Energy-Efficient Beam Coordination Strategies with Rate Dependent Processing Power

Oskari Tervo, *Student Member, IEEE*, Antti Tölli, *Senior Member, IEEE*, Markku Juntti, *Senior Member, IEEE*, and Le-Nam Tran, *Member, IEEE*

Abstract—This paper proposes energy-efficient coordinated beamforming strategies for multi-cell multi-user multiple-input single-output system. We consider a practical power consumption model, where part of the consumed power depends on the base station or user specific data rates due to coding, decoding and backhaul. This is different from the existing approaches where the base station power consumption has been assumed to be a convex or linear function of the transmit powers. Two optimization criteria are considered, namely network energy efficiency maximization and weighted sum energy efficiency maximization. We develop successive convex approximation based algorithms to tackle these difficult nonconvex problems. We further propose decentralized implementations for the considered problems, in which base stations perform parallel and distributed computation based on local channel state information and limited backhaul information exchange. The decentralized approaches admit closed-form solutions and can be implemented without invoking a generic external convex solver. We also show an example of the pilot contamination effect on the energy efficiency using a heuristic pilot allocation strategy. The numerical results are provided to demonstrate that the rate dependent power consumption has a large impact on the system energy efficiency, and, thus, has to be taken into account when devising energy-efficient transmission strategies. The significant gains of the proposed algorithms over the conventional low-complexity beamforming algorithms are also illustrated.

Index Terms—Coordinated beamforming, centralized algorithms, decentralized algorithms, energy efficiency, successive convex approximation, fractional programming, pilot contamination, circuit power, processing power.

I. INTRODUCTION

Multi-antenna technology has been a core underlying component in modern wireless communication systems and will certainly remain its vital role in the development of the future 5G networks. Although the multi-antenna techniques have been shown to provide huge spectral efficiency gains, they cause a serious concern over the increased power consumption due to the number of associated radio frequency (RF) elements. In addition to the power directly used to transmit data, a significant amount of indirect power is consumed for other

related operations, e.g., for running the base stations (BSs), baseband processing, RF processing, coding, decoding, and backhaul operations. In fact, in densely deployed networks, the data transmit power may be only a small part of the total power consumption according to 5G visions [1]. As a result, energy efficiency (EE) has become an important design criterion for future networks [2]–[7]. On the contrary to the conventional design criteria of either maximizing the sum rate or minimizing the required transmit power, energy efficiency optimization is to maximize the ratio between sum rate and total power consumption.

The main challenge in multiuser wireless communication systems design is due to multiuser interference caused by the use the same transmission resources. To this end, several interference coordination techniques have been proposed in the last decades. The idea of these methods is to mitigate the inter-cell interference by allowing cooperation between nearby cells. Among those, a powerful method adopted in current LTE-A systems is called *coordinated beamforming*, where base stations can jointly design their beamforming vectors without sharing their data [8]. Another widely studied technique is joint transmission, where the data can be coherently transmitted from multiple base stations [9]. However, since joint transmission requires a tight synchronization between the base stations, the focus of this paper is on the coordinated beamforming. It has been widely studied, e.g., for sum rate maximization [10]–[12], and transmit power minimization [13]–[15]. The energy-efficient coordinated beamforming strategy is highly dependent on the processing power resulting from the circuits and coordination operations of the base stations.

Related Work: The energy efficiency maximization (EEmax) problems belong to the class of fractional programs, which have been widely studied for both single-cell (SC) [5], [6], [16] and multi-cell (MC) system models [7], [9], [17], [18]. The problem of maximizing the minimum energy efficiency among base stations in a multi-cell multiuser multiple-input single-output (MISO) system was studied in [7]. Energy-efficient joint transmission for multi-cell OFDMA systems with limited backhaul capacity and single-antenna base stations was considered in [9], where the power and sub-carrier allocations were jointly optimized by assuming zero-forcing beamformers over all the base stations. Coordinated beamforming for network EEmax in multi-cell multi-antenna systems was studied in [17], [18] where the latter incorporated the data rate constraints of users to the optimization problem. The weighted sum energy efficiency (WsumEE) was suggested as a performance measure in [16], [19], [20], to satisfy the

heterogeneous energy efficiency requirements of different cells or users. However, all these works only considered the circuit power as a constant and failed to recognize the fact that this sort of power consumption heavily depends on the data rate. More specifically, the circuit power is an increasing function of the transmission rate, since a higher data rate requires a larger codebook which incurs higher power for encoding and decoding on baseband circuit boards [21]–[32].

The rate dependent power has been assumed to be either linear or non-linear convex increasing function of the data rate [21]–[32]. The linear case with uniform user rates in a single-cell system was investigated in [23], where zero-forcing beamforming with massive MIMO setup was shown to achieve maximal energy efficiency. However, in a multi-cell network where the inter-cell interference experienced by each user becomes significant, zero-forcing method is highly suboptimal because the degrees of freedom are used up for nulling both intra- and inter-cell interference. A general convex power consumption model in point-to-point MIMO orthogonal frequency-division multiplexing systems was considered in [24], [25], and in single-antenna energy-harvesting systems in [26], [28], [30]. The fact that the rate dependent power consumption can increase more than linearly with the data rate is shown, e.g., in [31], where the power consumption of the decoding process is investigated by using convolutional codes. Specifically, these codes can be decoded using a trellis representation of the encoder's state space, and the authors in [31] show that the decoding complexity for each time symbol increases exponentially with the information rate. This complexity result is caused by the fact that the state space and the number of possible state transitions per channel access in the decoder-trellis expands exponentially with the product of constraint length and the information rate.

Contributions: We study energy-efficient coordinated beamforming in multi-cell multi-user MISO systems with a general non-linearly increasing convex rate dependent power consumption model as in [24], [25]. This is different from the related research, which adopts either a simple (i.e., rate-independent) power consumption model [17], [19], [20] or a simplified beamforming technique (i.e., single-cell zero-forcing [23] or a point-to-point system [24], [25]). Because the power consumption model depends on a specific implementation (e.g., the decoder), we use a general model to emphasize that the proposed algorithms can be applied to any power consumption model which is a convex increasing function of the data rate. Two different optimization criteria are considered. The first one is the network energy efficiency maximization (NetEEmax) which yields the maximum achievable energy efficiency of the network. The second one is the weighted sum energy efficiency maximization (WsumEEmax), which maximizes the sum of energy efficiencies of the cells. The latter is particularly relevant in heterogeneous networks, since it can balance or adjust the energy efficiencies and data rates between the cells and can be implemented in a decentralized manner [20]. The framework for the proposed solutions is based on successive convex approximation (SCA) principle [33], which has been successfully applied in various wireless communications problems [6], [34], [35]. The SCA

is a local optimization method, where the main idea is to approximate the nonconvex part of the problem by proper convex bounds. In this way, the SCA method results in a sequence of convex subproblems that guarantee the feasibility of the iterates and monotonicity of the objective function. For the NetEEmax problem, we propose equivalent transformations to arrive at an iterative algorithm where a concave-convex fractional program is solved in each iteration using the Charnes-Cooper transformation [36]. The WsumEEmax problem is reformulated to derive an iterative method where a convex program is solved in each iteration. We also propose decentralized implementations in which base stations perform parallel and distributed computation based on local channel state information and limited backhaul information exchange. More specifically, we first reformulate the original problems by using the relation between signal-to-interference-plus-noise ratio (SINR) and mean squared error (MSE) when minimum MSE (MMSE) receiver is used. Then, we derive the optimality conditions for the approximated problem and propose an algorithm which combines the SCA and alternating optimization methods, and admits closed-form solutions.

To illustrate the pilot contamination effect on the energy efficiency, we propose a simple heuristic energy-efficient pilot allocation strategy. The idea of the proposed low-complexity method is to use (known) path gain information to calculate group-specific energy efficiency metrics and greedily allocate the pilot resources to the groups.

The numerical results illustrate that the proposed algorithms can provide significant energy efficiency gains (up to 60 % in the considered setting) over the methods where the rate dependent power is not taken into account, showing the importance of including it in the optimization framework. We numerically compare the proposed algorithms with various conventional transmission schemes, and show that our proposed algorithms outperform all the existing coordinated beamforming designs in terms of energy efficiency.

Parts of this paper have been published in our previous conference publication [37]. The following additional contributions can be found in the present paper. We propose decentralized closed-form designs for the problems and also provide additional implementation aspects for the centralized solutions. We have also extended the system model to take into account the pilot overhead and pilot contamination, and consider more specific power consumption model. Furthermore, we propose a heuristic pilot allocation algorithm to address the problem of pilot contamination. We also provide an alternative iterative second-order cone program (SOCP) approximation algorithm to solve the WsumEEmax problem. Finally, we provide more extensive simulations to illustrate the performance of the methods.

Organization and Notation: The rest of the paper is organized as follows. Section II presents the system model, power consumption model and the considered optimization problems. The centralized algorithms are provided in Section III, followed by the decentralized methods in Section IV. The computational complexity is discussed in Section V. Pilot allocation strategy is proposed in Section VI while numerical evaluation and conclusions are presented in Section VII and

VIII, respectively.

The following notations are used in this paper. $|x|$ denotes the cardinality of x if x is a set, and absolute value of x, otherwise. $||\mathbf{x}||_2$ is a ℓ_2 or Euclidean norm of $\mathbf{x}$ and boldcase letters without any superscripts or subscripts denote the set of variables. Otherwise, boldcase letters are vectors. $\mathbf{x}^T, \mathbf{x}^H, \mathrm{Re}(\mathbf{x})$ and $\mathrm{Im}(\mathbf{x})$ mean transpose, Hermitian transpose, real part and imaginary part of $\mathbf{x}$, respectively. $\mathrm{tr}(\mathbf{X})$ means trace of $\mathbf{X}$.

II. System Model and Problem Formulation

A. System Model

We consider the downlink (DL) of a multi-cell multiple-input single-output (MISO) system with B cells. Each base station $b \in \mathcal{B} = \{1, \dots, B\}$ equipped with N_b antennas transmits data to a group of K_b single-antenna users in its cell, represented by the set $\mathcal{K}_b$. Each user $k \in \mathcal{K} \triangleq \cup_{b \in \mathcal{B}} \mathcal{K}_b$ in the network is served only by a single BS which is denoted by $b_k \in \mathcal{B}$, i.e., $\mathcal{K}_b \cap \mathcal{K}_{b'} = \emptyset \; \forall b \neq b'$.

In the downlink, the data symbol s_k intended for user k is multiplied with the beamforming vector $\mathbf{w}_k \in \mathbb{C}^{N_b \times 1}$ before being transmitted. Accordingly, the received signal at user k is given by

$$y_k = \mathbf{h}_{b_k,k}^H \mathbf{w}_k s_k + \sum_{j \in \mathcal{K} \setminus \{k\}} \mathbf{h}_{b_j,k}^H \mathbf{w}_j s_j + n_k \quad (1)$$

where $\mathbf{h}_{b,k} \in \mathbb{C}^{N_b \times 1}$ is the channel vector from BS b to user k, and n_k is the background noise with distribution $\mathcal{CN}(0, \sigma^2)$. The data streams are assumed to be independent and have zero mean and unit power. As a result, we can write the signal-to-interference-plus-noise ratio of user k as

$$\Gamma_k(\mathbf{w}) \triangleq \frac{|\mathbf{h}_{b_k,k}^H \mathbf{w}_k|^2}{N_0 + \sum_{j \in \mathcal{K} \setminus \{k\}} |\mathbf{h}_{b_j,k}^H \mathbf{w}_j|^2} \quad (2)$$

where $\mathbf{w} \triangleq \{\mathbf{w}_k\}_{k \in \mathcal{K}}$ and $N_0 = \sigma^2 W$ by assuming that the system operates over bandwidth W Hz.

We consider a block transmission, where the user channels stay constant within time-frequency coherence blocks of $U = W_C T_C$ channel uses/symbols. W_C (in Hz) denotes the coherence bandwidth and T_C (in second) is the coherence time. For the notational simplicity, we assume $W_C = W$ in this paper. We assume time-division duplex (TDD) protocol where all the cells are operating in the downlink mode. The protocol is matched to the coherence block so that each block consists of downlink demodulation pilots, uplink (UL) pilots, and downlink data transmission. The BSs can estimate the downlink channels from the uplink pilots by exploiting the reciprocity. On the other hand, the users can estimate the effective channels (i.e., the combination of beamformer and channel) from the downlink demodulation pilots to decode the data. Let $\tau^{\mathrm{ul}} \geq \max_b K_b$ and $\tau^{\mathrm{dl}} \geq \max_b K_b$ be the total number of pilot resources in the network in uplink and downlink, respectively. This means that the total number of pilots in the coherence block is $\tau^{\mathrm{ul}} + \tau^{\mathrm{dl}}$. It is assumed that all the users are active across the transmission bandwidth, i.e., there is no frequency-domain scheduling. Thus, the rate expression of user k for a given (known) channel realization is given by

$$R_k(\mathbf{w}) = (1 - \tfrac{\tau^{\mathrm{ul}} + \tau^{\mathrm{dl}}}{U}) W \log(1 + \Gamma_k(\mathbf{w})) \quad (3)$$

where $(1 - \tfrac{\tau^{\mathrm{ul}} + \tau^{\mathrm{dl}}}{U})$ accounts for the pilot overhead in UL and DL.[1] As is well-known, the capacity expressions in information theory are based on the length of the codewords approaching infinity. Although this can never be exactly realized in practice, the state of the art powerful codes such as turbo, low density parity check (LDPC) and polar codes [38], [39] can approximately achieve the bounds with practical codeword block sizes (of less than 10000 bits) at low to moderate SNR levels with the accuracy sufficient for practical purposes.[2] In that sense, (3) is a practically relevant upper bound for the achievable rate. In other words, we focus on block transmission wherein the block is large enough in terms of data bits such that the information theoretic rate expression is a relevant upper bound for the achievable rate within a block of static channel state over coherence time; this rate value is deterministic given the known channel realization. Simultaneously, the block needs to be short enough in physical time units such that the channel remains constant over the block. Note that the optimization is performed for a fixed channel. When the final performance is evaluated via computer simulations, the rate bound in (3) can be averaged out over the fading channel states.

