REVIEW

Energy-efficient transmission strategies for CoMP downlink—overview, extension, and numerical comparison

Kien-Giang Nguyen[1*] [ID][†], Oskari Tervo[1†], Quang-Doanh Vu[1], Le-Nam Tran[2] and Markku Juntti[1]

Abstract

This paper focuses on energy-efficient coordinated multi-point (CoMP) downlink in multi-antenna multi-cell wireless communications systems. We provide an overview of transmit beamforming designs for various energy efficiency (EE) metrics including maximizing the overall network EE, sum weighted EE, and fairness EE. Generally, an EE optimization problem is a nonconvex program for which finding the globally optimal solutions requires high computational effort. Consequently, several low-complexity suboptimal approaches have been proposed. Here, we sum up the main concepts of the recently proposed algorithms based on the state-of-the-art successive convex approximation (SCA) framework. Moreover, we discuss the application to the newly posted EE problems including new EE metrics and power consumption models. Furthermore, distributed implementation developed based on alternating direction method of multipliers (ADMM) for the provided solutions is also discussed. For the sake of completeness, we provide numerical comparison of the SCA based approaches and the conventional solutions developed based on parametric transformations (PTs). We also demonstrate the differences and roles of different EE objectives and power consumption models.

Keywords: Energy efficiency, Generalized Dinkelbach's algorithm, Successive convex approximation, Fractional programming, Power consumption, Coordinated beamforming

1 Introduction

Fifth generation (5G) wireless network visions foresee the challenges of the data traffic demand caused by the upcoming explosive growth of wireless devices and applications [1]. The rapid expansion of mobile networks is increasing the energy consumption beyond sustainable limits. In the larger base stations (BSs), the most power-hungry components of the multi-antenna transmitters are the transmit power amplifiers (PAs), but the other circuits and components are also significant power consumers. In fact, they become even dominant in the smaller BSs, which are becoming more and more popular in the future dense networks. Nevertheless, this causes problems in terms of electricity costs for operators and the increase in greenhouse gas emission for the whole world [2, 3].

Consequently, energy efficiency (EE) has become an important design target for wireless access networks.

In wireless communications, energy efficiency is generally defined as the ratio of the total reliably transmitted data to the total energy consumption [4]. In other words, it equals the achievable data rate in bits per second divided by the consumed power in Watts. In either case, the basic unit of EE is bits per Joule (bits/J). It is worth mentioning that the classical transceiver optimization framework, on the other hand, typically focuses on maximizing the multi-user weighted sum rate or (area) spectral efficiency regardless of the proportionally rapid increase of total power consumed by the wireless network. The EE optimization deviates from this setup by making a controlled trade-off between the supported rate and the consumed power [2, 3, 5–7].

Variations of the EE objective have been proposed depending on the system constraints and design targets. The basic alternatives include *network EE* (NEE), *sum weighted EE* (SWEE), and *fairness EE* [5]. While the first

*Correspondence: giang.nguyen@oulu.fi
†Kien-Giang Nguyen and Oskari Tervo contributed equally to this work
[1]Centre for Wireless Communications, University of Oulu, P.O. Box 4500, FI-90014 Oulu, Finland
Full list of author information is available at the end of the article

metric optimizes the EE gain of the entire network, the others aim at satisfying the specific EE requirements on individual base stations or users involved.

In the presence of multi-user interference, an EE maximization (EEmax) problem belongs to a class of non-convex fractional programs for which finding a globally optimal solution is challenging. However, an optimal solution of the EEmax problem in multi-user multiple-input single-output (MISO) downlink system has been provided in [7] using a branch-reduce-and-bound approach. Even though this approach guarantees finding the global optimum, it still requires very high computational complexity. Therefore, low-complexity suboptimal designs have attracted more attention for practical applications.

Common suboptimal approaches for EE designs have been developed based on parametric transformation (PT) inspired by the fractional structure of the EE objectives [5, 8, 9]. However, such an approach leads to two-layer iterative procedures [9], which often have high computational complexity (as discussed in Section 3.1) and/or are not suitable for distributed implementation. In addition, analyzing the convergence of those methods has not been properly addressed [7].

Recently, novel algorithms have been developed based on the state-of-the-art local optimization toolbox, namely successive convex approximation (SCA) algorithm, which efficiently solves the EEmax problems; the proposed framework is a one-loop iterative procedure which finds out locally optimal solutions after a relatively small number of iterations and, thus, significantly reduces the complexity compared to the existing PT approach [10]; the convergence of the SCA-based methods is provably guaranteed [7, 10], and the procedure is also well suited for the implementation in a distributed manner [11].

In this paper, we consider coordinated multi-point (CoMP) downlink in multi-antenna multi-cell systems and focus on the applications of the SCA approach on the EEmax problems arising in the wireless access systems such as 4G and 5G cellular standards. The main contributions of this paper can be summarized as follows:

- Overview: We provide a summary of the basic concepts of the SCA-based algorithms; introduce some key transformations which turn the EEmax problems into representations that successfully leverage the principle of the SCA; revisit the problems of maximizing the NEE, SWEE, and maxminEE; and discuss how to arrive at efficient solutions. We also discuss how to distributively implement the solutions.
- Extension: We discuss the recently proposed weighted product EE (WPEE) objective function and a general model of power consumption. We show how to adopt the proposed framework to the EEmax problems involved.

- Numerical comparisons: We make several numerical comparisons on the algorithms. The most important one is the comparison between the existing and the proposed approaches in terms of convergence speed and average performances. Other evaluations have been made to illustrate the roles and benefits of different EE objectives and the impact of different power consumption models on the EE performance.

An initial version of the paper was published in [12]. Herein, we provide a more detailed and broader summary of the EE optimization and discussion on the differences of the SCA- and fractional programming-based approaches. We also extend the SCA framework to solve the problem of WPEE maximization. We further present four different approximations for the involved logarithmic functions, which enable the second-order programming formulations of the problems. Finally, we consider more detailed power consumption models and provide a significantly more extensive set of simulation results to evaluate different methods.

The rest of the paper is organized as follows. System model and several energy efficiency measures are presented in Section 2. Centralized solutions and their distributed implementation are provided in Section 3, followed by numerical results in Section 4. Conclusion is provided in Section 5.

Notation: Bold lower and upper case letters represent vectors and matrices, respectively; calligraphic letters denote sets; $|\cdot|$ represents the absolute value; $\|\cdot\|_2$ represents the l_2 norm; $\mathcal{CN}(0, a)$ denotes a zero-mean circularly symmetric complex Gaussian random variable with variance a; $\mathbb{C}^{a \times b}$ represents the space of complex matrices of dimensions given in superscript; $\Re(\cdot)$ represents real part of the argument; $\mathbb{E}\{\cdot\}$ denotes the expectation operator. $\mathbf{a}^T$ and $\mathbf{a}^H$ stand for the transpose and the Hermitian transpose of $\mathbf{a}$, respectively. $\langle \mathbf{a}, \mathbf{b} \rangle$ denotes the inner product of vectors $\mathbf{a}$ and $\mathbf{b}$. $\{\mathbf{a}_b\}_{b \in \mathcal{B}}$ refers to a composite vector containing all $\mathbf{a}_b$ where b belongs to the set $\mathcal{B}$. $\nabla_{\mathbf{x}} g(\mathbf{x})$ represents the partial derivative of function $g(\mathbf{x})$ with respect to the elements of $\mathbf{x}$. Other notations are defined at their first appearance.

2 System model and energy-efficient problem formulations

2.1 Channel and signal model

We consider a downlink transmission in multi-cell multi-user multiple-input single-output (MISO) system consisting of B BSs, each of which is equipped with M antennas. There are U single-antenna users in each cell and a total of UB users in the network[1]. We assume that the BSs operate following the coordinated beamforming mode, i.e., each BS only serves U users in its own cell.[2] The considered system model is illustrated in Fig. 1. The beamforming

