterministic equivalents for certain functionals of large random matrices," *Ann. Appl. Probab.*, vol. 17, no. 3, pp. 875–929, Mar. 2007.
- [33] S. Lakshminaryana, J. Hoydis, M. Debbah, and M. Assaad, "Asymptotic analysis of distributed multi-cell beamforming," in *Proc. IEEE Pers. Indoor Mobile Radio Commun.*, 2010, pp. 2105–2110.
- [34] H. Dahrouj and W. Yu, "Coordinated beamforming for the multicell multi-antenna wireless system," *IEEE Trans. Wireless Commun.*, vol. 9, no. 5, pp. 1478–1759, May 2010.
- [35] R. Zakhour and S. Hanly, "Min-max power allocation in cellular networks with coordinated beamforming," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 287–302, Feb. 2013.
- [36] D. Ng, E. 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.
- [37] H. Ngo, E. Larsson, and T. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.
- [38] E. Björnson, P. Zetterberg, M. Bengtsson, and B. Ottersten, "Capacity limits and multiplexing gains of MIMO channels with transceiver impairments," *IEEE Commun. Lett.*, vol. 17, no. 1, pp. 91–94, Jan. 2013.
- [39] E. Björnson, L. Sanguinetti, J. Hoydis, and M. Debbah, "Designing multi-user MIMO for energy efficiency: When is massive MIMO the answer?," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Istanbul, Turkey, Apr. 2014, pp. 242–247.
- [40] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency estimation and capacity limits," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 7112–7139, Nov. 2014.
- [41] S. Mohammed, "Impact of transceiver power consumption on the energy efficiency of zero-forcing detector in massive MIMO systems," *IEEE Trans. Commun.*, vol. 62, no. 11, pp. 3874–3890, Nov. 2014.
- [42] L. Tran, M. Juntti, M. Bengtsson, and B. Ottersten, "Weighted sum rate maximization for MIMO broadcast channels using dirty paper coding and zero-forcing methods," *IEEE Trans. Commun.*, vol. 61, no. 6, pp. 2362–2373, Jun. 2013.

- [43] 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.
- [44] S. Bandyopadhyay, S. Saha, U. Maulik, and K. Deb, "A simulated annealing-based multiobjective optimization algorithm: AMOSA," *IEEE Trans. Evol. Comput.*, vol. 12, no. 3, pp. 269–283, Jun. 2008.
- [45] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [46] L. Zhang, R. Zhang, Y. Liang, Y. Xin, and H. Poor, "On Gaussian MIMO BC-MAC duality with multiple transmit covariance constraints," *IEEE Trans. Inf. Theory*, vol. 58, no. 4, pp. 2064–2078, Apr. 2012.
- [47] G. Dartmann, X. Gong, W. Afzal, and G. Ascheid, "On the duality of the max-min beamforming problem with per-antenna and per-antenna-array power constraints," *IEEE Trans. Veh. Technol.*, vol. 62, no. 2, pp. 606–619, Feb. 2013.
- [48] J. Borwein and A. Lewis, *Convex Analysis and Nonlinear Optimization: Theory and Examples*. Berlin, Germany: Springer-Verlag, 2006.
- [49] A. Wiesel, Y. Eldar, and S. Shamai, "Linear precoding via conic optimization for fixed MIMO receivers," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 161–176, Jan. 2006.
- [50] W. Yu and T. Lan, "Transmitter optimization for the multi-antenna downlink with per-antenna power constraints," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2646–2660, Jun. 2007.
- [51] M. Schubert and H. Boche, "Solution of the multiuser downlink beamforming problem with individual SINR constraints," *IEEE Trans. Veh. Technol.*, vol. 53, no. 1, pp. 18–28, Jan. 2004.
- [52] P. Weeraddana, M. Codreanu, M. Latva-Aho, and A. Ephremides, "Resource allocation for cross-layer utility maximization in wireless networks," *IEEE Trans. Veh. Technol.*, vol. 60, no. 6, pp. 2790–2809, Jul. 2011.
- [53] S. Boyd, S. Kim, L. Vandenberghe, and A. Hassibi, "A tutorial on geometric programming," *Optim. Eng.*, vol. 8, no. 1, pp. 67–127, Mar. 2007.
- [54] MOSEK ApS. (2011). *The MOSEK Optimization Toolbox for MATLAB Manual*, Version 6.0 (Revision 103) [Online]. Available: <http://www.mosek.com>
- [55] X. Zhang and J. Lee, "Low complexity MIMO scheduling with channel decomposition using capacity upperbound," *IEEE Trans. Commun.*, vol. 56, no. 6, pp. 871–876, Jun. 2008.
- [56] J. Bibby, "Axiomatisations of the average and a further generalisation of monotonic sequences," *Glasgow Math. J.*, vol. 15, pp. 63–65, 1974.
- [57] "Aspects of WINNER+ spectrum preferences," *WINNER+Deliverable D3.2*, May 2009.