Pilot contamination: The pilot reuse results in the effect of pilot contamination. For simplicity, we focus on uplink pilot contamination and assume that the users have perfect effective channel information, i.e, we set $\tau^{\mathrm{dl}} = |\mathcal{K}|$ throughout the paper. However, in practice the pilot reuse is meaningful for the downlink demodulation pilots as well. If $\tau^{\mathrm{ul}} = |\mathcal{K}|$, the BSs have perfect channel information towards all the users, since it allows to allocate orthogonal pilot resources to each user. On the other hand, if $\tau^{\mathrm{ul}} < |\mathcal{K}|$, then some of the pilot resources need to be reused, which leads to the effect of pilot contamination. Assuming that the channels are perfectly estimated by each BS from the received uplink pilots, which are contaminated by the pilot signals of the users using the same pilot resources, the observed channel at BS b becomes

$$\tilde{\mathbf{h}}_{b,k} = \hat{\mathbf{h}}_{b,k} + \sum_{j \in \mathcal{K}^i} \hat{\mathbf{h}}_{b,j} \quad (4)$$

where $\hat{\mathbf{h}}_{b,k} \in \mathbb{C}^{N_b \times 1}$ is the (perfect) channel vector from BS b to user k, and $\mathcal{K}^i$ is the set of users using pilot resource i. The above expression uses a standard assumption of high-SNR pilots so that the noise term vanishes. As a result, the BS performs the beamformer optimization based on $\tilde{\mathbf{h}}_{b,k}$, instead of $\hat{\mathbf{h}}_{b,k}$.

B. Power Consumption Model

We combine the power consumption models from [23]–[25] and extend the general rate dependent power consumption model to account for the multi-cell multi-user transmission. As a result, the total power consumption of BS b writes as

$$P_{\mathrm{tot},b} = \frac{1}{\eta} \sum_{k \in \mathcal{K}_b} ||\mathbf{w}_k||_2^2 + P_{\mathrm{CP},b} + P_{\mathrm{RD}} \delta(r_b) \quad (5)$$

---

[1] The methods are easily extended to a frequency-selective case where the rate would be defined as $R_k(\mathbf{w}) = (1 - \tfrac{\tau^{\mathrm{ul}} + \tau^{\mathrm{dl}}}{U}) \sum_{m \in \mathcal{M}} W_C \log(1 + \Gamma_{k,m}(\mathbf{w}_m))$, where $\mathcal{M}$ is the set of coherence bands and the SINR expressions are per coherence band.

[2] For example, with 100 Mbit/s high rate mobile broadband, this would require the channel to remain static for a block size of 100 μs in time domain, which is not a hard assumption.

where the first term is the data transmit power in the downlink, $\eta \in [0, 1]$ is the power amplifier efficiency at the BS, $P_{\text{CP},b}$ is the total rate independent circuit power consumption of cell b, $P_{\text{RD}} \geq 0$ is a constant accounting for the coding, decoding and backhaul power consumption, and $\delta(r_b)$ is a differentiable, strictly increasing and convex function of the total sum rate r_b of BS b, satisfying $\delta(0) = 0$. The rate dependent power consumption of BS b could also be modeled as $\sum_{k \in \mathcal{K}_b} P_{\text{RD}} \delta(r_k)$, where $\delta(r_k)$ is a function of individual user rate r_k. In Appendix A, we show how to modify the algorithms proposed in the subsequent sections to deal with this alternative model. $P_{\text{CP},b}$ in (5) is decomposed as

$$P_{\text{CP},b} = P_{\text{FIX}} + P_{\text{TC},b} + P_{\text{CE}} + P_{\text{LP},b} \tag{6}$$

where P_{FIX} is a fixed power consumption required for site-cooling, control signaling, and the load-independent power of backhaul infrastructure and baseband processors, $P_{\text{TC},b}$ is the power consumption of the transceiver chains, P_{CE} is the power consumed in channel estimation for DL and UL, and $P_{\text{LP},b}$ is the power used for linear processing at the BS side. $P_{\text{TC},b}$ can be further decomposed as

$$P_{\text{TC},b} = N_b P_{\text{BS}} + P_{\text{SYN}} + K_b P_{\text{UE}} \tag{7}$$

where P_{BS} is the power per RF chain at each antenna, P_{SYN} is the power consumed by local oscillator and P_{UE} is the fixed circuit power of each user. To model the power consumed for computation of the beamformers and linear processing, we exploit the model used in [23]. In this regard, $P_{\text{LP},b}$ can be expressed as

$$P_{\text{LP},b} = W(1 - \frac{(\tau^{\text{ul}} + \tau^{\text{dl}})}{U}) \frac{2 N_b |\mathcal{K}_b|}{L_{\text{BS}}} + P_{\text{LP,c}}, \tag{8}$$

where L_{BS} [flops/W or flop/J] is the computational efficiency of a BS. The first term in (8) is the power consumed for the linear multiplication of the beamformers and the data symbols over the whole transmission bandwidth, while the second term $P_{\text{LP,c}}$ is the power consumed for the computation of the beamformers. The power consumption in relation to beamformer optimization depends on the specific algorithm. Since all the proposed methods are iterative, we can write

$$P_{\text{LP,c}} = Q P_{\text{LP,iter}}, \tag{9}$$

where Q is the required number of iterations to calculate the beamformers. The value of $P_{\text{LP,iter}}$ depends on the complexity of each iteration and is discussed in Section V.[3]

C. Problem Formulation

Let us define $\tilde{R}_b(\mathbf{w}) \triangleq \alpha \sum_{k \in \mathcal{K}_b} \log(1 + \Gamma_k(\mathbf{w}))$ to be a function denoting the total sum rate of BS b, where $\alpha \triangleq (1 - \frac{\tau^{\text{ul}} + \tau^{\text{dl}}}{U})W$. The first problem considered in this paper is called network energy efficiency maximization stated as

$$\max_{\mathbf{w}} \quad \frac{\sum_{b \in \mathcal{B}} \tilde{R}_b(\mathbf{w})}{g(\mathbf{w}) + P_{\text{RD}} \sum_{b \in \mathcal{B}} \delta(\tilde{R}_b(\mathbf{w}))} \tag{10a}$$

$$\text{s.t.} \quad \sum_{k \in \mathcal{K}_b} ||\mathbf{w}_k||_2^2 \leq P_b, \ \forall b \in \mathcal{B} \tag{10b}$$

where $g(\mathbf{w}) \triangleq \frac{1}{\eta} \sum_{k \in \mathcal{K}} ||\mathbf{w}_k||_2^2 + \sum_{b \in \mathcal{B}} P_{\text{CP},b}$ includes the power consumption which does not depend on the rate function. Conventionally, the denominator of the objective function has been either linear or convex function of the power values [23]–[25]. However, this assumption no longer holds for the problem in (10).

The second problem of interest is the one of weighted sum energy efficiency maximization with BS-specific power constraints written as

$$\max_{\mathbf{w}} \quad \sum_{b \in \mathcal{B}} \omega_b \frac{\tilde{R}_b(\mathbf{w})}{g_b(\tilde{\mathbf{w}}_b) + P_{\text{RD}} \delta(\tilde{R}_b(\mathbf{w}))} \tag{11a}$$

$$\text{s.t.} \quad \sum_{k \in \mathcal{K}_b} ||\mathbf{w}_k||_2^2 \leq P_b, \ \forall b \in \mathcal{B} \tag{11b}$$

where $\tilde{\mathbf{w}}_b \triangleq \{\mathbf{w}_k\}_{k \in \mathcal{K}_b}$, $g_b(\tilde{\mathbf{w}}_b) \triangleq \frac{1}{\eta} \sum_{k \in \mathcal{K}_b} ||\mathbf{w}_k||_2^2 + P_{\text{CP},b}$ refers to the power consumption which is independent of the rate function $\tilde{R}_b(\mathbf{w})$, and ω_b is the energy efficiency priority weighting factor for BS b. Despite the apparent similarity, the WsumEEmax problem is somewhat more difficult to tackle compared to (10), simply because the objective (11a) is a sum of fractional functions, which is not quasiconcave even if the numerators and denominators are linear. The energy efficiency metrics (10a) and (11a) above can be seen as achievable instantaneous EE values optimized per channel realization.

III. PROPOSED CENTRALIZED SOLUTIONS

A. Network Energy Efficiency Maximization

We remark that the problem in (10) is not a concave fractional program for which efficient methods are known [40], [36]. The obvious reason is that both the numerator and the denominator in (10a) are nonconvex. To find a more tractable reformulation, we introduce the following equivalent transformation of (10)

$$\max_{\mathbf{w}, \mathbf{r}} \quad \frac{\sum_{b \in \mathcal{B}} r_b}{g(\mathbf{w}) + P_{\text{RD}} \sum_{b \in \mathcal{B}} \delta(r_b)} \tag{12a}$$

$$\text{s.t.} \quad r_b \leq \alpha \sum_{k \in \mathcal{K}_b} \log(1 + \Gamma_k(\mathbf{w})), \forall b \in \mathcal{B} \tag{12b}$$

$$\sum_{k \in \mathcal{K}_b} ||\mathbf{w}_k||_2^2 \leq P_b, \ \forall b \in \mathcal{B} \tag{12c}$$

where (10) and (12) are equivalent because the constraints in (12b) are active at optimality, and $\mathbf{r} \triangleq \{r_b\}_{b \in \mathcal{B}}$ are new variables representing the sum rate of each base station b.[4] At this point, we note that the objective function has become a linear-convex fractional function and the difficulty in solving (12) is due to the constraint (12b). To this end, we introduce new variables $\gamma \triangleq \{\gamma_k\}_{k \in \mathcal{K}}$ to represent the SINR of each user k, and equivalently formulate (12) as

$$\max_{\mathbf{w}, \mathbf{r}, \gamma} \quad \frac{\sum_{b \in \mathcal{B}} r_b}{g(\mathbf{w}) + P_{\text{RD}} \sum_{b \in \mathcal{B}} \delta(r_b)} \tag{13a}$$

$$\text{s.t.} \quad r_b \leq \alpha \sum_{k \in \mathcal{K}_b} \log(1 + \gamma_k), \forall b \in \mathcal{B} \tag{13b}$$

$$\gamma_k \leq \frac{|\mathbf{h}_{b_k,k}^H \mathbf{w}_k|^2}{N_0 + \sum_{j \in \mathcal{K} \setminus \{k\}} |\mathbf{h}_{b_j,k}^H \mathbf{w}_j|^2}, \forall k \in \mathcal{K} \tag{13c}$$

$$\sum_{k \in \mathcal{K}_b} ||\mathbf{w}_k||_2^2 \leq P_b, \ \forall b \in \mathcal{B} \tag{13d}$$

---

[3]The fixed power values in general can be different for different BSs especially in heterogeneous networks, but for simplicity they are assumed to be equal for all the cells.

[4]The proposed algorithms can be straightforwardly extended to include data rate constraints, i.e., $r_b \geq R_b^{\text{target}}$ or $r_k \geq R_k^{\text{target}}$.

where SINR constraints in (13c) are still nonconvex. By introducing a new variables $\boldsymbol{\beta} \triangleq \{\beta_k\}_{k\in\mathcal{K}}$ for total interference-plus-noise of user k [34], [12], we can further rewrite (13) as

$$\max_{\mathbf{w},\mathbf{r},\boldsymbol{\gamma},\boldsymbol{\beta}} \quad \frac{\sum_{b\in\mathcal{B}} r_b}{g(\mathbf{w}) + P_{\text{RD}} \sum_{b\in\mathcal{B}} \delta(r_b)} \tag{14a}$$

$$\text{s. t.} \quad \gamma_k \leq |\mathbf{h}_{b_k,k}^H \mathbf{w}_k|^2 / \beta_k, \forall k \in \mathcal{K} \tag{14b}$$

$$r_b \leq \alpha \sum_{k\in\mathcal{K}_b} \log(1 + \gamma_k), \forall b \in \mathcal{B} \tag{14c}$$

$$\beta_k \geq N_0 + \sum_{j\in\mathcal{K}\setminus\{k\}} |\mathbf{h}_{b_j,k}^H \mathbf{w}_j|^2, \forall k \in \mathcal{K} \tag{14d}$$

$$\sum_{k\in\mathcal{K}_b} ||\mathbf{w}_k||_2^2 \leq P_b, \ \forall b \in \mathcal{B} \tag{14e}$$

which lends itself to the application of successive convex approximation framework [33], [41]. Specifically the right hand side of (14b) is called a quadratic-over-linear function which is jointly convex with respect to β_k and $\mathbf{w}_k$. Thus, we can use the first-order lower approximation for the right side of (14b) as [20]

$$|\mathbf{h}_{b_k,k}^H \mathbf{w}_k|^2 / \beta_k \geq 2\text{Re}((\mathbf{w}_k^{(n)})^H \mathbf{h}_{b_k,k} \mathbf{h}_{b_k,k}^H \mathbf{w}_k)/\beta_k^{(n)}$$
$$- (|\mathbf{h}_{b_k,k}^H \mathbf{w}_k^{(n)}|/\beta_k^{(n)})^2 \beta_k \triangleq \Psi_k^{(n)}(\mathbf{w}_k, \beta_k). \tag{15}$$