Fig. 1 CoMP system model
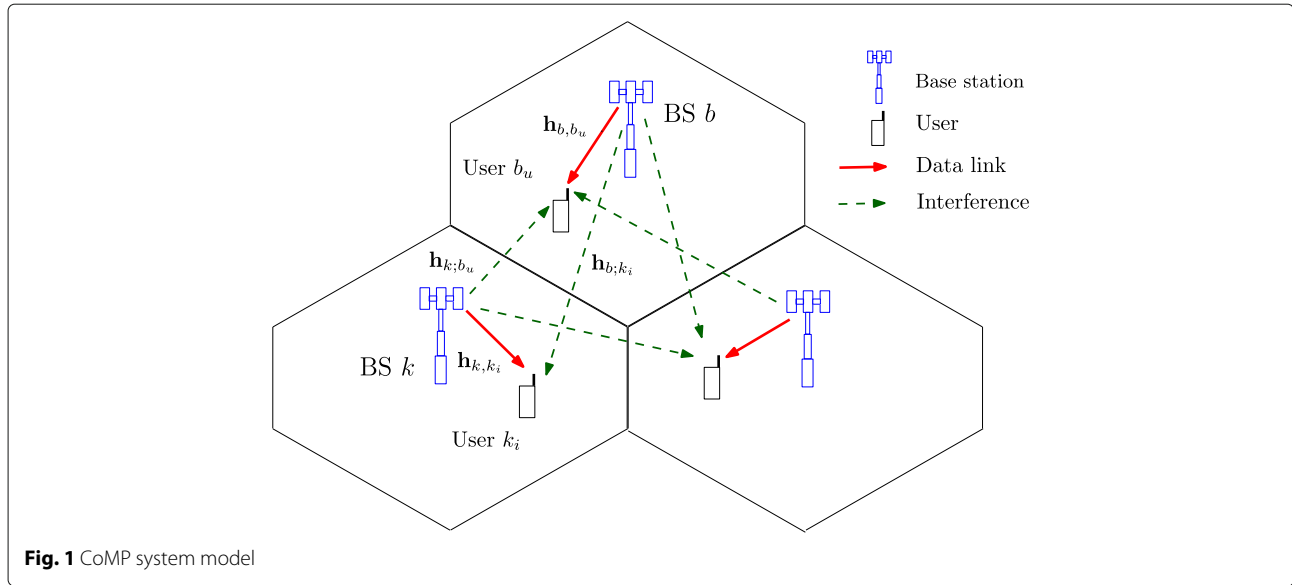
vectors are designed to control the interference between the cells so as to maximize a performance target [13]. Let us denote the set of BSs by $\mathcal{B} = \{1, \ldots, B\}$ and the set of users in cell b by $\mathcal{U}_b = \{1, \ldots, U\}$. User u in cell b is denoted compactly as b_u. Let s_{b_u} be an independent data symbol for user b_u which is assumed to have a unit energy, i.e., $\mathbb{E}\{|s_{b_u}|^2\} = 1$. Linear transmit precoding is adopted such that the signal transmitted to user b_u (from BS b) is a multiplication of s_{b_u} and transmit beamforming vector $\mathbf{v}_{b_u} \in \mathbb{C}^{M \times 1}$. Let $\mathbf{h}_{k,b_u} \in \mathbb{C}^{1 \times M}$ denote the flat-fading channel (row) vector between BS k and user b_u. The received signal at user b_u can be written as

$$y_{b_u} = \mathbf{h}_{b,b_u}\mathbf{v}_{b_u}s_{b_u} + \sum_{i \in \mathcal{U}_b \setminus \{b_u\}} \mathbf{h}_{b,b_u}\mathbf{v}_{b_i}s_{b_i} + \sum_{k \in \mathcal{B} \setminus \{b\}} \sum_{i \in \mathcal{U}_k} \mathbf{h}_{k,b_u}\mathbf{v}_{k_i}s_{k_i} + z_{b_u},$$
(1)

where z_{b_u} is the additive white Gaussian noise with distribution $z_{b_u} \sim \mathcal{CN}(0, \sigma_{b_u}^2)$, $\sigma_{b_u}^2 = WN_0$ is the noise power when using the transmission bandwidth W, and the noise power spectral density is N_0. In (1), the second and third terms represent the intra-cell and inter-cell interference, respectively. Let us denote by $G_{b_u}(\mathbf{v}) \triangleq \sum_{i \in \mathcal{U}_b \setminus \{b_u\}} |\mathbf{h}_{b,b_u}\mathbf{v}_{b_i}|^2 + \sum_{k \in \mathcal{B} \setminus \{b\}} \sum_{i \in \mathcal{U}_k} |\mathbf{h}_{k,b_u}\mathbf{v}_{k_i}|^2$ the power of interference at user b_u. As is common in the system optimization, we use the information theoretic rate expressions of the Gaussian channels. Those assume the use of Gaussian codebooks. Therefore, the multi-user interference terms can be modeled as additive colored Gaussian noise, and the signal-to-interference-plus-noise ratio (SINR) at user b_u is expressed as

$$\Gamma_{b_u}(\mathbf{v}) \triangleq \frac{|\mathbf{h}_{b,b_u}\mathbf{v}_{b_u}|^2}{G_{b_u}(\mathbf{v}) + \sigma_{b_u}^2}.$$
(2)

The data rate of user b_u is given by $r_{b_u}(\mathbf{v}) = W \log(1 + \Gamma_{b_u}(\mathbf{v}))$, and the total data rate over the network is given by

$$R(\mathbf{v}) \triangleq \sum_{b \in \mathcal{B}} \sum_{u \in \mathcal{U}_b} r_{b_u}(\mathbf{v}).$$
(3)

2.1.1 Transmit power constraints
Since the available power budget at the BSs is finite, the transmit power at each BS should satisfy

$$\sum_{u \in \mathcal{U}_b} \|\mathbf{v}_{b_u}\|_2^2 \le P_b, \forall b \in \mathcal{B},$$
(4)

where P_b is the transmit power budget at BS b. In practice, the power amplifier at each antenna chain is designed to operate over a specific power range, i.e, the output power should not exceed a predefined threshold. Thus, the power constraint for each antenna can be also imposed, i.e.,

$$\sum_{u \in \mathcal{U}_b} |[\mathbf{v}_{b_u}]_m|^2 \le P_b^m, \forall b \in \mathcal{B}, \ m = 1, 2, \ldots, M,$$
(5)

where $[\mathbf{x}]_m$ denotes the mth element of vector $\mathbf{x}$, and P_b^m is the maximum transmit power at the mth antenna of BS b. Several other power constraints could be applied, but we focus on these most common ones.

2.2 Power consumption model
The consumed power can be classified into three main categories: circuit operation power in network elements, signal processing power, and power dissipated on power amplifiers (PAs). Some of the power components are static (sta), while others are dynamic (dyn) or rate-dependent (RD). The power consumption model is sketched in Fig. 2.
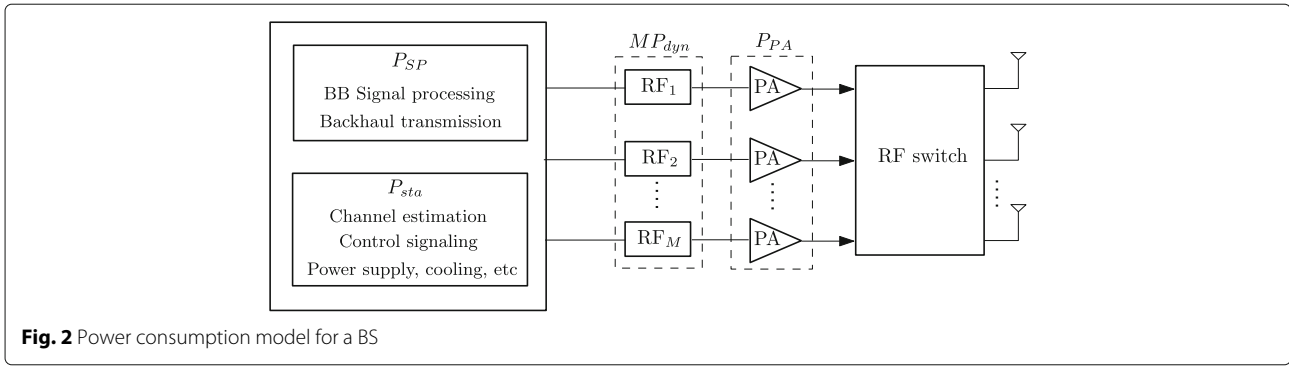
Nguyen *et al. EURASIP Journal on Wireless Communications and Networking* (2018) 2018:207

Page 4 of 18

Fig. 2 Power consumption model for a BS

2.2.1 Circuit power

A significant amount of power is used to operate the electronic circuits of network elements such as the base stations, site cooling, control signaling, backhaul infrastructure, and some parts of the baseband processors. In addition, a radio frequency (RF) chain including, for example, converters, filters, and mixers requires some operating power. In general, we can express the amount of power consumption for operating transceiver circuits in cell b as [5, 14, 15]

$$P_{\text{cir},b} \triangleq P_{\text{sta}} + MP_{\text{dyn}} + UP_{\text{Us}}, \tag{6}$$

where P_{sta} and P_{dyn} represent for static and dynamic power consumption at BS b, respectively, and P_{Us} accounts for power running a user device.

2.2.2 Signal processing power

The data needs to be encoded and modulated at the transmitter as well as demodulated and decoded at the receiver. Conventionally, the amount of power for these functionalities is assumed to be fixed [5, 8, 9, 16, 17]. However, generally, a higher data rate requires a larger codebook, and the larger number of bits incurs higher power for encoding and decoding on baseband circuit boards. Moreover, the backhaul is used to transmit data between the core network and the BSs, and the power consumed for the backhaul also increases with the data rate [18–20]. From this perspective, signal processing power consumption is rate-dependent and is assumed to be a linear function of the transmission rate [18]. Let us denote by $P_{\text{SP},b}(r_b(\mathbf{v}))$, where $r_b(\mathbf{v}) \triangleq \sum_{u \in \mathcal{U}_b} r_{b_u}(\mathbf{v})$ is the signal processing power for BS b. Then, we can write

$$P_{\text{SP},b}(r_b(\mathbf{v})) \triangleq \begin{cases} \text{constant} & \text{for fixed power model} \\ p_{\text{SP}} \sum_{u \in \mathcal{U}_b} r_{b_u}(\mathbf{v}) & \text{for rate-dependent power model} \end{cases}, \tag{7}$$

where p_{SP} is a constant coefficient with unit W/(Gbits/s).