Shiwen He (M'14) received the M.S. degree from Chengdu University of Technology, Chengdu, China, and the Ph.D. degree in information and communication engineering from Southeast University, Nanjing, China, in 2009 and 2013, respectively. From 2013 to 2015, he was a Postdoctoral Researcher with the State Key Laboratory of Millimeter Waves, Department of Radio Engineering, Southeast University. Since October 2015, he has been with the School of Information Science and Engineering, Southeast University. His research interests include multiuser

MIMO communication, cooperative communications, energy efficient communications, millimeter wave communication, and optimization theory.



Yongming Huang (M'10) received the B.S. and M.S. degrees from Nanjing University, Nanjing, China, and the Ph.D. degree in electrical engineering from Southeast University, Nanjing, China, in 2000, 2003 and 2007, respectively. Since March 2007, he has been a Faculty of the School of Information Science and Engineering, Southeast University. From 2008 to 2009, he visited the Signal Processing Laboratory, Electrical Engineering, Royal Institute of Technology (KTH), Stockholm, Sweden. His research interests include space time wireless communications, cooperative

wireless communications, energy efficient wireless communications and optimization theory. He serves as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, *EURASIP Journal on Advances in Signal Processing*, and *EURASIP Journal on Wireless Communications and Networking*.



Luxi Yang (M'96) received the M.S. and Ph.D. degree in electrical engineering from the Southeast University, Nanjing, China, in 1990 and 1993, respectively. Since 1993, he has been with the Department of Radio Engineering, Southeast University, where he is currently a Professor of information systems and communications, and the Director of Digital Signal Processing Division. He is the author or coauthor of 2 published books and more than 100 journal papers, and holds 10 patents. His research interests include signal processing for wireless communica-

tions, MIMO communications, cooperative relaying systems, and statistical signal processing. He received the first- and second-class prizes of the Science and Technology Progress Awards of the State Education Ministry of China, in 1998 and 2002, respectively. He is currently a member of the Signal Processing Committee of Chinese Institute of Electronics.



Björn Ottersten (S'87–M'89–SM'99–F'04) was born in Stockholm, Sweden, 1961. He received the M.S. degree in electrical engineering and applied physics from Linköping University, Linköping, Sweden, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, USA, in 1986 and 1989, respectively. He has held research positions with the Department of Electrical Engineering, Linköping University; the Information Systems Laboratory, Stanford University; the Katholieke Universiteit Leuven, Leuven; and the

University of Luxembourg, Luxembourg. From 1996 to 1997, he was the Director of Research at ArrayComm Inc., a start-up in San Jose, CA, USA, based on Ottersten's patented technology. In 1991, he was appointed as a Professor of signal processing with the Royal Institute of Technology (KTH), Stockholm, Sweden. From 1992 to 2004, he was the Head of the Department for Signals, Sensors, and Systems with KTH, and from 2004 to 2008, he was the Dean of the School of Electrical Engineering with KTH. Currently, he is the Director for the Interdisciplinary Centre for Security, Reliability, and Trust with the University of Luxembourg. He is a Board Member of the Swedish Research Council and as Digital Champion of Luxembourg, and acts as an Adviser to the European Commission. His research interests include security and trust, reliable wireless communications, and statistical signal processing. He has served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and on the Editorial Board of the *IEEE Signal Processing Magazine*. He is currently an Editor-in-Chief of *EURASIP Signal Processing Journal* and a member of the Editorial Boards of *EURASIP Journal of Applied Signal Processing* and *Foundations and Trends in Signal Processing*. He is a Fellow of the EURASIP and a member of the IEEE Signal Processing Society Board of Governors. He has coauthored journal papers that received the IEEE Signal Processing Society Best Paper Award in 1993, 2001, 2006, and 2013, and the 3 IEEE conference papers receiving Best Paper Awards. He was the recipient of the IEEE Signal Processing Society Technical Achievement Award in 2011. He is the first recipient of the European Research Council advanced research grant.



Wei Hong (M'92–SM'07–F'12) received the B.S. degree from the University of Information Engineering, Zhengzhou, China, and the M.S. and Ph.D. degrees from Southeast University, Nanjing, China, all in radio engineering, in 1982, 1985, and 1988, respectively. Since 1988, he has been with the State Key Laboratory of Millimeter Waves and has served for the director of the laboratory since 2003. He is currently a Professor and the Dean of the School of Information Science and Engineering, Southeast University. In 1993, 1995,

1996, 1997, and 1998, he was a short-term Visiting Scholar at the University of California at Berkeley, Berkeley, CA, USA, and at the University of California, Santa Cruz, CA, USA, respectively. He has been engaged in numerical methods for electromagnetic problems, millimeter wave theory and technology, antennas, RF technology for wireless communications etc. He has authored and coauthored over 300 technical publications and 2 books. He was the recipient of the first-class Science and Technology Progress Prizes (thrice) issued by the Ministry of Education of China and Jiangsu Province Government, etc. He was also the recipient of the Foundations for China Distinguished Young Investigators and for "Innovation Group" issued by NSF of China. He is a Fellow of CIE, MTT-S AdCom Member, a Vice-President of Microwave Society and Antenna Society of CIE, the Chairperson of the IEEE MTT-S/AP-S/EMC-S Joint Nanjing Chapter, and has served as an Associate Editor of the IEEE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES from 2007 to 2010, and an Editor Board member for *International Journal of Antennas and Propagation*, *China Communications*, and *Chinese Science Bulletin*, among others.