According to the SCA principle we will replace the right side of (14b) by a convex lower bound. From (15), the problem at iteration n of the proposed SCA-based algorithm can be expressed as

$$\max_{\mathbf{w},\boldsymbol{\gamma},\boldsymbol{\beta},\mathbf{r}} \quad \frac{\sum_{b\in\mathcal{B}} r_b}{g(\mathbf{w}) + P_{\text{RD}} \sum_{b\in\mathcal{B}} \delta(r_b)} \tag{16a}$$

$$\text{s. t.} \quad \gamma_k \leq \Psi_k^{(n)}(\mathbf{w}_k, \beta_k), \forall k \in \mathcal{K} \tag{16b}$$

$$r_b \leq \alpha \sum_{k\in\mathcal{K}_b} \log(1 + \gamma_k), \forall b \in \mathcal{B} \tag{16c}$$

$$\beta_k \geq N_0 + \sum_{j\in\mathcal{K}\setminus\{k\}} |\mathbf{h}_{b_j,k}^H \mathbf{w}_j|^2, \forall k \in \mathcal{K} \tag{16d}$$

$$\sum_{k\in\mathcal{K}_b} ||\mathbf{w}_k||_2^2 \leq P_b, \ \forall b \in \mathcal{B} \tag{16e}$$

which is a concave-convex fractional program. The common way of solving concave-convex fractional program is to use the Dinkelbach method [40] which requires iterative processing. Here, we use a parameter-free approach based on the Charnes-Cooper transformation [36]. Specifically, the concave-convex fractional program can be transformed to an equivalent convex program with the transformations $\bar{\mathbf{w}}_k = \phi\mathbf{w}_k, \bar{\gamma}_k = \phi\gamma_k, \bar{\beta}_k = \phi\beta_k, \bar{r}_b = \phi r_b$ and $\phi = \frac{1}{\frac{1}{\eta}\sum_{k\in\mathcal{K}} ||\mathbf{w}_k||_2^2 + \sum_{b\in\mathcal{B}}(P_{\text{RD}}\delta(r_b) + P_{\text{CP},b})}$. As a result, solving (16) boils down to solving the following convex program

$$\max_{\bar{\mathbf{w}},\bar{\boldsymbol{\gamma}},\bar{\boldsymbol{\beta}},\bar{\mathbf{r}},\phi} \quad \sum_{b\in\mathcal{B}} \bar{r}_b \tag{17a}$$

$$\text{s. t.} \quad \bar{\gamma}_k - \Psi_k^{(n)}(\bar{\mathbf{w}}_k, \bar{\beta}_k)) \leq 0, \forall k \in \mathcal{K} \tag{17b}$$

$$\bar{r}_b - \alpha \sum_{k\in\mathcal{K}_b} \phi\log(1 + \frac{\bar{\gamma}_k}{\phi}) \leq 0, \forall b \in \mathcal{B} \tag{17c}$$

$$\phi N_0 + \sum_{j\in\mathcal{K}\setminus\{k\}} \frac{|\mathbf{h}_{b_j,k}^H \bar{\mathbf{w}}_j|^2}{\phi} - \bar{\beta}_k \leq 0, \forall k \in \mathcal{K} \tag{17d}$$

$$\frac{1}{\eta} \sum_{k\in\mathcal{K}} \frac{||\bar{\mathbf{w}}_k||_2^2}{\phi} + \sum_{b\in\mathcal{B}}(P_{\text{RD}}\phi\delta(\frac{\bar{r}_b}{\phi}) + \phi P_{\text{CP},b}) \leq 1 \tag{17e}$$

$$\sum_{k\in\mathcal{K}_b} \frac{||\bar{\mathbf{w}}_k||_2^2}{\phi} \leq \phi P_b, \forall b \in \mathcal{B}. \tag{17f}$$

The optimal solutions for problem (16) can be extracted as $\mathbf{w}_k^* = \frac{\bar{\mathbf{w}}_k^*}{\phi^*}, \gamma_k^* = \frac{\bar{\gamma}_k^*}{\phi^*}, \beta_k^* = \frac{\bar{\beta}_k^*}{\phi^*}, r_b^* = \frac{\bar{r}_b^*}{\phi^*}$. In the proposed algorithm, we iteratively approximate (10) by (17) until convergence. The proposed algorithm is presented in Algorithm 1. The monotonic convergence of the objective function for Algorithm 1 is not difficult to see. Specifically, due to linear approximation (15), the constraints (17b) become loose after each update of $\bar{\mathbf{w}}^{(n)}, \bar{\boldsymbol{\beta}}^{(n)}$. Furthermore, the updating rules ensure feasibility of the next iteration. These facts guarantee that $\sum_{b\in\mathcal{B}} \bar{r}_b^{(n+1)} \geq \sum_{b\in\mathcal{B}} \bar{r}_b^{(n)}$. Finally, the power constraint bounds each $\bar{r}_b$ from above. More detailed convergence analysis of the sequence of iterates and convergence point for the problem with similar structure can be found in [34, Appendix A]. We remark that Alg. 1 applies to any convex function $\delta(\cdot)$.

Remark 1. The function $\phi\delta(\frac{\bar{r}_b}{\phi})$ in constraint (17e) is a perspective transformation of $\delta(\bar{r}_b)$ which is also convex due to the convexity of $\delta(\bar{r}_b)$ [42]. In practice, it is computationally efficient to reformulate (17e) in a form which can be handled by dedicated powerful convex solvers. This is entirely possible depending on the type of $\delta(\bar{r}_b)$. Specifically, if $\delta(y) = y$ or $\delta(y) = y^2$, then (17e) immediately admits a second-order cone (SOC) representation. Let us now consider a general power model considered in [25] where $\delta(y) = y^m$ with constant $m > 1$ is always convex in the domain $y \geq 0$. For all practical purposes we can assume that m is a rational number without loss of optimality. Thus, there exist k and c such that $k \geq 1$ and $m = c/k$. By introducing a slack variable x_b, we can equivalently express (17e) as

$$\frac{1}{\eta} \sum_{k\in\mathcal{K}} \frac{||\bar{\mathbf{w}}_k||_2^2}{\phi} + \sum_{b\in\mathcal{B}}(P_{\text{RD}}x_b + \phi P_{\text{CP},b}) \leq 1 \tag{18a}$$

$$-\bar{r}_b \geq -\phi^{1-\frac{k}{c}} x_b^{\frac{k}{c}}, \forall b \in \mathcal{B}. \tag{18b}$$

Note that the constraint in (18b) is an inequality involving rational powers [43, Eq. (11)] and can be implemented as a series of SOC constraints as shown after [43, Eq. (11)].

Remark 2. If the power consumption has a linear dependence on the rate, i.e., $\delta(r_b) = r_b$ and P_{RD} is the same for all the BSs, then $P_{\text{RD}}\delta(r_b)$ does not affect the optimal variables of (10). In this special case, (10a) is equal to $\min(g(\mathbf{w})/\sum_{b\in\mathcal{B}} \tilde{R}_b(\mathbf{w}) + P_{\text{RD}})$. As can be seen, P_{RD} becomes a constant in the objective function and could be ignored in the optimization process without loss of optimality (but not in the actual utility). This means that the optimal beamformers of (10) and the problem considered in [17] would be equal in this special case.

B. Weighted Sum Energy Efficiency Maximization

The objective function of (11) is a sum-of-ratios objective function. A common approach to solve a sum-of-ratios maximization problem with concave-convex ratios is to transform it to a parameterized form with some fixed parameters, and then search the optimal parameters by solving a series of convex subproblems [16], [19]. Specifically, the general form of sum-of-fractional program $\max_{\mathbf{x}} \sum \frac{f_i(\mathbf{x})}{g_i(\mathbf{x})}$ can be solved as a series of subproblems $\max_{\mathbf{x}} \sum_i \alpha_i(f_i(\mathbf{x}) - \beta_i g_i(\mathbf{x}))$, where

Algorithm 1 Proposed SCA-based beamformer design for the network energy efficiency maximization problem.

Initialization: Set $n = 0$, and generate initial points $(\bar{\mathbf{w}}^{(n)}, \bar{\boldsymbol{\beta}}^{(n)})$.

1: **repeat**
2:   Solve (17) with $(\bar{\mathbf{w}}^{(n)}, \bar{\boldsymbol{\beta}}^{(n)})$ and denote optimal values as $(\bar{\mathbf{w}}^*, \bar{\boldsymbol{\beta}}^*)$.
3:   Update $(\bar{\mathbf{w}}^{(n)} = \bar{\mathbf{w}}^*, \bar{\boldsymbol{\beta}}^{(n)} = \bar{\boldsymbol{\beta}}^*)$.
4:   $n := n + 1$.
5: **until** desired accuracy
Output: $\mathbf{w}_k^* = \frac{\bar{\mathbf{w}}_k^*}{\phi^*}, \forall k \in \mathcal{K}$

Algorithm 2 Proposed SCA-based beamformer design for the WsumEEmax problem.

Initialization: Set $n = 0$, and generate initial points $(\mathbf{w}^{(n)}, \boldsymbol{\beta}^{(n)}, \mathbf{z}^{(n)}, \mathbf{r}^{(n)})$.

1: **repeat**
2:   Solve (23) with $(\mathbf{w}^{(n)}, \boldsymbol{\beta}^{(n)}, \mathbf{z}^{(n)}, \mathbf{r}^{(n)})$ and denote optimal values as $(\mathbf{w}^*, \boldsymbol{\beta}^*, \mathbf{z}^*, \mathbf{r}^*)$.
3:   Update $(\mathbf{w}^{(n)} = \mathbf{w}^*, \boldsymbol{\beta}^{(n)} = \boldsymbol{\beta}^*, \mathbf{z}^{(n)} = \mathbf{z}^*, \mathbf{r}^{(n)} = \mathbf{r}^*)$.

4:   $n := n + 1$.
5: **until** desired accuracy
Output: $\mathbf{w}_k^*, \forall k \in \mathcal{K}$

α_i, β_i are some parameters. At each iteration, α_i and β_i are fixed. After solving the parameterized program, α_i, β_i are updated according to a damped Newton method [44] (see [44] for details). However, in problem (11), both the numerator $f_i(\mathbf{x})$ and the denominator $g_i(\mathbf{x})$ in each EE function are non-convex, meaning that such a parametric approach would result in a multi-level iterative algorithm due to the non-convexity of each parameterized subproblem. To avoid this drawback, we propose an approach with only one iteration loop.

Here we again apply the SCA framework to solve (11). As the first step, we introduce new variables $\mathbf{t} \triangleq \{t_b\}_{b \in \mathcal{B}}$ to represent the energy efficiency of each cell and arrive at the following equivalent transformation of (11)

$$\max_{\mathbf{t}, \mathbf{w}} \quad \sum_{b \in \mathcal{B}} \omega_b t_b \tag{19a}$$

$$\text{s.t.} \quad t_b \leq \frac{\tilde{R}_b(\mathbf{w})}{g_b(\tilde{\mathbf{w}}_b) + P_{\text{RD}}\delta(\tilde{R}_b(\mathbf{w}))}, \forall b \in \mathcal{B} \tag{19b}$$

$$\sum_{k \in \mathcal{K}_b} ||\mathbf{w}_k||_2^2 \leq P_b, \ \forall b \in \mathcal{B} \tag{19c}$$

which is in fact an epigraph form of (11). Next, by introducing new user-specific SINR variables $\{\gamma_k\}_{k \in \mathcal{K}}$ and BS-specific rate variables $\{r_b\}_{b \in \mathcal{B}}$ as in network EEmax problem, we can further equivalently reformulate (19) as

$$\max_{\mathbf{t}, \mathbf{w}, \mathbf{r}, \boldsymbol{\gamma}} \quad \sum_{b \in \mathcal{B}} \omega_b t_b \tag{20a}$$

$$\text{s.t.} \quad t_b \leq \frac{r_b^2}{g_b(\tilde{\mathbf{w}}_b) + P_{\text{RD}}\delta(r_b)}, \forall b \in \mathcal{B} \tag{20b}$$

$$\gamma_k \leq \Gamma_k(\mathbf{w}), \forall k \in \mathcal{K} \tag{20c}$$

$$\alpha \sum_{k \in \mathcal{K}_b} \log(1 + \gamma_k) \geq r_b^2, \forall b \in \mathcal{B} \tag{20d}$$

$$\sum_{k \in \mathcal{K}_b} ||\mathbf{w}_k||_2^2 \leq P_b, \ \forall b \in \mathcal{B}. \tag{20e}$$

We have used r_b^2 in (20d) rather than r_b as in (16c) so that we can directly present (20b) as a difference of convex (DC) constraint. The equivalence between (20) and (19) is guaranteed, since all the constraints (20b)-(20d) are active at optimality. The constraints in (20c) are equivalent to (13c) and can be handled as shown in (14) and (15). To find a tractable reformulation of nonconvex constraints (20b), we can equivalently split it into the following two constraints

$$t_b \leq \frac{r_b^2}{z_b}, \forall b \in \mathcal{B} \tag{21a}$$

$$z_b \geq g_b(\tilde{\mathbf{w}}_b) + P_{\text{RD}}\delta(r_b), \forall b \in \mathcal{B}, \tag{21b}$$

where we have introduced new variables $\mathbf{z} \triangleq \{z_b\}_{b \in \mathcal{B}}$. Now,

(21b) is convex and (21a) is a DC constraint. Similarly to (14b), we can use the first-order lower approximation for the right side of (21a) as

$$\frac{r_b^2}{z_b} \geq \frac{2r_b^{(n)}}{z_b^{(n)}}r_b - \left(\frac{r_b^{(n)}}{z_b^{(n)}}\right)^2 z_b \triangleq \varphi_b^{(n)}(r_b, z_b). \tag{22}$$

With the linear approximations (15) and (22), we obtain an SCA-based iterative algorithm for solving (11) where the problem at iteration n reads

$$\max_{\mathbf{t}, \mathbf{r}, \mathbf{z}, \mathbf{w}, \boldsymbol{\gamma}, \boldsymbol{\beta}} \quad \sum_{b \in \mathcal{B}} \omega_b t_b \tag{23a}$$

$$\text{s.t.} \quad t_b \leq \varphi_b^{(n)}(r_b, z_b), \forall b \in \mathcal{B} \tag{23b}$$

$$\gamma_k \leq \Psi_k^{(n)}(\mathbf{w}_k, \beta_k), \forall k \in \mathcal{K} \tag{23c}$$

$$(11b), (20d), (21b), (14d). \tag{23d}$$

The proposed algorithm for the WsumEEmax problem is summarized in Algorithm 2. The same convergence discussions as in Algorithm 1 applies to Algorithm 2 also. In Appendix B, we further show that the WsumEEmax problem can be approximated as an SOCP at each iteration. It is worth pointing out that the SCA-based method was also used in [6] to solve the energy efficiency maximization problem in a single-cell system. Herein, however, different transformations are used due to the rate dependent power consumption. The main differences are summarized in Appendix C.

C. Feasible Initial Points

Finding a feasible initial point is an important issue for an SCA-based algorithm. For the NetEEmax problem, we can generate any beamformers $\mathbf{w}_k^{(0)}$ which satisfy the power constraints (which is easily done by normalization), then replace the equality (14d) with equality, i.e., calculate $\beta_k^{(0)} = N_0 + \sum_{j \in \mathcal{K} \setminus \{k\}} |\mathbf{h}_{b_j,k}^H \mathbf{w}_j^{(0)}|^2$. The resulting initial points $\mathbf{w}^{(0)}, \boldsymbol{\beta}^{(0)}$ are feasible. The feasible initial values of $\mathbf{w}^{(0)}, \boldsymbol{\beta}^{(0)}$ are also feasible to the WsumEEmax problem. However, due to the additional approximation in (23b), we also need to find initial $\mathbf{r}^{(0)}, \mathbf{z}^{(0)}$. To this end, we can first calculate $\gamma_k = \Gamma_k(\mathbf{w}^{(0)})$ according to (20c), then $r_b^{(0)} = \sqrt{\sum_{k \in \mathcal{K}_b} \log(1 + \gamma_k)}$ according to (20d), and $z_b^{(0)} = g_b(\tilde{\mathbf{w}}_b^{(0)}) + \delta(r_b^{(0)})$ which result in feasible initial points $\mathbf{w}^{(0)}, \boldsymbol{\beta}^{(0)}, \mathbf{r}^{(0)}, \mathbf{z}^{(0)}$.