2.2.3 Power dissipated on PAs

The amount of power consumed by the PAs strongly depends on the power amplifier's efficiency. Conventionally, the efficiency of a PA is assumed to be a constant over operating range [5, 8, 9, 16, 17]. This assumption leads to the model

$$P_{\text{PA},b}(\mathbf{v}) \triangleq \frac{1}{\epsilon} \sum_{u \in \mathcal{U}_b} \|\mathbf{v}_{b_u}\|_2^2, \tag{8}$$

where $P_{\text{PA},b}$ denotes the PAs' dissipated power at BS b, and $\epsilon \in (0, 1)$ is a constant standing for the PA efficiency. However, in practice, PA efficiency is highly dependent on the output power region and the employed PA type. To account this, the nonlinear power consumption models of PAs have been introduced [21–24] in which the PA efficiency of RF chain m at BS b is expressed as

$$\epsilon_{b,m}(\{\mathbf{v}_{b_u}\}_u) \triangleq \tilde{\epsilon} \sqrt{\sum_{u \in \mathcal{U}_b} |[\mathbf{v}_{b_u}]_m|^2}, \tag{9}$$

where $\tilde{\epsilon} = \epsilon_{\max}/\sqrt{P_b^m}$, and $\epsilon_{\max} \in (0, 1)$ is the maximum PA's efficiency. We note that P_b^m and $\epsilon_{\max}$ depend on the employed PA techniques. For notational simplicity, we assume that $\tilde{\epsilon}$ is the same for all b and m. From (9), the total power consumption on the PAs at BS b can be written as

$$P_{\text{PA},b}(\mathbf{v}) \triangleq \sum_{m=1}^{M} \frac{\sum_{u \in \mathcal{U}_b} |[\mathbf{v}_{b_u}]_m|^2}{\epsilon_{b,m}(\{\mathbf{v}_{b_u}\}_u)} = \frac{1}{\tilde{\epsilon}} \sum_{m=1}^{M} \sqrt{\sum_{u \in \mathcal{U}_b} |[\mathbf{v}_{b_u}]_m|^2}. \tag{10}$$

2.2.4 General power consumption models

Based on the above discussion, the total power consumption model in cell b can be collectively written as

$$P_{\text{BS},b}(\mathbf{v}) \triangleq P_{\text{cir},b} + P_{\text{SP},b}(r_b(\mathbf{v})) + P_{\text{PA},b}(\mathbf{v}). \tag{11}$$

Hence, the total network power consumption is

$$P_{\text{total}}(\mathbf{v}) \triangleq \sum_{b \in \mathcal{B}} \left(P_{\text{cir},b} + P_{\text{SP},b}(r_b(\mathbf{v})) + P_{\text{PA},b}(\mathbf{v}) \right). \tag{12}$$

On the other hand, the power for the data transmission to a user is a favorable measure in some user-centric applications. Let

$$P_{\mathrm{SP},b}(r_{b_u}(\mathbf{v})) \triangleq \begin{cases} P_{\mathrm{SP},b}(r_b(\mathbf{v}))/U & \text{for fixed power model} \\ p_{\mathrm{SP}}r_{b_u}(\mathbf{v}) & \text{for rate-dependent power model} \end{cases}$$
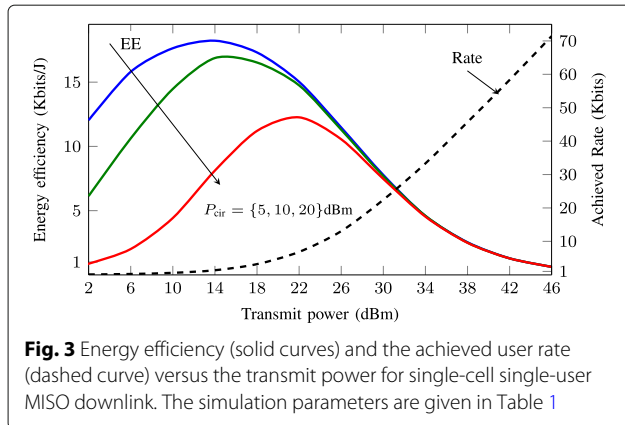
denote the signal processing power corresponding to user b_u. Then, the amount of consumed power corresponding to the data transmission to user b_u can be written as [17]

$$P_{\mathrm{Us},b_u}(\mathbf{v}) \triangleq \sum_{m=1}^{M} \frac{|[\mathbf{v}_{b_u}]_m|^2}{\tilde{\epsilon}\sqrt{\sum_{u \in \mathcal{U}_b}|[\mathbf{v}_{b_u}]_m|^2}}$$
$$+ P_{\mathrm{SP},b}(r_{b_u}(\mathbf{v})) + \frac{P_{\mathrm{sta}} + MP_{\mathrm{dyn}}}{U} + P_{\mathrm{Us}}$$

(13)

in which all users in a cell are assumed to be evenly responsible for the operating power of their serving BS.

2.3 Energy-efficiency metrics

The EE measures the number of bits reliably transmitted by a unit energy. In other words, it can be defined as the ratio of the achievable data rate to the total power consumption. The ratio quantifies the trade-off between the network throughput and the power consumption. This is illustrated via a simple single-cell single-user MISO downlink example in Fig. 3. The energy efficiency and the achieved rate are plotted versus the transmit power. We observe that, for all cases of the operating circuit power, when the transmit power increases, the EE first increases, reaching a maximum, and then decreases. In other words, when the circuit power plays a non-negligible role and the rate is penalized by the overall power consumption, the optimum performance is not achieved by using all available power budget. This observation gives rise to the systematic development of the optimization algorithms as detailed below, where four different widely considered EE metrics are introduced and discussed.

Fig. 3 Energy efficiency (solid curves) and the achieved user rate (dashed curve) versus the transmit power for single-cell single-user MISO downlink. The simulation parameters are given in Table 1

2.3.1 Network energy efficiency

The NEE metric quantifies the EE performance of the entire network [5, 25]. It is defined as

$$\mathrm{NEE}(\mathbf{v}) \triangleq \frac{R(\mathbf{v})}{P_{\mathrm{total}}(\mathbf{v})} \tag{14}$$

We remark that in scenarios where cellular BSs with different features and specifications co-exist, e.g., heterogeneous network, NEE might lack relevance, since neither EE requirement for each cell/user nor the fairness among all parties of the network can be guaranteed.

2.3.2 Sum weighted energy efficiency

The SWEE metric can satisfy the specific demand on EE of each network node. For the considered system model, SWEE can be expressed as

$$\mathrm{SWEE}(\mathbf{v}) \triangleq \begin{cases} \sum_{b \in \mathcal{B}} \omega_b \dfrac{\sum_{u \in \mathcal{U}_b} r_{b_u}(\mathbf{v})}{P_{\mathrm{BS},b}(\mathbf{v})} & \text{for the BS-centric network} \\ \sum_{b \in \mathcal{B}} \sum_{u \in \mathcal{U}_b} \omega_{b_u} \dfrac{r_{b_u}(\mathbf{v})}{P_{\mathrm{Us},b_u}(\mathbf{v})} & \text{for the user-centric network} \end{cases}$$

(15)

where $\omega_b \in (0,1]$ and $\omega_{b_u} \in (0,1]$ are parameters representing the priority for cell b and user b_u, respectively.

2.3.3 Weighted product EE

The WPEE metric also takes into account the individual demand on EE of each node which is defined as [5, 26]

$$\mathrm{WPEE}(\mathbf{v}) \triangleq \begin{cases} \prod_{b \in \mathcal{B}} \left(\dfrac{\sum_{u \in \mathcal{U}_b} r_{b_u}(\mathbf{v})}{P_{\mathrm{BS},b}(\mathbf{v})} \right)^{\omega_b} & \text{for the BS-centric network} \\ \prod_{b \in \mathcal{B}} \prod_{u \in \mathcal{U}_b} \left(\dfrac{r_{b_u}(\mathbf{v})}{P_{\mathrm{Us},b_u}(\mathbf{v})} \right)^{\omega_{b_u}} & \text{for the user}-\text{centric network} \end{cases}$$

(16)

It is worth noting that although the WPEE metric does not give the same EE-unit (bits/J) as such, it has been used in the literature to achieve fairness in EE. Specifically, it is not difficult to see that none of the BSs experiences EE close to zero when WPEE is considered.

2.3.4 Max-min fairness energy efficiency

The max-min fairness EE metric provides the best fairness for the considered nodes compared to others. This metric is preferable to the scenarios where EE is critical for each cell, e.g., in cellular networks where BSs are not connected to a fixed electricity grid. The definition of the metric is given as [10]

$$\mathrm{minEE}(\mathbf{v}) \triangleq \begin{cases} \min_{b \in \mathcal{B}} \dfrac{\sum_{u \in \mathcal{U}_b} r_{b_u}(\mathbf{v})}{P_{\mathrm{BS},b}(\mathbf{v})} & \text{for the BS-centric network} \\ \min_{\substack{b \in \mathcal{B} \\ u \in \mathcal{U}_b}} \dfrac{r_{b_u}(\mathbf{v})}{P_{\mathrm{Us},b_u}(\mathbf{v})} & \text{for the user}-\text{centric network} \end{cases}$$

(17)

Nguyen *et al. EURASIP Journal on Wireless Communications and Networking* (2018) 2018:207

Page 6 of 18

2.4 Energy efficiency optimization problems

From the above discussions, the problems of beam-forming design for EE maximization can be generally written as

$$\underset{\mathbf{v}}{\text{maximize}}\ f_{\text{EE}}(\mathbf{v}) \quad \text{subject to } \{(4), (5)\} \tag{18}$$

where the objective function $f_{\text{EE}}(\mathbf{v})$ represents one of the NEE($\mathbf{v}$), SWEE($\mathbf{v}$), WPEE($\mathbf{v}$), and minEE($\mathbf{v}$).