IV. PROPOSED DECENTRALIZED SOLUTIONS

The algorithms presented in the preceding section require centralized processing. For the WsumEEmax problem, the

rate function $\tilde{R}_b(\mathbf{w})$ in (11a) couples all the cells due to the inter-cell interference (in the approximated problem, inter-cell interference appears in constraint (14d)). In the NetEEmax problem, also the total power consumption in the objective function (10a) (the objective (16a) in the approximated problem) couples all the cells. Due to the special structure of the WsumEEmax problem, we propose an alternative decentralized formulation which can be solved efficiently only relying on local channel state information and (scalar) backhaul information exchange. In particular, the proposed approach admits closed-form solutions and, thus, can be solved without invoking a generic external convex solver. We will also discuss the challenge and also possibility to solve the NetEEmax problem in a decentralized manner in the end of the section.

Let us start from problem (20). By adding 1 to both sides of (20c) and using fact that when optimal MMSE receiver is used, it holds that $1 + \Gamma_k(\mathbf{w}) = \frac{1}{\epsilon_k(\mathbf{w}, u_k)}$, where

$$\epsilon_k(\mathbf{w}, u_k) = |u_k|^2 \left(\sum_{j \in \mathcal{K}} |\mathbf{h}_{b_j,k}^H \mathbf{w}_j|^2 + N_0 \right) \\ - 2\mathrm{Re}(u_k \mathbf{h}_{b_k,k}^H \mathbf{w}_k) + 1 \quad (24)$$

is MSE and

$$u_k = \left(\sum_{j \in \mathcal{K}} \mathbf{h}_{b_j,k}^H \mathbf{w}_j \mathbf{w}_j^H \mathbf{h}_{b_j,k} + N_0 \right)^{-1} \mathbf{h}_{b_k,k}^H \mathbf{w}_k \quad (25)$$

is the MMSE receiver of user k [10]. Then, we can equivalently rewrite (20) as

$$\max_{\mathbf{t},\mathbf{r},\mathbf{z},\mathbf{w},\gamma,\mathbf{u}} \quad \sum_{b \in \mathcal{B}} \omega_b t_b \quad (26a)$$

$$\mathrm{s.t.} \quad t_b \le \frac{r_b^2}{z_b}, \forall b \in \mathcal{B} \quad (26b)$$

$$z_b \ge g_b(\tilde{\mathbf{w}}_b) + P_{\mathrm{RD}}\delta(r_b), \forall b \in \mathcal{B} \quad (26c)$$

$$\epsilon_k(\mathbf{w}, u_k) \le \frac{1}{1+\gamma_k}, \forall k \in \mathcal{K} \quad (26d)$$

$$\alpha \sum_{k \in \mathcal{K}_b} \log(1+\gamma_k) \ge r_b^2, \forall b \in \mathcal{B} \quad (26e)$$

$$\sum_{k \in \mathcal{K}_b} ||\mathbf{w}_k||_2^2 \le P_b, \ \forall b \in \mathcal{B}. \quad (26f)$$

The above problem is still nonconvex even for fixed receivers u_k. However, if the receivers u_k are fixed, all the other constraints are convex while (26b) and (26d) are DC constraints. The convex right hand side of (26b) can be linearized as in (22). To deal with (26d), we linearize $\frac{1}{1+\gamma_k}$ around the point $\gamma_k^{(n)}$ as

$$\frac{1}{1+\gamma_k} \ge (1+\gamma_k^{(n)})^{-1} - \frac{1}{(1+\gamma_k^{(n)})^2}(\gamma_k - \gamma_k^{(n)}). \quad (27)$$

As a result, problem (26) to find the beamformers for fixed receivers can be approximated as

$$\max_{\mathbf{t},\mathbf{r},\mathbf{z},\mathbf{w},\gamma} \quad \sum_{b \in \mathcal{B}} \omega_b t_b \quad (28a)$$

$$\mathrm{s.t.} \quad t_b \le \frac{2r_b^{(n)}}{z_b^{(n)}} r_b - \left(\frac{r_b^{(n)}}{z_b^{(n)}}\right)^2 z_b, \forall b \in \mathcal{B} \quad (28b)$$

$$z_b \ge g_b(\tilde{\mathbf{w}}_b) + P_{\mathrm{RD}}\delta(r_b), \forall b \in \mathcal{B} \quad (28c)$$

$$\epsilon_k(\mathbf{w}, u_k) \le \frac{1}{(1+\gamma_k^{(n)})} - \frac{\gamma_k - \gamma_k^{(n)}}{(1+\gamma_k^{(n)})^2}, \forall k \in \mathcal{K} \quad (28d)$$

$$\alpha \sum_{k \in \mathcal{K}_b} \log(1+\gamma_k) \ge r_b^2, \forall b \in \mathcal{B} \quad (28e)$$

$$\sum_{k \in \mathcal{K}_b} ||\mathbf{w}_k||_2^2 \le P_b, \ \forall b \in \mathcal{B}. \quad (28f)$$

Thus, for fixed receivers, the above convex problem can be run until convergence using the SCA approach. However, the monotonic convergence of the objective function is guaranteed even if we solve problem (28) only once after receiver update, then update the linearization point ((28b) and (28d)), followed again by receiver update and continue the procedure until convergence [34]. We refer to this algorithm as centralized method in the numerical results (e.g., Fig. 8). Problem (28) still requires centralized processing, since all the cells are coupled due to interference terms in the left hand side of (28d). To enable decentralized processing, we resort to Karush-Kuhn-Tucker (KKT) conditions [42] of (28). The Lagrangian of (28) can be written as

$$L(\mathbf{w}, \mathbf{t}, \mathbf{z}, \gamma, \mathbf{a}, \mathbf{c}, \mathbf{d}, \mathbf{f}, \mathbf{s}) =$$

$$\sum_{b \in \mathcal{B}} \omega_b t_b - \sum_{b \in \mathcal{B}} a_b(t_b - \frac{2r_b^{(n)}}{z_b^{(n)}} r_b + (\frac{r_b^{(n)}}{z_b^{(n)}})^2 z_b)$$

$$- \sum_{b \in \mathcal{B}} c_b(\frac{1}{\eta} \sum_{k \in \mathcal{K}_b} ||\mathbf{w}_k||_2^2 + P_{\mathrm{RD}}\delta(r_b) + P_{\mathrm{CP},b} - z_b)$$

$$- \sum_{k \in \mathcal{K}} d_k(|u_k|^2 (\sum_{j \in \mathcal{K}} |\mathbf{h}_{b_j,k}^H \mathbf{w}_j|^2 + N_0)$$

$$- 2\mathrm{Re}(u_k \mathbf{h}_{b_k,k}^H \mathbf{w}_k) + 1 - (1+\gamma_k^{(n)})^{-1}$$

$$+ (1+\gamma_k^{(n)})^{-2}(\gamma_k - \gamma_k^{(n)}))$$

$$- \sum_{b \in \mathcal{B}} f_b(r_b^2 - \alpha \sum_{k \in \mathcal{K}_b} \log(1+\gamma_k))$$

$$- \sum_{b \in \mathcal{B}} s_b(\sum_{k \in \mathcal{K}_b} ||\mathbf{w}_k||_2^2 - P_b) \quad (29)$$

where $\mathbf{w}, \mathbf{t}, \mathbf{z}, \mathbf{r}, \gamma$ are primal variables and $\mathbf{a}, \mathbf{c}, \mathbf{d}, \mathbf{f}, \mathbf{s}$ are dual variables related to constraints (28b) - (28f). The KKT conditions of (28) can be written as

$$\frac{\partial L}{\partial w_k} = \frac{1}{\eta} c_b \mathbf{w}_k + \sum_{j \in \mathcal{K}} d_j |u_j|^2 \mathbf{h}_{b_k,j} \mathbf{h}_{b_k,j}^H \mathbf{w}_k \\ - u_k \mathbf{h}_{b_k,k} d_k + s_b \mathbf{w}_k = 0, \forall k \in \mathcal{K} \quad (30a)$$

$$\frac{\partial L}{\partial t_b} = \omega_b - a_b = 0, \forall b \in \mathcal{B} \quad (30b)$$

$$\frac{\partial L}{\partial r_b} = \frac{2r_b^{(n)}}{z_b^{(n)}} a_b - c_b P_{\mathrm{RD}}\delta'(r_b) - 2f_b r_b = 0, \forall b \in \mathcal{B} \quad (30c)$$

$$\frac{\partial L}{\partial \gamma_k} = \gamma_k + 1 - \frac{f_b \alpha (1+\gamma_k^{(n)})^2}{d_k} = 0, \forall k \in \mathcal{K} \quad (30d)$$

$$\frac{\partial L}{\partial z_b} = c_b - a_b (\frac{r_b^{(n)}}{z_b^{(n)}})^2 = 0, \forall b \in \mathcal{B} \quad (30e)$$

$$\mathbf{c} \ge 0, \mathbf{d} \ge 0, \mathbf{f} \ge 0, \mathbf{s} \ge 0, \mathbf{a} \ge 0 \quad (30f)$$

$$a_b(t_b - \frac{2r_b^{(n)}}{z_b^{(n)}} r_b + (\frac{r_b^{(n)}}{z_b^{(n)}})^2 z_b) = 0, \forall b \in \mathcal{B} \quad (30g)$$

$$c_b(\frac{1}{\eta} \sum_{k \in \mathcal{K}_b} ||\mathbf{w}_k||_2^2 + P_{\mathrm{RD}}\delta(r_b) + P_{\mathrm{CP},b} - z_b) = 0, \forall b \in \mathcal{B} \quad (30h)$$

$$d_k(|u_k|^2 (\sum_{j \in \mathcal{K}} |\mathbf{h}_{b_j,k}^H \mathbf{w}_j|^2 + N_0) - 2\mathrm{Re}(u_k \mathbf{h}_{b_k,k}^H \mathbf{w}_k) \\ + 1 - (1+\gamma_k^{(n)})^{-1} + (1+\gamma_k^{(n)})^{-2}(\gamma_k - \gamma_k^{(n)})) = 0, \\ \forall k \in \mathcal{K} \quad (30i)$$

$$f_b(r_b^2 - \alpha \sum_{k \in \mathcal{K}_b} \log(1+\gamma_k)) = 0, \forall b \in \mathcal{B} \quad (30j)$$

$$s_b(\sum_{k \in \mathcal{K}_b} ||\mathbf{w}_k||_2^2 - P_b) = 0, \forall b \in \mathcal{B} \quad (30k)$$

$$(28b), (28c), (28d), (28e), (28f). \quad (30l)$$

In (30c), $\delta'(r_b)$ is derivative of $\delta(r_b)$. Equations (30b) and

(30e) immediately imply that $a_b = \omega_b$ and $c_b = \omega_b(\frac{r_b^{(n)}}{z_b^{(n)}})^2$. The beamformers can be solved from (30a) as

$$\mathbf{w}_k = d_k u_k(\sum_{j \in \mathcal{K}} d_j|u_j|^2 \mathbf{h}_{b_k,j}\mathbf{h}_{b_k,j}^H + s_b\mathbf{I} + \frac{1}{\eta}c_b\mathbf{I})^{-1}\mathbf{h}_{b_k,k} \quad (31)$$

where d_k and s_b are dual variables related to MSE constraints (28d) and power constraints (28f). The dual variables s_b are chosen to satisfy the power constraints (28f) using the bisection algorithm [11]. The MMSE receivers u_k are solved as given in (25). Since (28b)-(28e) hold with equality at the optimum, we can write

$$\gamma_k = -\epsilon_k(\mathbf{w}, u_k)(1 + \gamma_k^{(n)})^2 + (1 + 2\gamma_k^{(n)}) \quad (32a)$$

$$r_b = \sqrt{\alpha \sum_{k \in \mathcal{K}_b} \log(1 + \gamma_k)} \quad (32b)$$

$$z_b = g_b(\tilde{\mathbf{w}}_b) + P_{\text{RD}}\delta(r_b) \quad (32c)$$

$$t_b = \frac{2r_b^{(n)}}{z_b^{(n)}}r_b - (\frac{r_b^{(n)}}{z_b^{(n)}})^2 z_b. \quad (32d)$$

The dual variable f_b can be computed from (30c) as

$$f_b = \frac{2\frac{r_b^{(n)}}{z_b^{(n)}}\omega_b - \omega_b(\frac{r_b^{(n)}}{z_b^{(n)}})^2 P_{\text{RD}}\delta'(r_b)}{2r_b}. \quad (33)$$

Since the dual variables f_b and d_k depend on each other in (30d), one has to be fixed to optimize for the other. In the proposed method, d_k is fixed to evaluate f_b using (33). By solving d_k from (30d), the dual variable $d_k^{(i)}$ at iteration i is a point in the line segment between $d_k^{(i-1)}$ and $\frac{f_b^{(i)}(1+\gamma_k^{(n)})^2}{\gamma_k^{(i)}+1}$ determined by using a diminishing or a fixed step size $\rho^{(i)} \in [0,1]$, i.e.,

$$d_k^{(i)} = d_k^{(i-1)} + \rho^{(i)}(\frac{f_b^{(i)}\alpha(1+\gamma_k^{(n)})^2}{\gamma_k^{(i)}+1} - d_k^{(i-1)}). \quad (34)$$

As can be seen, $d_k^{(i)} = \frac{f_b^{(i)}\alpha(1+\gamma_k^{(n)})^2}{\gamma_k^{(i)}+1}$ is satisfied when d_k converges.

To summarize, the updates in the iterative algorithm are:

$$\mathbf{w}_k^{(i)} = d_k^{(i-1)}u_k^{(i-1)}(\sum_{j \in \mathcal{K}} d_j^{(i-1)}|u_j^{(i-1)}|^2\mathbf{h}_{b_k,j}\mathbf{h}_{b_k,j}^H$$
$$+ s_b\mathbf{I} + \frac{1}{\eta}c_b^{(i-1)}\mathbf{I})^{-1}\mathbf{h}_{b_k,k} \quad (35a)$$

$$u_k^{(i)} = (\sum_{j \in \mathcal{K}} \mathbf{h}_{b_j,k}^H\mathbf{w}_j^{(i)}(\mathbf{w}_j^{(i)})^H\mathbf{h}_{b_j,k} + N_0)^{-1}\mathbf{h}_{b_k,k}^H\mathbf{w}_k^{(i)} \quad (35b)$$

$$\gamma_k^{(i)} = -\epsilon_k(\mathbf{w}^{(i)}, u_k^{(i)})(1 + \gamma_k^{(i-1)})^2 + (1 + 2\gamma_k^{(i-1)}) \quad (35c)$$

$$r_b^{(i)} = \sqrt{\alpha \sum_{k \in \mathcal{K}_b} \log(1 + \gamma_k^{(i)})} \quad (35d)$$

$$z_b^{(i)} = \frac{1}{\eta}\sum_{k \in \mathcal{K}_b}||\mathbf{w}_k^{(i)}||_2^2 + P_{\text{CP},b} + P_{\text{RD}}\delta(r_b^{(i)}) \quad (35e)$$

$$t_b^{(i)} = \frac{2r_b^{(i-1)}}{z_b^{(i-1)}}r_b^{(i)} - (\frac{r_b^{(i-1)}}{z_b^{(i-1)}})^2 z_b^{(i)} \quad (35f)$$

$$f_b^{(i)} = \frac{2\frac{r_b^{(i-1)}}{z_b^{(i-1)}}\omega_b - \omega_b(\frac{r_b^{(i-1)}}{z_b^{(i-1)}})^2 P_{\text{RD}}\delta'(r_b^{(i)})}{2r_b^{(i)}} \quad (35g)$$

$$d_k^{(i)} = d_k^{(i-1)} + \rho(\frac{f_b^{(i)}\alpha(1+\gamma_k^{(i-1)})^2}{\gamma_k^{(i)}+1} - d_k^{(i-1)}) \quad (35h)$$

$$c_b^{(i)} = \omega_b(\frac{r_b^{(i-1)}}{z_b^{(i-1)}})^2. \quad (35i)$$

Note that (35f), (35g), (35h), and (35i) involve also $r_b^{(i-1)}$,

Algorithm 3 Proposed decentralized beamformer design for the WsumEEmax problem.

Initialization: Set $n = 0$, and generate initial points $(\gamma^{(n)}, \mathbf{z}^{(n)}, \mathbf{r}^{(n)})$.

1: **repeat**
2:     BS $b, \forall b$: Update $\mathbf{w}_k^{(i)}$ using (35a) and transmit precoded downlink pilots.
3:     User $k, \forall k$: Using the effective channel information, calculate $u_k^{(i)}$ using (35b)
4:     User $k, \forall k$: Signal $u_k^{(i)}$ to BSs using precoded pilots
5:     BS $b, \forall b$: Evaluate $\gamma_k^{(i)}, \forall k \in \mathcal{K}_b$, $r_b^{(i)}$, $f_b^{(i)}$, $z_b^{(i)}$, $t_b^{(i)}$, $c_b^{(i)}$ and $d_k^{(i)}, \forall k \in \mathcal{K}_b$
6:     BS $b, \forall b$: Share $d_k^{(i)}, \forall k \in \mathcal{K}_b$ to other BSs via backhaul

7:     $n := n + 1$.
8: **until** convergence or $n >$ predefined maximum number of iterations.

Output: $\mathbf{w}_k^*, \forall k \in \mathcal{K}$

$z_b^{(i-1)}$, and $\gamma_k^{(i-1)}$, i.e., these values are not updated until for the next iteration due to the fixed SCA step. The beamformer structure in (35a) resembles the one provided by the weighted MMSE method for weighted sum rate maximization (WSRmax) [10], [35]. Here, however, the beamformer involves additional scaling factor $\frac{1}{\eta}c_b$ to reflect the EE objective. Also, the stream specific scaling factor d_k reflects the EE utility (not just the inverse of MSE as for WSRmax). The proposed decentralized algorithm is summarized in Algorithm 3.[5]

Required Signaling: The beamformer update in (35a) requires information of the effective channels $u_j\mathbf{h}_{b_k,j}$ and dual variables d_k. The effective channels $u_j\mathbf{h}_{b_k,j}$ can be signaled to the BSs using precoded pilots, i.e., the UL pilots (sequences) are multiplied by the complex scalars (or vectors in multi-antenna case) [11]. Due to the fact that the power consumption of BS b depends on all the beamformers of BS b, d_k can be only evaluated at the BS as such. In this case, d_k should be shared via backhaul signaling to other BSs. This means that each BS needs to send K_b scalar values per iteration to other BSs via backhaul. In a centralized method, a central controller requires the global channel information, i.e., all the channel vectors in the network. Each complex channel coefficient consists of two scalar coefficients. Let us assume an equal number of L users per cell and N antennas at each BS. Each BS needs to share $2(B-1)LN$ scalars to a central controller which performs all the processing, and it is $2(B-1)N$ times the decentralized case. It is worth observing that the signaling overhead of the centralized method scales both with the number of BS antennas N and the number of users L, while only with L for the decentralized method. Thus, if the number of antennas is large, the decentralized method requires less signaling even when the channels are static for longer time. However, the benefits of the decentralized method become more important in the time-varying channels because the BSs can easily acquire local channel information, and they only need to exchange the weights d_k [14]. The centralized

---

[5]The proposed method can be used directly to the case with multiantenna receivers as well.

method would require sharing all the channel information every time when the channels change, which causes significant signaling overhead and delays. Another option would be to signal f_b to the users of own cell, and then d_k to BSs using uplink precoded pilots which could be done without any backhaul exchange. However, this would incur additional pilot resources. In practice, it can be more beneficial to perform multiple beamformer updates for fixed u_k's to improve the convergence speed in terms of over-the-air iterations and reduce the signaling load. We study the convergence behaviour numerically in Fig. 8.

Notes on Convergence of Algorithm 3: For the centralized algorithm (see discussions after (28)), the monotonic convergence of the objective function (28a) can be guaranteed similarly to the queue deviation problem in [34]. Because we solve the KKT conditions of (28) iteratively, the convergence of the distributed method could be guaranteed in a one special case. Specifically, we should first fix the receivers and linearization point $(\boldsymbol{\gamma}^{(n)}, \mathbf{z}^{(n)}, \mathbf{r}^{(n)})$, solve KKT conditions (30) (with fixed or diminishing step size ρ in (34)) until convergence, update linearization point and receivers again, and continue the procedure by solving KKT conditions with fixed receivers and linearization point. In fact, this kind of method would be equivalent to the centralized algorithm. However, since we combine all the updates into a single iteration loop (i.e., update of linearization point, receivers, and the iterative line search for the dual variables d_k) to improve the convergence speed, the formal convergence of the objective cannot be guaranteed. Fortunately, we have experimentally observed the convergence of the objective function to the centralized solution if the step size ρ is properly chosen. The choice of ρ in (34) also affects the convergence speed of the algorithm. In the simulations, we set $\rho = 0.15$. The convergence behaviour is numerically illustrated in Fig. 8.

Remark 3. Note that in order to enable decentralized implementation, BSs and users should alternately optimize the transmitters and receivers. If this is done sequentially, we need more pilot resources which reduces the available (time) resources for actual data transmission. Thus, there is a trade-off between the available over-the-air iterations and EE performance, depending on the coherence time of the channel (i.e., the size of the coherence block). In practice, we have to define the maximum number of over-the-air iterations which leaves sufficient amount of resources for data transmission [45].

Network EEmax problem: The problem of network EEmax is more challenging to implement in a decentralized manner, due to the fact that both sum rate and sum power couple the entire network since they appear in a single fraction in (10). However, by allowing more backhaul signaling, it is possible to arrive at a slightly modified method of Algorithm 3. Specifically, for the network EEmax, we replace (26) as

$$\max_{\mathbf{t},\mathbf{r},\mathbf{z},\mathbf{w},\boldsymbol{\gamma},p,\mathbf{u}} \quad \sum_{b\in\mathcal{B}} t_b \tag{36a}$$

$$\text{s.t.} \quad t_b \leq \frac{r_b^2}{p}, \forall b \in \mathcal{B} \tag{36b}$$

$$p \geq \sum_{b\in\mathcal{B}} z_b \tag{36c}$$

$$(26c) - (26f) \tag{36d}$$

where we have added new variable p for the ease of notation. Problem (36) is similar to the one in (26) except now the sum power p appears in constraint (36b) and, thus, links all the cells. By following the same steps to arrive at the updates in (35), all the other updates remain the same except $a_b = 1$, and $t_b^{(i)}, f_b^{(i)}$ and $c_b^{(i)}$ in (35f), (35g) and (35i) are replaced with

$$t_b^{(i)} = \frac{2r_b^{(i-1)}}{p^{(i-1)}} r_b^{(i)} - \left(\frac{r_b^{(i-1)}}{p^{(i-1)}}\right)^2 p^{(i)} \tag{37a}$$

$$f_b^{(i)} = \frac{2\frac{r_b^{(i-1)}}{p^{(i-1)}} - c_b^{(i-1)} P_{\text{RD}} \delta'(r_b^{(i)})}{2r_b^{(i)}} \tag{37b}$$

$$c_b^{(i)} = \sum_{b\in\mathcal{B}} \left(\frac{r_b^{(i-1)}}{p^{(i-1)}}\right)^2, \tag{37c}$$

respectively. Furthermore, we need to add update for p as

$$p^{(i)} = \sum_{b\in\mathcal{B}} z_b^{(i)}. \tag{38}$$

Equation (37c) implies that $c_b^{(i)}$ is actually the same for all the BSs, so c_b in the beamformer update (35a) reflects the total network EE, instead of BS specific EE as in the WsumEEmax problem. We can see that due to the coupling in (37c) and (38), the BS specific scalars $\left(\frac{r_b^{(i-1)}}{p^{(i-1)}}\right)^2$ and power consumption values z_b need to be exchanged between the BSs. This is additional signaling overhead compared to the WsumEEmax problem. Although we have numerically observed the convergence of the method (as in the WsumEEmax problem), it may be difficult to implement in practice due to the extra signaling overhead. However, it is a numerically efficient method, which can be realized without invoking any optimization solver.

V. Computational Complexity

Evaluating the computational complexity for Algorithms 1 and 2 is challenging because they are based on solving optimization problems. Moreover, the worst-case complexity models (e.g., in [46]) are in general rather conservative and do not give a realistic view. However, for Algorithm 3 and the closed-form solution for the NetEEmax presented in previous section, we can estimate the per-iteration complexity due to their closed-form structure. Thus, we use those as a baseline schemes for the complexity analysis. The proposed algorithms basically result in a similar beamformer structure (see, e.g., Eq. (31)) as the MMSE precoding in [23]. In fact, the expression in (31) is a weighted MMSE beamformer, where the weights d_k reflect the EE objective. The algorithm complexity is dominated by the beamformer expression, i.e., calculation of all the vector-vector multiplications and the matrix inversions. Thus, we can approximate the per-BS power consumption resulting from the beamformer optimization per iteration as

$$P_{\text{LP,iter}}^{\text{Alg. 3}} = \frac{W}{U}\left(\frac{N_b^3}{3L_{\text{BS}}} + \frac{3|\mathcal{K}|N_b^2 + 2N_b^2|\mathcal{K}_b| + \mathcal{O}(|\mathcal{K}|)}{L_{\text{BS}}}\right). \tag{39}$$

The above expression is different from the one used in [23]. Specifically, the work of [23] assumed massive MIMO setup with $|\mathcal{K}| < N_b$. In this paper, $|\mathcal{K}| > N_b$ throughout the simulations, and, thus, the zero-forcing beamformer does not exist. In the above expression, the power consumption scales cubicly with the number of antennas per BS in contrast to

the number of users as in [23]. Another difference is that a single-cell system was considered in [23], while in the multi-cell case considered herein, $|\mathcal{K}|$ in the expression $P_{\text{LP,iter}}$ is the total number of users in the network. $P_{\text{LP,iter}}$ assumes that the beamformer equation is solved by the standard Cholesky decomposition and forward/backward substitutions [47]. It is interesting to note that in fact the vector-vector multiplications $\mathbf{h}_k^H \mathbf{h}_k$ (which cause the term $3|\mathcal{K}|N_b^2$ in (39)) dominate the complexity when $|\mathcal{K}| > N_b$, because it is the term scaling with the total number of users in the network. All the other related computations are of linear complexity $\mathcal{O}(|\mathcal{K}|)$ and small compared to the other terms. It is worth observing that the complexities of the decentralized methods are significantly smaller than in Algorithms 1 and 2, which makes them tractable for energy-efficient processing.

VI. PILOT ALLOCATION STRATEGY

The pilot contamination in the network can be reduced by using more orthogonal pilot resources. However, this in turn decreases the available resources for data transmission. That is, there exists a trade-off between achieved energy efficiency and the number of pilot resources. In this section, we present a simple heuristic pilot allocation algorithm, which can achieve the energy efficiency trade-off explained above. A pilot allocation strategy has a significant impact on the pilot contamination, but a more detailed study of this topic is left for future work due to the space limitation.

If channel state information is perfectly known, the optimal pilot allocation could be found by exhaustive search, i.e., solving the energy efficiency optimization problem (network EEmax or WsumEEmax) for each possible pilot allocation combination. However, this is a combinatorial problem and intractable for a large network size. Moreover, in practice the optimization problem cannot be solved to find the best allocation because channel information is not known prior to pilot transmission. Thus, a pragmatic goal for the pilot contamination problem is to find a good pilot allocation strategy based on other prior information. In this paper, we propose a heuristic energy-efficient pilot allocation scheme which works as follows. If τ pilot resources are available (either in uplink or downlink), we assume that the users are divided into two different group sizes, $M_{\text{max}} = \left\lceil \frac{|\mathcal{K}|}{\tau} \right\rceil$ and $M_{\text{min}} = M_{\text{max}} - 1$, respectively. In each group of size M_{min} and M_{max}, M_{min} and M_{max} users share one pilot resource, respectively. To minimize the effect of pilot contamination, the number of smaller size groups X_{min} should be made as large as possible. Explicitly, we set $X_{\text{max}} = |\mathcal{K}| - (M_{\text{max}} - 1)\tau$ to be the number of larger size groups and $X_{\text{min}} = (|\mathcal{K}| - X_{\text{max}}M_{\text{max}})$. For example, if $|\mathcal{K}| = 21$ and $\tau = 12$, then $M_{\text{max}} = 2, M_{\text{min}} = 1, X_{\text{max}} = 9$, and $X_{\text{min}} = 3$, i.e., 3 users get orthogonal pilot resource and the others are divided into pairs. The first step is to find a good strategy to allocate the users in the smaller groups which can be expected to have larger impact on the total energy efficiency. In order to get a good guess of the best users, we form all the possible user combinations of size M_{min} and evaluate initial group specific energy efficiency metrics for each group based on known information (if $M_{\text{min}} = 1$, then

---

Algorithm 4 Proposed energy-efficient pilot allocation strategy.