In general, (18) is a nontractable fractional program.[3] In the next section, we briefly review the conventional approaches, which suboptimally solve the EEmax problems, and then provide the recently proposed SCA framework which improves efficiently the solution quality.

3 Centralized methods for energy-efficient transmissions: review and extension

3.1 Conventional fractional programming approaches

Most of the existing solutions for the EEmax problems are based on conventional fractional programming methods, i.e., parameterized approaches [5, 8, 9] or the parameter-free approach based on the Charnes-Cooper transformation. We briefly sketch the idea of these approaches for solving fractional programs below.

In general, a fractional program is expressed as

$$\underset{\mathbf{x} \in \mathcal{S}}{\text{maximize}}\ \sum_{i=1}^{L} \frac{f_i(\mathbf{x})}{g_i(\mathbf{x})} \tag{19}$$

where $L \geq 1$, $\mathcal{S}$, $f_i(\mathbf{x})$ and $g_i(\mathbf{x})$ are a convex set, concave and convex functions respect to variable vector $\mathbf{x} \in \mathbb{C}^N$, respectively.

3.1.1 Single-ratio fractional programs

When $L = 1$, the problem can be transformed into a parameterized form. That is, one can consider the following problem with parameter ω: $H(\omega) = \max_{\mathbf{x} \in \mathcal{S}}\{f_1(\mathbf{x}) - \omega g_1(\mathbf{x})\}$. Due to the fact that $H(\omega)$ is continuous and

strictly monotonically decreasing [27], $H(\omega) = 0$ has a unique solution ω^*. The optimal solution to the original fractional program is $\mathbf{x}^* = \arg\max_{\mathbf{x} \in \mathcal{S}}\{f_1(\mathbf{x}) - \omega^* g_1(\mathbf{x})\}$. Thereby, the problem can be solved by finding ω such that $H(\omega) = 0$. A parametric approach exploits the Newton method to find root of $H(\omega)$ (often called as the Dinkelbach method or the Newton-Rhapson method). The method first initializes $\omega^{(0)} = \frac{f_1(\mathbf{x}^{(0)})}{g_1(\mathbf{x}^{(0)})}$. Subsequently, the problem $H(\omega^{(0)})$ is solved, the solution of which is then used to update $\omega^{(n)} = \frac{f_1(\mathbf{x}^{(n-1)})}{g_1(\mathbf{x}^{(n-1)})}$, and this procedure is repeated until convergence. Besides the well-known Dinkelbach method, the problem can be also solved as a single convex program using the Charnes-Cooper transformation [28].

3.1.2 Multi-ratio fractional programs

When $L > 1$, (19) is a sum-of-ratios fractional program. A conventional heuristic strategy for solving this type of problems 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 [9, 29, 30]. Specifically, the solutions for (19) can be obtained by solving as a series of subproblems $H(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \max_{\mathbf{x} \in \mathcal{S}} \sum_{i=1}^{L} \alpha_i(f_i(\mathbf{x}) - \beta_i g_i(\mathbf{x}))$, where $\{\alpha_i\}_i$ and $\{\beta_i\}_i$ are parameters. Similarly to the single-ratio case, $\{\alpha_i\}_i$ and $\{\beta_i\}_i$ are first fixed and the subproblem is solved for given parameters. Then, $\{\alpha_i\}_i$ and $\{\beta_i\}_i$ are updated according to a damped Newton method.

Nevertheless, the advantages of the parametric approaches are hardly recognized when they are applied to wireless communications problems because $f_i(\mathbf{x})$ and $g_i(\mathbf{x})$ are often nonconvex. Implicitly, the parametric subproblem is nonconvex, and its optimal solutions are difficult to find. To cope with this, the SCA or alternating optimization method based on iterative weighted minimum mean square error (WMMSE) approach is often combined with the parametric method leading to multi-level iterative algorithms. Thus, these algorithms need a very high number of iterations to converge. Moreover, likely local optimality for each parametric problem is achieved which means that parametric approaches may not always guarantee the convergence.

To avoid the multi-level iterative procedure, we present below the framework developed recently based on the SCA method. The algorithms derived from the approach are provably and fast convergent; thus, they overcome the issues raised by the earlier solutions.

3.2 SCA principle

We first briefly review the SCA principles before presenting their applications to the EEmax problems. The central idea of the SCA method is to iteratively approximate the nonconvex constraints of an optimization problem by

Table 1 Simulation parameters

Parameters	Value
Path loss and shadowing	$38 \log_{10}(d[\text{m}]) + 34.5 + \mathcal{N}(0, 8)$
Inter-BS distance	$D = 1$ km
Static power consumption P_{sta}	33 dBm
Static power consumption P_{dyn}	30 dBm
Power amplifier efficiency ϵ	0.35
Number of BSs B	3
Number of users per cell U_b	2
Number of Tx antennas M	4
Signal bandwidth W	10 kHz
Power spectral density of noise	-174 dBm/Hz

Nguyen *et al. EURASIP Journal on Wireless Communications and Networking* (2018) 2018:207

Page 7 of 18

Algorithm 1 SCA procedure solving (20)

Initialization: Set $n := 0$, choose an initial feasible point $\mathbf{x}^{(n)}$.
repeat
  Solve (21) and obtain optimal value $\mathbf{x}^*$
  Update $\mathbf{x}^{(n+1)} := \mathbf{x}^*$
  Update $n := n + 1$
until Convergence
Output: $\mathbf{x}^{(n)}$

---

proper convex ones [31]. In particular, let us consider a general optimization program given by

$$\underset{\mathbf{x} \in \mathcal{S}}{\text{minimize}}\ f(\mathbf{x}) \quad \text{subject to} \quad \{g_i(\mathbf{x}) \leq 0,\ i = 1, \ldots, L\} \tag{20}$$

where $f(\mathbf{x})$ is convex and $\{g_i(\mathbf{x})\}_i$ are nonconvex functions in a convex set $\mathcal{S}$ w.r.t variable vector $\mathbf{x}$. At iteration n, given a feasible point $\mathbf{x}^{(n)}$, function $g_i(\mathbf{x})$ is approximated by its convex approximation function $\hat{g}_i(\mathbf{x}, \mathbf{x}^{(n)})$ for all i such that

(a) $g_i(\mathbf{x}) \leq \hat{g}_i\left(\mathbf{x}, \mathbf{x}^{(n)}\right)$
(b) $g_i(\mathbf{x}^{(n)}) = \hat{g}_i\left(\mathbf{x}^{(n)}, \mathbf{x}^{(n)}\right)$
(c) $\nabla_{\mathbf{x}} g_i(\mathbf{x}^{(n)}) = \nabla_{\mathbf{x}} \hat{g}_i\left(\mathbf{x}^{(n)}, \mathbf{x}^{(n)}\right)$

for all $\mathbf{x} \in \tilde{\mathcal{S}} \triangleq \{\mathbf{x} \in \mathcal{S} | g_i(\mathbf{x}) \leq 0,\ i = 1, \ldots, L\}$. Properties (a) and (b) are to guarantee the monotonic (objective) convergence behavior for the SCA algorithm; properties (b) and (c) guarantee that the Karush-Kuhn-Tucker (KKT) optimality conditions are satisfied by convergent points [31]. By the replacement, we arrive at the following convex subproblem

$$\underset{\mathbf{x} \in \mathcal{S}}{\text{minimize}}\ f(\mathbf{x}) \quad \text{subject to} \quad \{\hat{g}_i(\mathbf{x}, \mathbf{x}^{(n)}) \leq 0,\ i = 1, \ldots, L\}. \tag{21}$$

The optimal solution $\mathbf{x}^*$ of (21) belongs to the set $\tilde{\mathcal{S}}$ due to (a) and (b). Thus, $\mathbf{x}^*$ is used as the feasible point for the next iteration, i.e., $\mathbf{x}^{(n+1)} = \mathbf{x}^*$. The process is iteratively carried out until convergence is established. The SCA procedure solving (20) is outlined in Algorithm 1. We note that $f(\mathbf{x}^*) \leq f(\mathbf{x}^{(n)})$ for all n, i.e. sequence $\{f(\mathbf{x}^{(n)})\}_n$ decreases monotonically. Thus, $\{f(\mathbf{x}^{(n)})\}_n$ converges if it is bounded below by a finite value in the set $\tilde{\mathcal{S}}$. The following remark shows a well-known method for arriving $\hat{g}_i(\mathbf{x}, \mathbf{x}^{(n)})$, which is widely used in this paper.

Remarks 1 *Let $g(\mathbf{x})$ be a concave function w.r.t $\mathbf{x}$, then its convex upper bound satisfying (a)–(c) can be achieved by the mean of the first-order Taylor approximation as*

$$g(\mathbf{x}) \leq \underbrace{g(\mathbf{x}^{(n)}) + \left\langle \nabla_{\mathbf{x}} g(\mathbf{x}^{(n)}), \mathbf{x} - \mathbf{x}^{(n)} \right\rangle}_{\triangleq \hat{g}(\mathbf{x}; \mathbf{x}^{(n)})}. \tag{22}$$

Example 1 *Consider the quadratic-over-linear function $g(x, y) = \frac{-x^2}{y}, y > 0$, which is concave w.r.t the involved variables. From (22), a convex upper bound of $g(x, y)$ at $(x^{(n)}, y^{(n)}), y^{(n)} > 0$, is written as*

$$\hat{g}(x, y; x^{(n)}, y^{(n)}) = -\left(\frac{\left(x^{(n)}\right)^2}{y^{(n)}} + \frac{2x^{(n)}}{y^{(n)}}\left(x - x^{(n)}\right) - \frac{\left(x^{(n)}\right)^2}{\left(y^{(n)}\right)^2}\left(y - y^{(n)}\right)\right)$$

$$= -\frac{2x^{(n)}}{y^{(n)}}x + \frac{(x^{(n)})^2}{(y^{(n)})^2}y.$$

It can be easily justified that $\hat{g}\left(x, y; x^{(n)}, y^{(n)}\right)$ satisfies properties (a)–(c) for all $(x, y > 0)$.

3.3 SCA-based solutions for EEmax problems

In this section, we present how to adopt the procedure discussed in Section 3.2 to the EEmax problems posted in Section 2.4. It is worth mentioning that directly applying the SCA method to these problems seems challenging, because deriving convex approximations for nonconvex parts in the problems that satisfy conditions (a)–(c) is very difficult. Thus, the necessary step is to transform the EEmax problems into more tractable representations, which preserve the optimality of the original one as well as are amenable to the SCA method.

3.3.1 Network EEmax problem

We first provide the SCA solutions for the problem with network EE metric which contains single-ratio fractional objective. Replacing $f_{\text{EE}}(\mathbf{v})$ in (18) by NEE(**v**), we get the following problem

$$\underset{\mathbf{v}}{\text{maximize}}\ \frac{R(\mathbf{v})}{P_{\text{total}}(\mathbf{v})} \quad \text{subject to} \quad \{(4), (5)\}. \tag{23}$$

For translating (23) to a more tractable form, we exploit the epigraph transformation [32]. Let us introduce new slack variables η, z, t, and $\{g_{b_u}\}_{b_u}$ and rewrite (23) as

$$\underset{\mathbf{v},\eta,z,t,\{g_{b_u}\}}{\text{maximize}} \quad \eta \tag{24a}$$

$$\text{subject to} \quad \eta \leq \frac{z^2}{t} \tag{24b}$$

$$t \geq P_{\text{total}}(\mathbf{v}) \tag{24c}$$

$$z^2 \leq \sum_{b \in \mathcal{B}, u \in \mathcal{U}_b} \log(1 + g_{b_u}) \tag{24d}$$

$$\frac{|\mathbf{h}_{b,b_u}\mathbf{v}_{b_u}|^2}{g_{b_u}} \geq G_{b_u}(\mathbf{v}) + \sigma_{b_u}^2, \ \forall b \in \mathcal{B}, u \in \mathcal{U}_b \tag{24e}$$

$$(4), (5). \tag{24f}$$

The relationship between (23) and (24) is stated in the following lemma.

Lemma 1 *Problems* (23) *and* (24) *are equivalent at optimality.*

The proof of the lemma is given in the Appendix. Let us now apply the SCA method to solve (24). First, we observe that constraints (24b) and (24e) are nonconvex while the others are convex. Second, the nonconvex parts in (24b) and (24e) are in the form of quadratic-over-affine function mentioned in Example 1. Therefore, the valid convex approximations for (24b) and (24e) are given as

$$\frac{2z^{(n)}}{t^{(n)}}z - \frac{(z^{(n)})^2}{(t^{(n)})^2}t \geq \eta \tag{25}$$

$$\frac{2\Re(\tilde{\mathbf{h}}_{b,b_u}^{(n)}\mathbf{v}_{b_u})}{g_{b_u}^{(n)}} - \frac{|\mathbf{h}_{b,b_u}\mathbf{v}_{b_u}^{(n)}|^2 g_{b_u}}{(g_{b_u}^{(n)})^2} \geq G_{b_u}(\mathbf{v}) + \sigma_{b_u}^2 \tag{26}$$

respectively, where $\tilde{\mathbf{h}}_{b,b_u}^{(n)} \triangleq \left(\mathbf{v}_{b_u}^{(n)}\right)^H \mathbf{h}_{b,b_u}^H \mathbf{h}_{b,b_u}$ and $\left(\mathbf{v}^{(n)}, z^{(n)}, t^{(n)}, \{g_{b_u}^{(n)}\}\right)$ is some feasible point of (24). As a result, we arrive at the approximate convex program at iteration n as

$$\underset{\mathbf{v},\eta,z,t,\{g_{b_u}\}}{\text{maximize}} \ \eta \quad \text{subject to} \ \{(4), (5), (24c), (24d), (25), (26)\}. \tag{27}$$

For the rate-dependent signal processing model, due to the following relation $\arg\max_{\mathbf{v}} \frac{R(\mathbf{v})}{P_{\text{total}}(\mathbf{v})} = \arg\min_{\mathbf{v}} \frac{P_{\text{total}}(\mathbf{v})}{R(\mathbf{v})} = \arg\min_{\mathbf{v}} \frac{\sum_{b \in \mathcal{B}}(P_{\text{cir},b} + P_{\text{PA},b}(\mathbf{v}))}{R(\mathbf{v})} + p_{\text{SP}} = \arg\max_{\mathbf{v}} \frac{R(\mathbf{v})}{\sum_{b \in \mathcal{B}}(P_{\text{cir},b} + P_{\text{PA},b}(\mathbf{v}))}$, we can ignore the term of rate-dependent power in the optimization process without loss of optimality. Consequently, the denominator of the objective becomes a convex function w.r.t. $\mathbf{v}$, and thus, the solutions can be obtained following the above discussion.

3.3.2 Sum weighted EEmax problem
We focus on the problem of SWEE maximization from the perspective of the BSs. The SWEE maximization problem from the user perspective is treated similarly. Replacing

$f_{\text{EE}}(\mathbf{v})$ in (18) by SWEE($\mathbf{v}$), we arrive at the problem

$$\underset{\mathbf{v}}{\text{maximize}} \ \sum_{b=1}^{B} \omega_b \frac{\sum_{u \in \mathcal{U}_b} r_{b_u}(\mathbf{v})}{P_{\text{BS},b}(\mathbf{v})} \quad \text{subject to} \ \{(4), (5)\}. \tag{28}$$

As the first step, we introduce new variables $\{\eta_b\}_b$, $\{z_b\}_b$, $\{t_b\}_b$, and $\{g_{b_u}\}_{b_u}$ and write (28) in equivalent form as

$$\underset{\mathbf{v},\{\eta_b\},\{z_b\},\{t_b\},\{g_{b_u}\}}{\text{maximize}} \quad \sum_{b \in \mathcal{B}} \omega_b \eta_b \tag{29a}$$

$$\text{subject to} \quad \eta_b \leq \frac{z_b^2}{t_b}, \ \forall b \in \mathcal{B} \tag{29b}$$

$$t_b \geq P_{\text{BS},b}(\mathbf{v}), \ \forall b \in \mathcal{B} \tag{29c}$$

$$z_b^2 \leq \sum_{u \in \mathcal{U}_b} \log(1 + g_{b_u}), \ \forall b \in \mathcal{B} \tag{29d}$$

$$\frac{|\mathbf{h}_{b,b_u}\mathbf{v}_{b_u}|^2}{g_{b_u}} \geq G_{b_u}(\mathbf{v}) + \sigma_{b_u}^2, \ \forall b \in \mathcal{B}, u \in \mathcal{U}_b \tag{29e}$$

$$(4), (5). \tag{29f}$$

The equivalence between (28) and (29) can be easily justified following the procedure in the proof for Lemma 1. The nonconvex parts of problem (29) lie in (29b) and (29e) which can be approximated in convex forms as

$$\frac{2z_b^{(n)}}{t_b^{(n)}}z_b - \frac{\left(z_b^{(n)}\right)^2}{\left(t_b^{(n)}\right)^2}t_b \geq \eta_b, \ \forall b \in \mathcal{B} \tag{30}$$

$$\frac{2\Re\left(\tilde{\mathbf{h}}_{b,b_u}^{(n)}\mathbf{v}_{b_u}\right)}{q_{b_u}^{(n)}} - \frac{|\mathbf{h}_{b,b_u}\mathbf{v}_{b_u}^{(n)}|^2 q_{b_u}}{\left(q_{b_u}^{(n)}\right)^2} \geq G_{b_u}(\mathbf{v}) + \sigma_{b_u}^2, \ \forall b \in \mathcal{B}, u \in \mathcal{U}_b \tag{31}$$

respectively. Then, the subproblem solved in iteration n is

$$\underset{\mathbf{v},\{\eta_b\},\{z_b\},\{t_b\},\{g_{b_u}\}}{\text{maximize}} \sum_{b \in \mathcal{B}} \omega_b \eta_b \tag{32}$$

$$\text{subject to} \ \{(4), (5), (29c), (29d), (30), (31)\}.$$

For the rate-dependent signal processing model, we replace constraints in (29d) and (29c) by

$$t_b + z_b^2 p_{\text{SP}} \leq \frac{z_b^2}{\eta_b}, \ \forall b \in \mathcal{B}$$
$$t_b \geq P_{\text{cir},b} + P_{\text{PA},b}(\mathbf{v}), \ \forall b \in \mathcal{B}.$$

The same transformation can be applied also to the following problems.