1: Set $M_{\text{max}} = \left\lceil \frac{|\mathcal{K}|}{\tau} \right\rceil$, $M_{\text{min}} = M_{\text{max}} - 1$, $X_{\text{max}} = |\mathcal{K}| - (M_{\text{max}} - 1)\tau$, $X_{\text{min}} = (|\mathcal{K}| - X_{\text{max}}M_{\text{max}})$, and form all the possible group combinations of size M_{min}.
2: For each group $\mathcal{J}_l$, evaluate energy efficiency metrics κ_l according to (40).
3: Allocate the first X_{min} user groups having the largest κ_l's to the first X_{min} pilot resources.
4: For the remaining users, form all possible combinations of user groups of size M_{max}, and evaluate the metrics (40) for each group.
5: Allocate the remaining X_{max} pilot resources to user groups having the largest κ_l's.

---

these are user-specific metrics). Towards this end, let us define the group specific energy efficiency metric for group $\mathcal{J}_l$ as

$$\kappa_l = \frac{\sum_{k \in \mathcal{J}_l} \tilde{r}_k}{\frac{1}{\eta} \sum_{k \in \mathcal{J}_l} \frac{P_{b_k}}{|\mathcal{K}_{b_k}|} + |\mathcal{J}_l| \frac{\sum_{b \in \mathcal{B}} P_{\text{CP},b}}{|\mathcal{K}|} + P_{\text{RD}} \sum_{k \in \mathcal{J}_l} \tilde{r}_k^m}$$

(40)

where $\tilde{r}_k \triangleq \alpha \log(1 + \frac{\frac{P_{b_k}}{|\mathcal{K}_{b_k}|} \zeta_{b_k,k}}{\sum_{j \in \mathcal{J}_l \setminus \{k\}} \frac{P_{b_j}}{|\mathcal{K}_{b_j}|} \zeta_{b_j,k} + N_0})$, and $\zeta_{b_k,k}$ is the path gain (including shadowing) from BS b_k to user k. Intuitively, we suggest to use $\frac{P_{b_k}}{|\mathcal{K}_{b_k}|} \zeta_{b_k,k}$ as a metric to indicate the desired channel power (assuming equal power allocation per user) and $\frac{P_{b_j}}{|\mathcal{K}_{b_j}|} \zeta_{b_j,k}$ the interfering channel power from the serving BS of user j to user k. The second term in the denominator of (40) represents the average circuit power for the user group $\mathcal{J}_l$. By the proposed metric, the first X_{min} user groups are chosen based on having the largest κ_l's. After the smaller size groups have been allocated, we again form all the possible user combinations of size M_{max} from the remaining users which have not been allocated yet, and calculate the metrics (40) for each group $\bar{\mathcal{J}}_l$. By following the idea of the first phase, the remaining pilot resources are allocated for the user groups having the largest κ_l until all the users have been allocated. For the sake of completeness, the proposed energy-efficient pilot allocation strategy is presented in Algorithm 4.

VII. NUMERICAL RESULTS

We evaluate the performance for a quasistatic frequency flat Rayleigh fading channel model and consider 7-cell wrap-around model, where each user suffers interference from six neighboring base stations. We also assume a small-cell setup where the inter-BS distance d_{B} is 120 m. The radius of each cell is assumed to be $\frac{d_{\text{B}}}{2}$, i.e., the cell edges are overlapping and the users are randomly dropped to the cell edges to enable fairness for the users. The path loss in dB is modeled as 35 + 30log(d) with distance d in meters and the shadowing is modeled as log-normal distribution with a standard deviation of 8 dB [48]. We set $|\mathcal{K}_b| = L, N_b = N$ for all b, i.e., L is the number of users per cell and N is the number of antennas at each BS, and the algorithms are stopped when the change over the last five iterations is smaller than $\xi = 10^{-4}$. We set $\delta(y_b) = y_b^m$, where y_b is the data rate of BS b and $m \geq 1$

is any rational number, and the other simulation parameters adopted from [49], [23], [24], [25] are presented in Table I.

In all the figures except 5 and 6, we focus on algorithmic behavior, and, thus, use the same complexity model (39) for all the algorithms. In this regard, Figures 5 and 6 then demonstrate a more realistic view on the effect of complexity on the achieved EE of the proposed algorithms. In Figures 1-8, we set $\tau^{\text{ul}} = \tau^{\text{dl}} = |\mathcal{K}|$, i.e., there is no pilot contamination, and in Figure 9, the effect of pilot contamination is demonstrated.

TABLE I
SIMULATION PARAMETERS.

Parameter	Value
Inter-BS distance	120 m
Cell radius	60 m
W	20 MHz
Path loss model	$35 + 30\log(d[\text{m}]) + \mathcal{N}(0, 8)$ dB
P_b	27 dBm
U	100
ω_b	1
N_0	-98 dBm
P_{FIX}	3 Watt
P_{BS}	0.4 Watt
P_{SYN}	1 Watt
P_{UE}	0.1 Watt
P_{LP}	Eq. (8)
P_{CE}	0.05 Watt
L_{BS}	12.8 Gflops/W
Q	20
η	0.2

Figures 1 and 2 plot the convergence and cumulative distribution functions (CDFs) of Algs. 1 and 2. We can see that both algorithms converge relatively fast in the considered setting and the obtained solution is insensitive to initial points. We have numerically observed that the proposed algorithms are stabilized (i.e., fairly close to the convergent point) after around 10 iterations. The convergence speed could be further increased by good initial points. We also show the cumulative distribution function of the number of iterations for different values of ξ which also verifies a fast convergence speed.

Figure 3 illustrates the average energy efficiency as a function of P_{RD} for different exponent values m. We compare the proposed algorithm with the existing method in [17] (dubbed as "DB-WMMSE", referring to the combination of the Dinkelbach and WMMSE algorithm) where the rate dependent power consumption is not taken into account in the optimization problem. However, to have a fair comparison, after solving the energy efficiency problem with $P_{\text{RD}} = 0$, the EE value plotted in Fig. 3 includes also the impact of rate dependent power. As analyzed mathematically in Remark 2, the rate dependent term does not affect the solution of the network EEmax when $m = 1$. However, for a general model $m > 1$, the proposed algorithm can provide up to 60% gain in the considered setting, showing the importance of including the rate dependent power consumption in the optimization. Note that the gains of Alg. 1 depend on the setup. Larger gains can be achieved in the systems with low transmit power, where the rate dependent signal processing power consumption is a significant part of the total power consumption. Figure 3 also reveals that the power model has a huge impact on the energy efficiency which shows the importance of accurate power modeling.
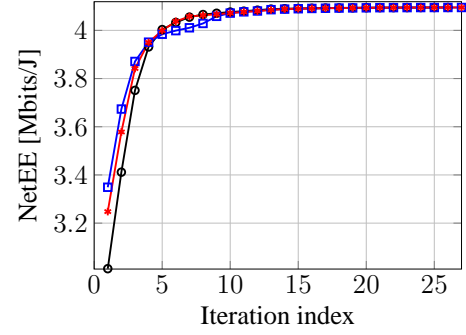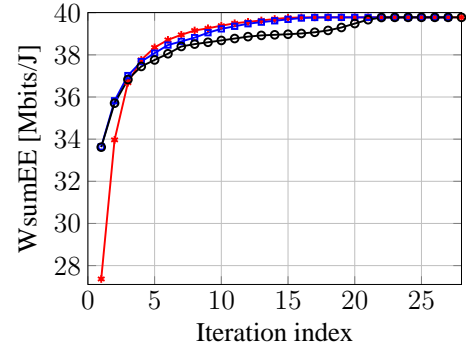

(a) Convergence illustration for three different initial points
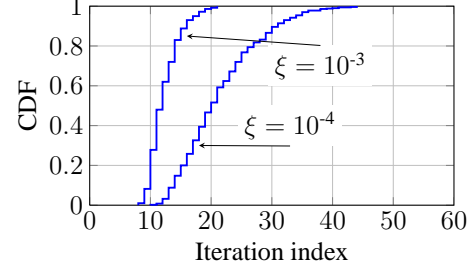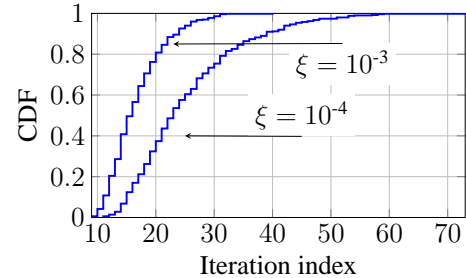

(b) Average cumulative distribution function

Fig. 1. Convergence illustration of Algorithm 1 with $N = 4$, $L = 2$, $m = 1.2$, $P_{\text{RD}} = 2.4$ [W/(Gbits/s)^m].

(a) Convergence illustration for three different initial points

(b) Average cumulative distribution function

Fig. 2. Convergence illustration of Algorithm 2 with $N = 4$, $L = 2$, $m = 1.2$, $P_{\text{RD}} = 2.4$ [W/(Gbits/s)^m].

Figure 4 compares the WsumEE performance with different exponent values m for the same simulation parameters as in Fig. 3. Here, we similarly compare Alg. 2 with the existing method in [19] (labelled by "Parametric") where the rate dependent power consumption is not taken into account in the optimization problem. We can see that for the WsumEE,
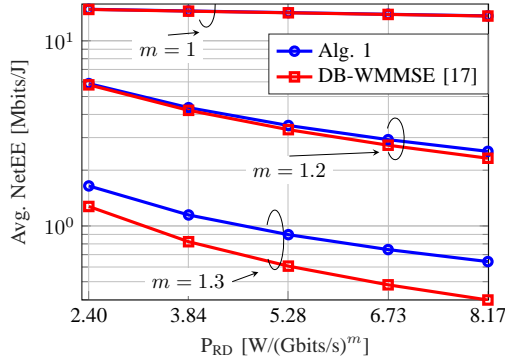
Fig. 3. Energy efficiency comparison of the algorithms for different rate dependent power consumption models with $N = 4$, $L = 2$.

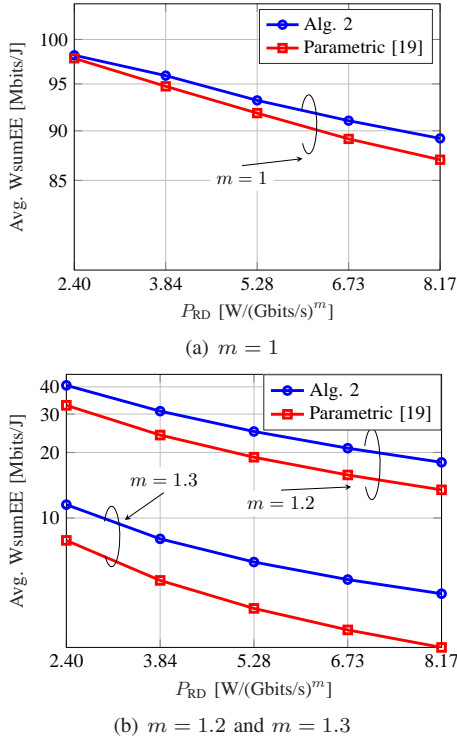(a) $m = 1$

(b) $m = 1.2$ and $m = 1.3$

Fig. 4. Weighted sum energy efficiency comparison of the algorithms for different rate dependent power consumption models with $N = 4$, $L = 2$.

Algorithm 2 offers performance improvement compared to the traditional method even when linear rate dependent power consumption model is used (i.e., $m = 1$).

Figure 5 shows the average energy efficiency for the different numbers of transmit antennas N. Since the complexity evaluation is difficult for Algs. 1 and 2 as discussed in Section V, we plot the performance of the closed-form solutions presented in Section IV. The closed-form solution for the NetEEmax presented in the end of Section IV is labeled as *'netEE, KKT'*. Note that in this figure, the actual network EE is also shown for Alg. 3. Specifically, the beamformers are calculated with Alg. 3, i.e., using the WsumEEmax as a design criterion, and the obtained beamformers are then used to calculate the objective of NetEEmax. We compare the proposed algorithms with various conventional beamforming methods such as the uncoordinated method, the orthogonal access method, and MMSE beamforming with multi-

cell and single-cell processing. The uncoordinated and the orthogonal methods require no coordination, i.e., each BS tries to maximize its own EE. This means that their power consumption due to the complexity scales only with the number of users in each cell, i.e., $|\mathcal{K}|$ is replaced with $|\mathcal{K}_b|$ in (39). In the uncoordinated method, all the BSs use all the bandwidth without any coordination, thereby causing severe inter-cell interference. In orthogonal access, the bandwidth is divided into 7 orthogonal sub-bands so that each BS occupies $W/7$ bandwidth, and, thus, the noise power level is also 7 times lower. In the MMSE precoding, we define $\mathbf{w}_k = \sqrt{p_k}\tilde{\mathbf{w}}_k$, where p_k is the transmit power for user k and $\tilde{\mathbf{w}}_k = (\mathbf{I}_N + \sum_{j\in\mathcal{K}} \frac{P_b}{LN_0}\mathbf{h}_{b_k,j}\mathbf{h}_{b_k,j}^H)^{-1}\mathbf{h}_{b_k,k}/||(\mathbf{I}_N + \frac{P_b}{LN_0}\sum_{j\in\mathcal{K}} \mathbf{h}_{b_k,j}\mathbf{h}_{b_k,j}^H)^{-1}\mathbf{h}_{b_k,k}||_2$ is the normalized MMSE beamforming vector. Then, the expressions $\tilde{\mathbf{w}}_k$ are plugged into problem (14), and the problem is iteratively solved by optimizing the power allocation to maximize the EE, while keeping $\tilde{\mathbf{w}}_k$ fixed. The *multi-cell MMSE* optimizes the power allocation to maximize the network EE, while the *single-cell MMSE* tries to maximize BS-specific energy efficiencies. For the sake of fair comparison, it is assumed that the MMSE precoding method is a non-iterative method, i.e., $Q = 1$ in the power consumption model (39), while we fix $Q = 20$ in the proposed algorithms. Thus, the MMSE consumes 1/20 fraction of the computational power of Alg. 3 and the 'netEE, KKT' method. We can see that the energy efficiency increases with the number of antennas even though the power consumption due to the complexity (and also the number of RF chains) increases rapidly (see (39)). It is also observed that significant energy efficiency gains are achieved over all the other methods even though power consumption related to the computational complexity is higher. The large gains of the proposed methods over the MMSE precoding also reveal that optimizing the beamforming directions has a significant impact on the energy efficiency. The performance differences between the MMSE method and the proposed algorithms can be partly explained by looking, e.g., the MMSE beamformer structure and the optimized beamformer structure in (31). We can see that the MMSE beamformer gives equal weights to all the user channels, while the proposed beamformer structure maximizes the desired EE objective by adjusting the weighting towards each direction. More specifically, the relative importance of the cancelled/suppressed directions can be controlled by adjusting the user specific weights depending on the scenario. For example, in some scenarios it might be beneficial to assign (near) zero weights for a subset of users to maximize the EE objective. This would enable better interference controlling capability towards users with nonzero weights. This flexibility does not exist in the MMSE beamformer case, where the weights are assumed to be equal. In practice, there exists a trade-off between the number of iterations for the beamformer calculation and the achieved energy efficiency, because each iteration consumes some power but produces better beamformers toward the final ones. This trade-off is illustrated in the next experiment.