3.3.3 Weighted product EEmax problem
WPEE metric has been considered in power control problems so far [5, 26]. However, to the best of our

knowledge, beamforming designs for WPEE maximization have not been yet investigated. We show below that the proposed framework can be straightforwardly applied to the problem with such metric. The problem of beamforming designs for WPEE maximization reads

$$\underset{\mathbf{v}}{\text{maximize}} \quad \prod_{b \in \mathcal{B}} \left(\frac{\sum_{u \in \mathcal{U}_b} r_{b_u}(\mathbf{v})}{P_{\text{BS},b}(\mathbf{v})} \right)^{\omega_b} \quad \text{subject to} \quad \{(4), (5)\}. \tag{33}$$

Also, we introduce new slack variables $\{\eta_b\}, \{z_b\}, \{t_b\}$, and $\{g_{b_u}\}$ then translate (33) into a more tractable form given as

$$\underset{\mathbf{v}, \{\eta_b\}, \{z_b\}, \{t_b\}, \{g_{b_u}\}}{\text{maximize}} \quad \prod_{b \in \mathcal{B}} (\eta_b)^{\omega_b} \tag{34}$$
$$\text{subject to} \{(4), (5), (29b), (29c), (29d), (29e)\}.$$

Again, we can justify the equivalence between (33) and (34) at the optimum similar to that for Lemma 1. We note that the objective function of (34) is generally neither concave nor convex since the exponents $\{\omega_b\}_b$ are arbitrary positive values. A simple way to overcome the issue is to scale the exponents so that the objective function turns into a concave monomial function which is conic quadratic representable [33]. Particularly, we can always find $\alpha > 1$ such that $\tilde{\omega}_b = \frac{\omega_b}{\alpha}$ for all b and $\sum_b \tilde{\omega}_b \leq 1$. Then, $\prod_{b \in \mathcal{B}} (\eta_b)^{\tilde{\omega}_b}$ is concave monomial. We also note that the optimal solution to (34) stays the same under the scale. Now, we are ready to arrive at the convex subproblem solved at iteration n of the SCA algorithm given as

$$\underset{\mathbf{v}, \{\eta_b\}, \{z_b\}, \{t_b\}, \{g_{b_u}\}}{\text{maximize}} \quad \prod_{b \in \mathcal{B}} (\eta_b)^{\tilde{\omega}_b} \tag{35}$$
$$\text{subject to} \{(4), (5), (29c), (29d), (30), (31)\}.$$

3.3.4 Max-min fairness energy efficiency
The problem of maxminEE is given by

$$\underset{\mathbf{v}}{\text{maximize}} \quad \underset{b \in \mathcal{B}}{\min} \frac{\sum_{u \in \mathcal{U}_b} r_{b_u}(\mathbf{v})}{P_{\text{BS},b}(\mathbf{v})} \quad \text{subject to} \quad \{(4), (5)\}. \tag{36}$$

We note that the centralized and distributed solutions for (36) have been provided in [10] and [11], respectively. Here, for complete discussion and self containment, we provide the main steps of solving the problem. Specifically, with the newly introduced variables $\eta, \{z_b\}, \{t_b\}$, and $\{g_{b_u}\}$, (36) can be equivalently written as

$$\underset{\mathbf{v}, \eta, \{z_b\}, \{t_b\}, \{g_{b_u}\}}{\text{maximize}} \quad \eta$$
$$\text{subject to} \left\{ \eta \leq \frac{z_b^2}{t_b}, \forall b \in \mathcal{B}; (4), (5), (29c), (29d), (29e) \right\} \tag{37}$$

where η represents the minimum EE among all BSs. A convex approximation of (37) solved in iteration n is

$$\underset{\mathbf{v}, \eta, \{z_b\}, \{t_b\}, \{g_{b_u}\}}{\text{maximize}} \quad \eta \quad \text{subject to} \{(4), (5), (29c), (29d), (30), (31)\}. \tag{38}$$

3.4 SOCP formulations of approximate programs
It is clear that the convex approximate problems (27), (32), (35), and (38) are general convex programs due to the logarithmic constraints, i.e., (24d) and (29d), and the nonlinear model of PA's efficiency in (24c) and (29c). Although off-the-shelf solvers are applicable to solve such programs, the computational complexity to output the solutions is relatively high in general [33]. Interestingly, it turns out that these constraints can be represented by second-order cone (SOC) constraints which can take the advantages of more powerful SOCP solvers to reduce the computational effort. In the rest of this section, we discuss the methods that can invoke the hidden SOC representation of the approximated convex programs.

We first consider constraint (29c) which can be equivalently transformed as

$$(29c) \Leftrightarrow \begin{cases} \sqrt{\sum_{u \in \mathcal{U}_b} |[\mathbf{v}_{b_u}]_m|^2} \leq u_{b,m}, \forall b \in \mathcal{B}, m = 1, \ldots, M \\ t_b \geq \sum_{m=1}^{M} u_{b,m} + P_{\text{cir},b}, \forall b \in \mathcal{B} \end{cases} \tag{39}$$

where $\{u_{b,m}\}_{b,m}$ are slack variables. We can see that the first type of constraint in the equivalent formulation is SOC while, the second one is linear. Constraint (24c) is treated similarly and skipped for conciseness.

We now focus on constraint (29d) whose equivalent formulation is given as

$$(29d) \Leftrightarrow \begin{cases} z_b^2 \leq \sum_{u \in \mathcal{U}_b} \beta_{b_u}, \forall b \in \mathcal{B} \\ \log(1 + g_{b_u}) \geq \beta_{b_u}, \forall b_u \end{cases} \tag{40}$$

where $\{\beta_{b_u}\}_{b_u}$ are slack variables. Remark that (40) is SOC representable ([33], Section 3.3). Because the first type of constraint on the right side is SOC representable, we only have to deal with the second one. From now on, for notational convenience, we consider constraint $\log(1 + x) \geq y$ where x and y are positive variables. In the following, we provide four different approaches translating the constraint into SOC representations.

3.4.1 Conic approximation of exponential cone
The first approach approximates $\log(1 + x) \geq y$ by a set of conic constraints based on the result in ([34], Example 4), which has been particularly applied to reduce complexity of solving EE problems in ([7], (31)) and ([10], (13)). The detailed formulation of the conic constraints approximating logarithmic function is omitted here due to the space limitation.

In some settings, using conic approximation of exponential cone could cause a significant increase in per-iteration complexity due to a large number of additional slack variables. This issue is avoided by the approaches presented following.

3.4.2 Equivalently SCA-applicable constraint

The second approach equivalently rewrites $\log(1 + x) \geq y$ as a nonconvex but SCA-applicable constraint. To see this, let us multiply both sides of the constraint by x, i.e.,

$$x \log(1 + x) \geq xy \qquad (41)$$

Since $x \log(1 + x)$ is convex, we can apply SCA principles on (41). A lower bound of $x \log(1 + x)$ is given as

$$x \log(1 + x) \geq d^{(n)} x - c^{(n)} \qquad (42)$$

where $c^{(n)} \triangleq \frac{(x^{(n)})^2}{x^{(n)} + 1}, d^{(n)} \triangleq \frac{x^{(n)}}{x^{(n)} + 1} + \log\left(1 + x^{(n)}\right)$, and $x^{(n)}$ is some positive value. Then, an approximation of (41) is

$$d^{(n)} x - c^{(n)} \geq xy \qquad (43)$$

which can be represented as SOC constraint, i.e,

$$\| [\, x + y - d^{(n)} \quad 2\sqrt{c^{(n)}} \,] \|_2 \leq x - y + d^{(n)}. \qquad (44)$$

3.4.3 Concave lower bound of the logarithm

We can use the well-known inequality of logarithmic function given as

$$\log(1 + z) \geq z(1 + z)^{-1} \qquad (45)$$

for all $z > -1$. By replacing z on both sides of (45) by $\frac{x - x^{(n)}}{x^{(n)} + 1}$ we arrive at

$$\log(1 + x) \geq \log(1 + x^{(n)}) + (x - x^{(n)})(1 + x)^{-1} \qquad (46)$$

for all $x \geq 0$. Now, we can easily check that (46) satisfies the three conditions (a)–(c). Thus, the valid approximate of $\log(1 + x) \geq y$ is

$$\log\left(1 + x^{(n)}\right) + \left(x - x^{(n)}\right)(1 + x)^{-1} \geq y. \qquad (47)$$

Interestingly, (47) contains a hidden SOCP representation given as

$$\|2\sqrt{1 + x^{(n)}}, \log(1 + x^{(n)}) - y - x\|_2 \leq \log(1 + x^{(n)}) - y + x + 2. \qquad (48)$$

3.4.4 Quadratic lower-bound of the logarithm

We can directly approximate (29d) under SCA principles without requiring the transformation step (40). Specifically, we use the following concave quadratic lower-bound derived based on the Lipschitz continuity of the logarithm [25]

$$\log(1 + x) \geq \log\left(1 + x^{(n)}\right) + \frac{(x - x^{(n)})}{1 + x^{(n)}} - \frac{C}{2}\left(x - x^{(n)}\right)^2. \qquad (49)$$

With $C \geq 1$, the inequality holds for all $x \geq 0$ and $x^{(n)} \geq 0$. As a result, an approximation of (29d) can be written as

$$\substack{2 \\ b} + \frac{C}{2}\left(g_{b_u} - g_{b_u}^{(n)}\right)^2 \leq \sum_{u \in \mathcal{U}_b} \log\left(1 + g_{b_u}^{(n)}\right) + \frac{\left(g_{b_u} - g_{b_u}^{(n)}\right)}{1 + g_{b_u}^{(n)}} \qquad (50)$$

which is indeed a rotated SOC constraint. It is worth noting that constant C has large impact on the tightness of the approximation (49), and thus, it influences the convergence speed of the iterative algorithm. More specifically, a smaller value of C implies a tighter approximation and may increase the convergence speed (the discussion is numerically justified in Fig. 7).