Fig. 6 demonstrates the effect of algorithm complexity on the average energy efficiency of Algorithm 3. More specif-
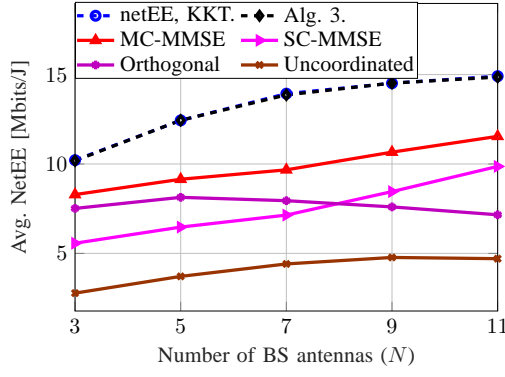
Fig. 5. Average energy efficiency vs. the number of TX antennas N with $L = 3$, $P_{RD} = 2.4$ [W/(Gbits/s)^m], $m = 1$.
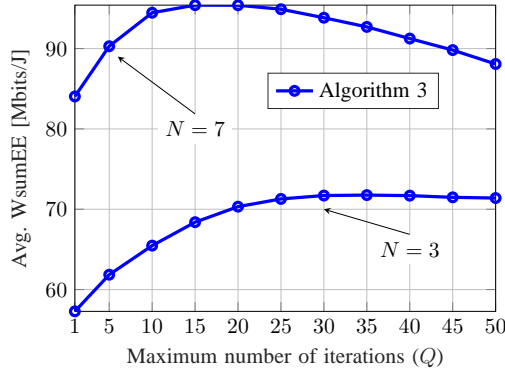

Fig. 6. Average WsumEE vs. the maximum number of iterations Q with $L = 3$, $P_{RD} = 2.4$ [W/(Gbits/s)^m], $m = 1$.
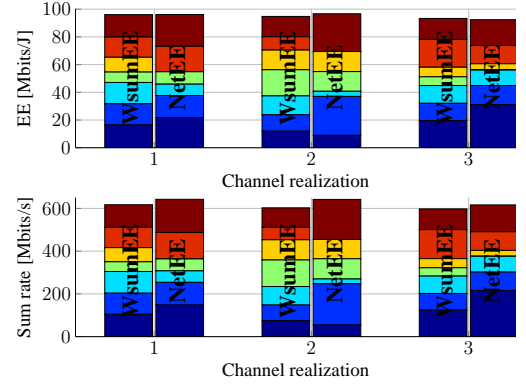

Fig. 7. Comparison of individual energy efficiencies (upper figure) and sum rates (lower figure) of the cells achieved in network EEmax (right bar) and weighted sum EEmax (left bar) with $N = 4$, $L = 2$, $m = 1$, $P_{RD} = 2.4$ [W/(Gbits/s)^m]. The colors denote different base stations.
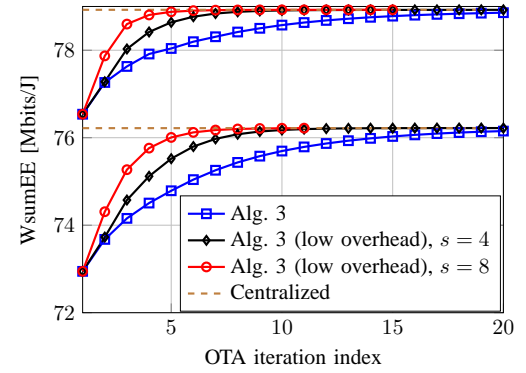

Fig. 8. Convergence of Alg. 3 in terms of over-the-air iterations with $\rho = 0.15$, $N = 4$, $L = 2$, $P_{RD} = 2.4$ [W/(Gbits/s)^m], $m = 1$.

ically, since each iteration of the algorithm consumes some amount power according to the power consumption model in (39), we set maximum number of iterations Q and stop the algorithm after Q iterations which is shown in the x-axis. It is observed that the best number of iterations depends on the number of antennas. For example, the best WsumEE is achieved approximately in 15-20 iterations when $N = 7$, while more iterations can be run for smaller number of antennas, because the complexity is lower. The performance decreases after a particular number of iterations due to the fact that the power consumption increase starts to dominate the achieved gain from the beamformer updates.

Figure 7 compares individual base station energy efficiencies (upper figure) and sum rates (lower figure) achieved in the NetEEmax and WsumEEmax problems for three random channel realizations. As can be seen, when equal weights are used for all the BSs, the WsumEEmax clearly balances the energy efficiencies and rates between the cells with only small performance degradation in the network energy efficiency (also see Figure 5 for further discussions). As far as fair resource allocation is concerned, the WsumEEmax design criterion proves to be a good choice.

Figure 8 illustrates the convergence of Alg. 3 in terms of over-the-air (OTA) iterations for two different channel realizations. By OTA iteration we basically mean receiver update because it is a step which requires over-the-air downlink and uplink signaling. We also consider a variant of Alg. 3 (dubbed as "Alg. 3 (low overhead)" in Fig. 8) where after each receiver

update, the beamformers are updated s times at the BS side using only backhaul exchange. This means that steps 3-4 of Alg. 3 are performed less frequently, i.e., the pilot overhead due to precoded pilot transmission is reduced. As can be seen, all the variants of Alg. 3 converge to the centralized solution which is calculated by centralized alternating optimization between the receivers and beamformers (see the discussion after (28)). We can see that the OTA signaling overhead can be reduced significantly with the variant method.

In Figure 9, we illustrate the effect of pilot contamination on the network energy efficiency performance using a heuristic pilot allocation strategy provided in Section VI, i.e., we use Algorithm 1 together with the pilot allocation algorithm. Let us recall that the pilot contamination in the network can be reduced by using more orthogonal pilot resources. However, this in turn decreases the available resources for data transmission. Thus, there exists an energy efficiency trade-off between the number of pilot resources and the effect of pilot contamination. Due to the limited resources, orthogonal allocation may not be possible in practice. Fortunately, it is unnecessary to use orthogonal resources for all the users because there may be some groups of users in which the users cause only small interference to each other possibly due to large spatial separation. Accordingly, these user groups could be allocated by the same resources to save resources for data transmission. This, however, requires advanced pilot allocation
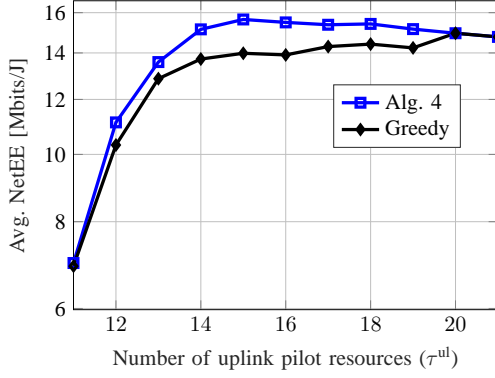
Fig. 9. The effect of pilot contamination on the energy efficiency with $N = 5, L = 3, P_{RD} = 2.4$ [W/(Gbits/s)^m], $m = 1$.

schemes. In Figure 9, we focus on uplink pilot allocation, and simply keep the downlink pilot allocation orthogonal. The proposed Algorithm 4 is compared with the conventional "greedy" method, where the orthogonal users are first allocated based on maximum path gains (including shadowing), and then the remaining user pairs are allocated according to maximum sum path gain metrics, i.e., the metric κ_l in steps 2-5 of Alg. 4 is replaced with $\sum_{k \in \mathcal{J}_k} \zeta_{b_k,k}$, and the user group having the largest κ_l is allocated first. We can see that the proposed Algorithm 4 achieves better energy efficiency than the greedy method, and it is obtained with non-orthogonal allocation using $\tau^{ul} = 15$ pilot resources. Note that when $\tau^{ul} = 15$, 36% of resources are wasted on DL and UL pilot signaling while the orthogonal allocation requires 42% of the resources. This is the reason why reducing the UL pilot signaling can exceed the loss from less accurate beamforming. This interesting result verifies the existence of a trade-off between the number of pilot resources and the effect of pilot contamination with Algorithm 4. When the available number of pilot resources is large, the greedy method gives only slightly worse performance than Alg. 4 because there are so many orthogonal users that it is sufficient to rely on path gains when allocating the resources. However, when the number of pilot resources becomes smaller, the performance of the greedy method is significantly inferior. Note that Fig. 9 is just a heuristic comparison of the pilot contamination effect, and a more detailed study would require its own body of work, which we would like to leave for future work. The existing state-of-the-art literature for this matter can be found, e.g., in [50]–[53].

VIII. CONCLUSIONS

This paper has studied multi-cell energy-efficient coordinated beamforming with a rate dependent power consumption model. We have considered two different optimization criteria: network energy efficiency maximization and weighted sum energy efficiency maximization. The framework for the proposed solutions has been based on the successive convex approximation principle. We have further proposed alternative formulations which enable decentralized closed-form implementations using only local channel state information and limited backhaul signaling. To reduce the effect of pilot contamination, a heuristic pilot allocation strategy has also been proposed in the paper. The numerical results have illustrated

that the rate dependent power consumption has a significant impact on the energy efficiency and has to be taken into account when devising energy-efficient transmission strategies. The proposed methods have been shown to outperform various conventional beamformer designs. We have also demonstrated that the pilot contamination has a significant impact on the EE performance and showed that the proposed energy-efficient pilot allocation strategy can be used to achieve significant improvements when non-orthogonal pilot resources are used.

APPENDIX A
REQUIRED CHANGES FOR ALTERNATIVE POWER CONSUMPTION MODEL

In case the rate dependent power consumption of BS b is modeled as $\sum_{k \in \mathcal{K}_b} P_{RD} \delta(r_k)$, where $\delta(r_k)$ is a function of individual user rate r_k, we replace r_b with r_k and apply this to the corresponding constraints so that the network EEmax problem (17) is replaced by

$$\max_{\bar{\mathbf{w}}, \bar{\boldsymbol{\gamma}}, \bar{\boldsymbol{\beta}}, \bar{\mathbf{r}}, \phi} \quad \sum_{k \in \mathcal{K}} \bar{r}_k \tag{41a}$$

$$\text{s. t.} \quad \bar{r}_k - \phi \log(1 + \tfrac{\bar{\gamma}_k}{\phi}) \le 0, \forall k \in \mathcal{K} \tag{41b}$$

$$\sum_{k \in \mathcal{K}} \left(\tfrac{1}{\eta} \tfrac{\|\bar{\mathbf{w}}_k\|_2^2}{\phi} + P_{RD} \phi \delta(\tfrac{\bar{r}_k}{\phi}) \right) + \phi \sum_{b \in \mathcal{B}} P_{CP,b} \le 1 \tag{41c}$$

$$(17b), (17d), (17f) \tag{41d}$$

where $\bar{\mathbf{r}} \triangleq \{r_k\}_{k \in \mathcal{K}}$. Following the same principle, the WsumEEmax problem in (23) is replaced by

$$\max_{\mathbf{t}, \mathbf{z}, \mathbf{w}, \boldsymbol{\gamma}, \boldsymbol{\beta}, \mathbf{r}} \quad \sum_{b \in \mathcal{B}} \omega_b t_b \tag{42a}$$

$$\text{s. t.} \quad t_b \le \varphi_b^{(n)}(\{r_k\}_{k \in \mathcal{K}_b}, z_b), \forall b \in \mathcal{B} \tag{42b}$$

$$z_b \ge g_b(\tilde{\mathbf{w}}_b) + P_{RD} \sum_{k \in \mathcal{K}_b} \delta(r_k), \forall b \in \mathcal{B} \tag{42c}$$

$$\log(1 + \gamma_k) \ge r_k^2, \forall k \in \mathcal{K} \tag{42d}$$

$$(11b), (14d), (23c) \tag{42e}$$

where $\varphi_b^{(n)}(\{r_k\}_{k \in \mathcal{K}_b}, z_b)$ is now a first-order approximation of $\sum_{k \in \mathcal{K}_b} r_k^2 / z_b$ instead of r_b^2 / z_b.