3.5 Distributed implementation

The algorithms in Section 3.3 are designed in a centralized fashion under the assumption that each BS (or a central controller) perfectly knows all the channel state information in the network. From the practical perspective, distributed solutions may be more attractive. Note that the conventional approaches are not suitable for decentralized implementation, since updating the parameterized values requires a central node [35, 36]. In contrast, the SCA-based algorithms can be easily carried out in distributed manner. In fact, distributed implementations of the SCA solutions for EEmax problems have been provided in [25, 37, 38]. We remark that distributed implementation is preferred to SWEE, WPEE, and EE-fairness due to their goal of achieving EE of individual node. In what follows, we present how to solve SWEE maximization problem (28) distributively and note that the procedure can be applied to WPEE and EE-fairness problems with slight modifications.

We first assume that each BS has (perfect) CSIs of the channels from itself to all users in the network, which is referred to as local CSI. This is a basic assumption in the distributed setting which has been adopted in [25, 37, 38]. The main idea of the proposed approach is to solve the convex subproblem (32) distributively using the alternating direction method of multipliers (ADMM) [39]. To do so, the vital step is to recognize the terms which need to be decoupled. From (32), we observe that excluding the inter-cell interference terms $\{G_{b_u}(\mathbf{v})\}$, all the other terms are readily local. For clarity, let us rewrite (31) as

$$\frac{2\Re(\tilde{\mathbf{h}}_{b,b_u}^{(n)} \mathbf{v}_{b_u})}{q_{b_u}^{(n)}} - \frac{|\mathbf{h}_{b,b_u} \mathbf{v}_{b_u}^{(n)}|^2 q_{b_u}}{(q_{b_u}^{(n)})^2}$$

$$\geq \bar{G}_{b_u}(\bar{\mathbf{v}}_b) + \sum_{k \in \mathcal{B} \setminus \{b\}} \sum_{i \in \mathcal{U}_k} |\mathbf{h}_{k,b_u} \mathbf{v}_{k_i}|^2, \ \forall b \in \mathcal{B}, u \in \mathcal{U}_b \qquad (51)$$

Nguyen *et al. EURASIP Journal on Wireless Communications and Networking*   (2018) 2018:207

Page 11 of 18

where $\bar{\mathbf{v}}_b \triangleq \left[\mathbf{v}_{b_1}^T, \mathbf{v}_{b_2}^T, ..., \mathbf{v}_{b_U}^T\right]^T$, and $\bar{G}_{b_u}(\bar{\mathbf{v}}_b) \triangleq \sum_{i\in\mathcal{U}_b\setminus\{u\}}$ $|\mathbf{h}_{b,b_u}\mathbf{v}_{b_i}|^2 + \sigma_{b_u}^2$ represents the intra-cell interference plus noise. To deal with the inter-cell interference, let us introduce the variables $\{\theta_{k,b_u}\}_{k,b_u}$. Then, (51) is equivalent to the following set of constraint

$$\frac{2\Re\left(\tilde{\mathbf{h}}_{b,b_u}^{(n)}\mathbf{v}_{b_u}\right)}{q_{b_u}^{(n)}} - \frac{|\mathbf{h}_{b,b_u}\mathbf{v}_{b_u}^{(n)}|^2 q_{b_u}}{\left(q_{b_u}^{(n)}\right)^2}$$

$$\geq \bar{G}_{b_u}(\bar{\mathbf{v}}_b) + \sum_{k\in\mathcal{B}\setminus\{b\}}\theta_{k,b_u}, \ \forall b\in\mathcal{B}, u\in\mathcal{U}_b \qquad (52)$$

$$\theta_{k,b_u} \geq \sum_{i\in\mathcal{U}_k}|\mathbf{h}_{k,b_u}\mathbf{v}_{k_i}|^2, \ \forall b\in\mathcal{B}, u\in\mathcal{U}_b, k\in\mathcal{B}\setminus\{b\}. \qquad (53)$$

With the transformation, we turn to handling the term $\sum_{k\in\mathcal{B}\setminus\{b\}}\theta_{k,b_u}$ in (52) for distributed implementation since the constraints in (53) can be treated locally. To this end, let us introduce new local variables for each interference term θ_{b,k_i} as $\tilde{\theta}_{b,k_i}^b$ and $\tilde{\theta}_{b,k_i}^k$ optimized at BS b and BS k, respectively. To make sure that these local variables are equal to each other, we further add an equality constraint

$$\theta_{b,k_i} = \tilde{\theta}_{b,k_i}^b = \tilde{\theta}_{b,k_i}^k, \forall b\in\mathcal{B}, k\in\mathcal{B}\setminus\{b\}, i\in\mathcal{U}_k. \qquad (54)$$

Now, we can write (32) equivalently as

$$\text{maximize} \quad \sum_{b\in\mathcal{B}}\omega_b\eta_b \qquad (55a)$$

$$\text{subject to} \ \left(\bar{\mathbf{v}}_b, \{\eta_b\}, \{z_b\}, \{t_b\}, \{g_{b_u}\}, \{\tilde{\boldsymbol{\theta}}_b\}\right)\in\mathcal{S}_b, \forall b\in\mathcal{B} \qquad (55b)$$

$$\boldsymbol{\theta}_b = \tilde{\boldsymbol{\theta}}_b, \forall b\in\mathcal{B} \qquad (55c)$$

where $\boldsymbol{\theta}_b \triangleq \{\{\theta_{k,b_u}\}_{k\in\mathcal{B}\setminus\{b\},u\in\mathcal{U}_b}, \{\theta_{b,k_i}\}_{k_i\in\bar{\mathcal{U}}_b}\}, \tilde{\boldsymbol{\theta}}_b \triangleq \{\{\tilde{\theta}_{k,b_u}^b\}_{k\in\mathcal{B}\setminus\{b\},u\in\mathcal{U}_b}, \{\tilde{\theta}_{b,k_i}^b\}_{k_i\in\bar{\mathcal{U}}_b}\}, \bar{\mathcal{U}}_b \triangleq \bigcup_{k\in\mathcal{B}\setminus\{b\}}\mathcal{U}_k$, and

$$\mathcal{S}_b \triangleq \Bigg\{(\bar{\mathbf{v}}_b, \{\eta_b\}, \{z_b\}, \{t_b\}, \{g_{b_u}\}, \{\tilde{\boldsymbol{\theta}}_b\})\Big|$$

$$\frac{2\Re\left(\tilde{\mathbf{h}}_{b,b_u}^{(n)}\mathbf{v}_{b_u}\right)}{q_{b_u}^{(n)}} - \frac{|\mathbf{h}_{b,b_u}\mathbf{v}_{b_u}^{(n)}|^2 q_{b_u}}{\left(q_{b_u}^{(n)}\right)^2} \geq \bar{G}_{b_u}(\bar{\mathbf{v}}_b)$$

$$+ \sum_{k\in\mathcal{B}\setminus\{b\}}\theta_{k,b_u}^b, \ \forall u\in\mathcal{U}_b$$

$$\theta_{b,k_i}^b \geq \sum_{u\in\mathcal{U}_b}|\mathbf{h}_{b,k_i}\mathbf{v}_{b_u}|^2, \ \forall k_i\in\bar{\mathcal{U}}_b$$

$$\{(4), (5), (29c), (29d), (30), (31)\}_b\Bigg\} \qquad (56)$$

is the feasible set for the local variables at BS b. In (56), the notation $\{\}_b$ means to take only the constraints related to BS b. Problem (55) is in a form of global consensus problem and can be optimally solved using the standard ADMM procedure. We refer interested readers to [11, 37] for further details.

4 Numerical result

We evaluate the performances of the different algorithms presented above. The general (fixed) simulation parameters are taken from Table 1, and the ones which are changed in the simulations are given in the caption of the corresponding figures. A network of B base stations is considered, and U_b users are randomly dropped to the coverage area of each base station. The user channels follow the Rayleigh distribution.