APPENDIX B
ITERATIVE SOCP APPROXIMATION OF THE WSUMEEMAX PROBLEM

We note that (21b) is similar to (17e) and thus admits SOC representation if we consider a power model $\delta(y) = y^m$ as in Remark 1 (see Remark 1 for details). When this model is used, it is straightforward to show that all the constraints in (23) admit SOC form except the one in (20d) which is an exponential cone. In our recent work [37], we showed that an exponential cone can be approximated as a system of SOC constraints which results from the Taylor expansion (to several orders) of the exponential function. However, the number of slack variables introduced in such an approximation increases quickly with the number of users. Thus, herein we propose a novel way to approximate (20d) as SOC constraint. In light of SCA principle, we need to find a concave lower bound for $\log(1 + x)$. Let us denote $h(x) = -\log(1 + x)$. Then we have

$$\|\nabla h(x_1) - \nabla h(x_2)\|_2 = \left\|-\frac{1}{1+x_1} + \frac{1}{1+x_2}\right\|_2$$

$$= \left\|\frac{x_1 - x_2}{(1+x_1)(1+x_2)}\right\|_2 \overset{(a)}{\leq} \|x_1 - x_2\|_2 \quad (43)$$

for $x_1, x_2 \geq 0$. Note that the inequality (a) above is due to $(1+x_1)(1+x_2) > 1$ for $x_1, x_2 > 0$. That is to say, the gradient of $h(x)$ is Lipschitz continuous with parameter $L = 1$ on the domain $x \geq 0$. As a result, the function $f(x) = \frac{L}{2}x^2 - h(x)$ is convex, and thus we have [54]

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) \quad (44)$$

which is equivalent to

$$h(y) \leq h(x) + \nabla h(x)^T (y - x) + \frac{L}{2}(y - x)^2 \quad (45)$$

for all $x, y \geq 0$. Substituting $h(x) = -\log(1 + x)$ into the above inequality leads to

$$\log(1 + y) \geq \log(1 + x) + \frac{1}{1+x}(y - x) - \frac{L}{2}(y - x)^2 \quad (46)$$

Now it is clear that we have the following inequality

$$\sum_{k \in \mathcal{K}_b} \log(1 + \gamma_k) \geq \sum_{k \in \mathcal{K}_b} \big(\log(1 + \gamma_k^{(n)}) + \frac{1}{1+\gamma_k^{(n)}}(\gamma_k - \gamma_k^{(n)}) - \frac{L}{2}(\gamma_k - \gamma_k^{(n)})^2\big). \quad (47)$$

Note the above inequality is tight and the first derivative of both sides with respect to γ_k are equal when $\gamma_k = \gamma_k^{(n)}, \forall k \in \mathcal{K}_b$, which satisfies the conditions of SCA framework. In summary, (20d) can be approximated as

$$\sum_{k \in \mathcal{K}_b} \alpha\big(\log(1 + \gamma_k^{(n)}) + \frac{\gamma_k - \gamma_k^{(n)}}{1 + \gamma_k^{(n)}} - \frac{L}{2}(\gamma_k - \gamma_k^{(n)})^2\big) \geq r_b^2 \quad (48)$$

which is an SOC constraint. Using the approximation (47), we replace convex constraints (20d) with (48) in (23), and arrive at an SOCP in step 2 of Algorithm 2, which is much more efficiently solved than the generic convex formulation in (23).

APPENDIX C
COMPARATIVE DISCUSSIONS TO [6]

In [6], the successive convex approximation method was also used to solve the energy efficiency maximization problem in a single-cell system. In principle, one could extend the ideas from [6] to solve the problems considered in this paper. However, herein we use different transformations which have many advantages compared to the one in [6], even when the rate dependent power would be ignored. Here we summarize the main differences.

Let us first consider the network EEmax problem. First, in [6], the objective function is linearized in Eq. (22), followed by an introduction for new variables for the energy efficiency and the total power. In the proposed method, we only introduce new variables for the rate function in (12), which implies that the proposed approach has fewer variables. Second, the linearization of the objective function in Eq. (22) of [6] also yields the fact that two different sets of approximated constraints are required in order to use the SCA method. In the proposed method, we use linear approximation for

one set of constraints only, which can be expected to result in better performance in a general case. Third, in [6], the linear approximations are performed for the jointly concave geometric mean functions $\sqrt{ab}$. In the proposed method, we use linear approximations for the jointly convex quadratic-over-linear function $\frac{a^2}{b}$. In [6], the right side approximation in (30b) can be written as

$$\sqrt{(a-1)b} \leq \sqrt{(a^{(n)} - 1)b^{(n)}} + \frac{1}{2}\big(\frac{a^{(n)}-1}{b^{(n)}}\big)(b - b^{(n)}) + \frac{1}{2}\big(\frac{b^{(n)}}{a^{(n)}-1}\big)(a - a^{(n)}) \quad (49)$$

In the above approximation, the denominator in $\frac{1}{2}\big(\frac{b^{(n)}}{a^{(n)}-1}\big)$ can go to zero if the rate goes to zero, which can lead to a numerical problem (i.e., division by zero). In the proposed method, the denominator in (15) does not have this problem because β_k (denoting the total interference-plus-noise) is always bounded below by the noise power. This is a particularly important point that needs to be accounted for in the considered problem, because we do impose no user-specific quality-of-service constraints.

In case of the WsumEEmax problem, since the objective function is a sum of fractional functions and the Charnes-Cooper transformation cannot be used, we linearize the objective function as done in [6]. However, we again approximate quadratic-over-linear functions instead of the geometric mean functions, which yields the fact that the constraints due to zero rates (as discussed above) cannot go to infinity in any case. Also, in [6], to approximate each convex problem as an SOCP, the exponential cone $x \geq e^y$ was approximated as a set of SOC constraints. However, this way introduces a lot of slack variables as we discuss in Appendix B above. In the proposed method presented in Appendix B, we propose a concave lower bound for $\log(1 + x)$, which does not introduce any extra variables, meaning that the complexity is significantly reduced compared to the method in [6].

REFERENCES

[1] A. Osseiran, F. Boccardi, V. Braun, K. Kusume, P. Marsch, M. Maternia, O. Queseth, M. Schellmann, H. Schotten, H. Taoka, H. Tullberg, M. Uusitalo, B. Timus, and M. Fallgren, "Scenarios for 5G mobile and wireless communications: the vision of the METIS project," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 26–35, May 2014.

[2] C. Isheden, Z. Chong, E. Jorswieck, and G. Fettweis, "Framework for link-level energy efficiency optimization with informed transmitter," *IEEE Trans. Wireless Commun.*, vol. 11, no. 8, pp. 2946–2957, Aug. 2012.

[3] 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.

[4] L. Venturino, A. Zappone, C. Risi, and S. Buzzi, "Energy-efficient scheduling and power allocation in downlink OFDMA networks with base station coordination," *IEEE Trans. Wireless Commun.*, vol. 14, no. 1, pp. 1–14, Jan. 2015.

[5] C. Jiang and L. Cimini, "Downlink energy-efficient multiuser beam-forming with individual SINR constraints," in *IEEE MILCOM*, 2011, pp. 495–500.

[6] O. Tervo, L.-N. Tran, and M. Juntti, "Optimal energy-efficient transmit beamforming for multi-user MISO downlink," *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5574–5588, Oct. 2015.

[7] K. Nguyen, L. Tran, O. Tervo, Q. Vu, and M. Juntti, "Achieving energy efficiency fairness in multicell MISO downlink," *IEEE Commun. Lett.*, vol. 19, no. 8, pp. 1426–1429, Aug. 2015.

16

[8] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H.-P. Mayer, L. Thiele, and V. Jungnickel, "Coordinated multipoint: Concepts, performance, and field trial results," *IEEE Commun. Mag.*, vol. 49, no. 2, pp. 102–111, Feb. 2011.

[9] 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.

[10] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331–4340, Sep. 2011.

[11] P. Komulainen, A. Tölli, and M. Juntti, "Effective CSI signaling and decentralized beam coordination in TDD multi-cell MIMO systems," *IEEE Trans. Signal Process.*, vol. 61, no. 9, pp. 2204–2218, May 2013.

[12] L. Tran, M. Hanif, A. Tölli, and M. Juntti, "Fast converging algorithm for weighted sum rate maximization in multicell MISO downlink," *IEEE Signal Process. Lett.*, vol. 19, no. 12, pp. 872–875, Dec. 2012.

[13] H. Dahrouj and W. Yu, "Coordinated beamforming for the multicell multi-antenna wireless system," *IEEE Trans. Wireless Commun.*, vol. 9, no. 5, pp. 1748–1759, May 2010.

[14] A. Tölli, H. Pennanen, and P. Komulainen, "Decentralized minimum power multi-cell beamforming with limited backhaul signaling," *IEEE Trans. Wireless Commun.*, vol. 10, no. 2, pp. 570–580, Feb. 2011.

[15] H. Pennanen, A. Tölli, and M. Latva-aho, "Decentralized coordinated downlink beamforming via primal decomposition," *IEEE Signal Process. Lett.*, vol. 18, no. 11, pp. 647–650, Nov 2011.

[16] Q. Wu, W. Chen, D. W. K. Ng, J. Li, and R. Schober, "User-centric energy efficiency maximization for wireless powered communications," *IEEE Trans. Wireless Commun.*, vol. 15, no. 10, pp. 6898–6912, Oct 2016.

[17] 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, 2013.

[18] Y. Li, Y. Tian, and C. Yang, "Energy-efficient coordinated beamforming under minimal data rate constraint of each user," *IEEE Trans. Veh. Technol.*, vol. 64, no. 6, pp. 2387–2397, Jun. 2015.

[19] 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.

[20] O. Tervo, L. N. Tran, and M. Juntti, "Decentralized coordinated beamforming for weighted sum energy efficiency maximization in multi-cell MISO downlink," in *Proc. IEEE Global Conf. Signal Inform. Process.*, Dec. 2015, pp. 1387–1391.

[21] Z. Zhang, V. Anantharam, M. J. Wainwright, and B. Nikolic, "An efficient 10GBASE-T ethernet LDPC decoder design with low error floors," *IEEE J. Solid-State Circuits*, vol. 45, no. 4, pp. 843–855, Apr. 2010.

[22] S. Ranpara and D. S. Ha, "A low-power Viterbi decoder design for wireless communications applications," in *Proc. IEEE Int. ASIC/SOC Conference*, 1999, pp. 377–381.

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

[24] T. Wang and L. Vandendorpe, "On the optimum energy efficiency for flat-fading channels with rate-dependent circuit power," *IEEE Trans. Commun.*, vol. 61, no. 12, pp. 4910–4921, Dec. 2013.

[25] Z. Wang, I. Stupia, and L. Vandendorpe, "Energy efficient precoder design for MIMO-OFDM with rate-dependent circuit power," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2015, pp. 1897–1902.

[26] A. Arafa and S. Ulukus, "Optimal policies for wireless networks with energy harvesting transmitters and receivers: Effects of decoding costs," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 12, pp. 2611–2625, Dec 2015.

[27] K. Tutuncuoglu and A. Yener, "Communicating with energy harvesting transmitters and receivers," in *Proc. Inf. Theory App. Workshop*, Feb 2012, pp. 240–245.

[28] H. Mahdavi-Doost and R. D. Yates, "Energy harvesting receivers: Finite battery capacity," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul 2013, pp. 1799–1803.

[29] P. Grover, K. Woyach, and A. Sahai, "Towards a communication-theoretic understanding of system-level power consumption," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 8, pp. 1744–1755, Sep 2011.

[30] J. Rubio, A. Pascual-Iserte, and M. Payaro, "Energy-efficient resource allocation techniques for battery management with energy harvesting nodes: a practical approach," in *Proc. European Wireless Conf.*, Apr 2013, pp. 1–6.

[31] P. Rost and G. Fettweis, "On the transmission-computation-energy tradeoff in wireless and fixed networks," in *Proc. IEEE Globecom Workshops*, Dec 2010, pp. 1394–1399.

[32] K. Ganesan, P. Grover, and J. Rabaey, "The power cost of over-designing codes," in *Proc. IEEE Workshop Signal Process. Systems*, Oct 2011, pp. 128–133.

[33] A. Beck, A. Ben-Tal, and L. Tetruashvili, "A sequential parametric convex approximation method with applications to nonconvex truss topology design problems," *Journal of Global Optimization*, vol. 47, no. 1, pp. 29–51, 2010.

[34] G. Venkatraman, A. Tölli, M. Juntti, and L. N. Tran, "Traffic aware resource allocation schemes for multi-cell MIMO-OFDM systems," *IEEE Trans. Signal Process.*, vol. 64, no. 11, pp. 2730–2745, Jun. 2016.

[35] J. Kaleva, A. Tölli, and M. Juntti, "Decentralized sum rate maximization with QoS constraints for interfering broadcast channel via successive convex approximation," *IEEE Trans. Signal Process.*, vol. 64, no. 11, pp. 2788–2802, Jun. 2016.

[36] S. Schaible, "Fractional Programming. I, Duality," *Management Science*, vol. 22, no. 8, pp. 858–867, 1976.

[37] O. Tervo, A. Tölli, M. Juntti, and L. N. Tran, "Energy-efficient coordinated beamforming with rate dependent processing power," in *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun.*, Jul. 2016, pp. 1–5.

[38] A. Balatsoukas-Stimming, P. Giard, and A. Burg, "Comparison of polar decoders with existing low-density parity-check and turbo decoders," in *Proc. IEEE Wireless Commun. Netw. Conf. Workshops*, March 2017, pp. 1–6.

[39] E. Dahlman, S. Parkvall, and J. Sköld, in *4G LTE-Advanced Pro and The Road to 5G (Third Edition)*. Academic Press, 2016.

[40] W. Dinkelbach, "On nonlinear fractional programming," *Management Science*, vol. 13, no. 7, pp. 492–498, 1967.

[41] B. R. Marks and G. P. Wright, "A general inner approximation algorithm for nonconvex mathematical programs," *Operations Research*, vol. 26, no. 4, pp. 681–683, 1977.

[42] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[43] F. Alizadeh and D. Goldfarb, "Second-order cone programming," *Mathematical Programming*, vol. 95, pp. 3–51, 2001.

[44] Y.-C. Jong, "An efficient global optimization algorithm for nonlinear sum-of-ratios problem." [Online]. Available: http://www.optimization-online.org/DB_HTML/2012/08/3586.html

[45] P. Jayasinghe, A. Tölli, and M. Latva-aho, "Bi-directional signaling strategies for dynamic TDD networks," in *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun.*, Jun. 2015, pp. 540–544.

[46] A. Ben-Tal and A. S. Nemirovskiaei, *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2001.

[47] S. Boyd and L. Vandenberghe, "Numerical linear algebra background." [Online]. Available: http://www.seas.ucla.edu/~vandenbe/ee236b/lectures/num-lin-alg.pdf

[48] ETSI Technical Report 136.931, "Radio Frequency (RF) requirements for LTE Pico Node B," Oct. 2012. [Online]. Available: http://www.etsi.org/deliver/etsi_tr/136900_136999/136931/11.00.00_60\/tr_136931v110000p.pdf

[49] "Energy efficiency analysis of the reference systems, areas of improvements and target breakdown," EARTH project, Deliverable D2.3 v2. [Online]. Available: https://bscw.ict-earth.eu/pub/bscw.cgi/d71252/EARTH_WP2_D2.3_v2.pdf

[50] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug 2011.

[51] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 264–273, Feb 2013.

[52] F. Fernandes, A. Ashikhmin, and T. L. Marzetta, "Inter-cell interference in noncooperative TDD large scale antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 192–201, Feb 2013.

[53] L. You, X. Gao, X. G. Xia, N. Ma, and Y. Peng, "Pilot reuse for massive MIMO transmission over spatially correlated rayleigh fading channels," *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3352–3366, Jun 2015.

[54] N. Parikh and S. Boyd, "Proximal algorithms," *Found. Trends Optim.*, vol. 1, no. 3, pp. 127–239, Jan. 2014.