4.1 Comparison on the convergence and the performance

In the first set of simulations, we compare the SCA methods against the conventional based on the fractional program (FP) ones in terms of the convergence rate and achieved EE performances. The following beamforming designs are considered in the evaluation:
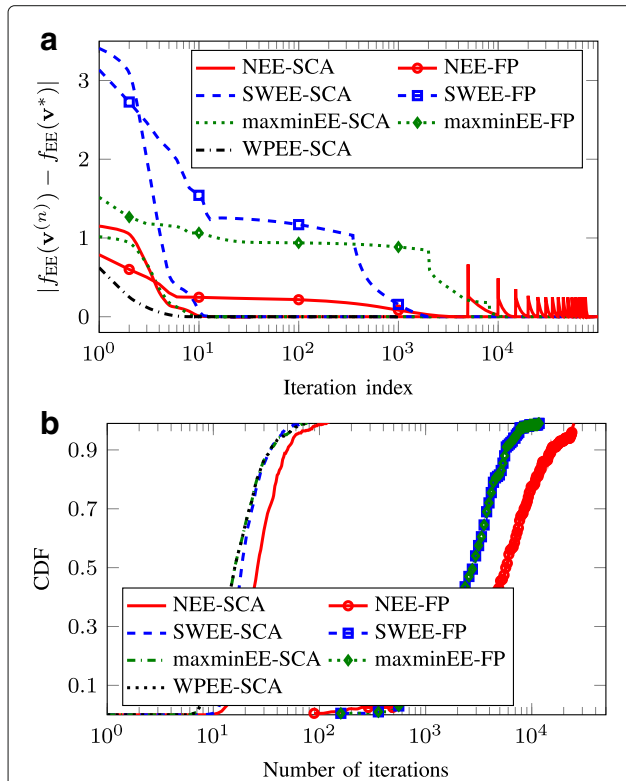
- NEE-SCA—the SCA procedure for solving (23).
- SWEE-SCA—the SCA procedure for solving (28). We set equal priorities for all the nodes, i.e, $\omega_b = 1, \forall b$.
- WPSEE-SCA—the SCA procedure for solving (33). Similar to SWEE-SCA, $\omega_b = 1, \forall b$.
- maxminEE-SCA—the SCA procedure for solving (36).
- NEE-FP—the beamforming design based on FP proposed in [16].
- SWEE-FP—the beamforming design based on FP proposed in [9]. This scheme has been studied for a MIMO channel. However, it can be easily simplified for MISO channels by setting the number of receive antenna to one.
- maxmin-FP—the beamforming design based on FP proposed in [17]. Although this approach has been proposed for a multi-cell joint transmission system, we can easily simplify it for the multi-cell coordinated beamforming case by properly rewriting the signal and interference terms [17] Remark 2.

We note that for the WPEE metric, only the performance of the SCA-based method is studied as beamforming designs for this metric based on the FP framework have not yet been proposed to the best of our knowledge. In addition, to reduce the computational burden for the simulations, we terminate the iterative processes of all the considered algorithms either when the increase in the objective between two consecutive iteration is less than 10^{-5} or after 10^6 iterations. Also, for a fair comparison, we only consider the conventional power model as in [5, 8, 9, 16, 17], i.e., fixed signal processing power and PAs' efficiency. The results for the general power consumption model are reported in Section 4.6.

4.2 Convergence comparison of the SCA and PF algorithms

Figure 4 presents the convergence behavior of the considered EE approaches. Particularly, Fig. 4a compares the convergence speed of the SCA frameworks and the FP methods in terms of the number of iterations for one random channel realization by showing the gap between the current objective value at iteration n and the achieved objective value after the termination of the algorithms, i.e., $|f_{EE}(\mathbf{v}^n) - f_{EE}(\mathbf{v}^*)|$. As can be seen, the SCA-based methods have steady monotonic converge properties, e.g., achieve the objective value of the convergence point after ten iterations in the considered setting. For the FP-based approaches, even hundreds or thousands of iterations can be required to reach the convergence, while the monotonic convergence is not always guaranteed, e.g., for the NEE-FP method.

To complete the comparison in terms of convergence speed between the SCA- and FP-based methods, we provide in Fig. 4b the cumulative distribution function (CDF) of the total number of iterations needed for convergence. It is observed that for 90% of channel realizations, the SCA converges after 30 iterations while the FP methods need even thousands of iterations to terminate. This

observation again shows the superiority of the SCA algorithms in terms of complexity compared to the conventional approaches.

4.3 EE performance comparison of the SCA and FP algorithms

Figure 5 plots the average performances of the SCA- and FP-based methods in terms of the achieved NEE, the sum

Fig. 4 Convergence behavior of different EE schemes with $P_b = 30$ dBm. **a** Convergence rate for one channel realization. **b** CDF of the numbers of required iterations over 500 channel realizations
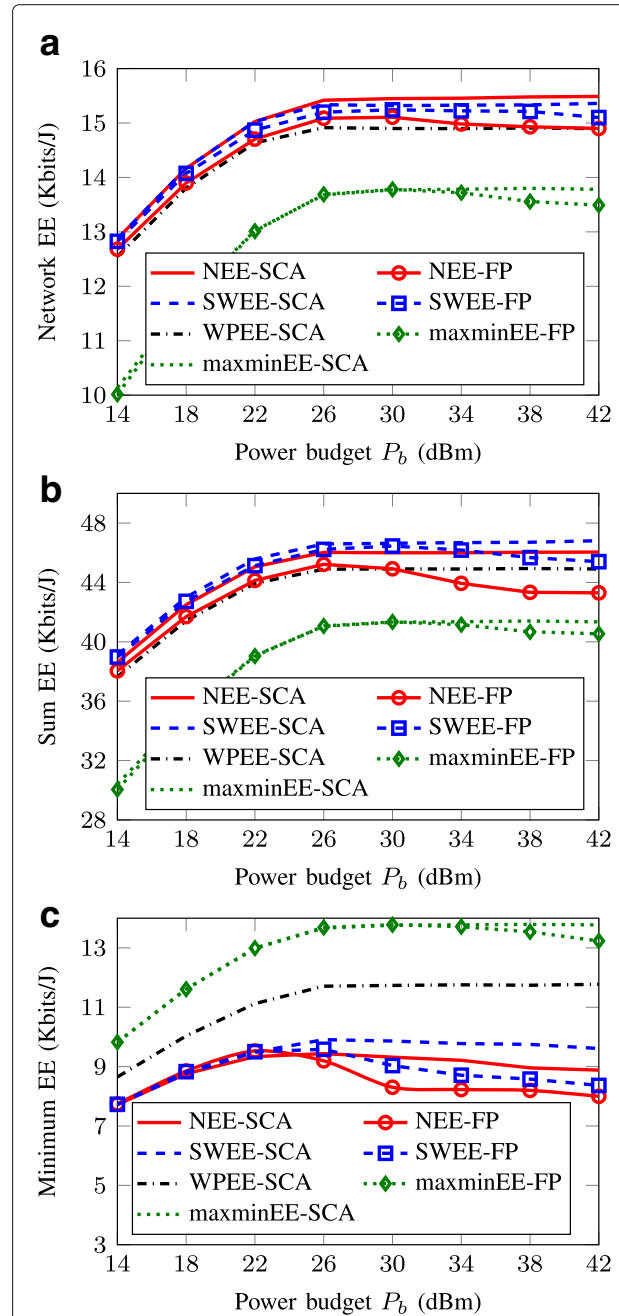
Fig. 5 Achieved EE performance versus power budget P_b. **a** Network EE. **b** Sum EE. **c** Minimum EE

EE, and the minimum EE versus the maximum transmit power budget P_b. Our first observation is that the SCA method maximizing a specific EE metric achieves approximately the same EE performance compared to the corresponding FP method in small and moderate power regions. This again implies the effectiveness of the SCA framework in solving the EE maximization problems as it can offer similar performance compared to the conventional ones but with much reduced complexity. Another observation is that the achieved EE with all the approaches saturates when the power budget is sufficiently large. This is because in the large power regime, the data rate logarithmically scales with the transmit power while the power consumption increases linearly with the transmit power. Thus, whenever the gain in achieved throughput cannot compensate for the increase of power consumption, the EE methods do not use the excess power to further increase data rate so as to maintain a high value EE. This fact has been discussed in many EE maximization related works [5, 8, 40]. On the other hand, we can see that the EEs achieved by the FP methods slightly downgrade for large value of P_b. The reason can be explained as follows. When P_b increases, the feasible set of the EE problems is expanded which results in the increasing number of iterations required for the convergence. However, due to the threshold on the maximum iterations for the iterative process, the FP methods may not reach the suboptimal solutions within 10^6 iterations. Consequently, they may output poor performances leading to the decrease of average achieved EE value. This observation again points out the drawback of the two-layer iterative procedure in practice. Let us then evaluate the achieved performances of the considered methods with respect to each EE metric. It is obvious that the NEE maximization methods outperform the other schemes in terms of the achieved NEE (in Fig. 5a), while the SWEE methods offer the best sum-EE values (in Fig. 5b). In terms of minimum EE, the maxminEE methods achieve the best performance as they aim at maintaining the balance of EE among all parties (in Fig. 5c). However, the maxminEE approaches suffer a loss in NEE and sum-EE performances. The WPEE metric, as expected, offers a better minimum EE than the SWEE and NEE criteria. We note that individual EE is one of the key features in many network scenarios (e.g., heterogeneous networks), and thus, a per-node EE performance and EE fairness will be the main focus in the next numerical experiment.

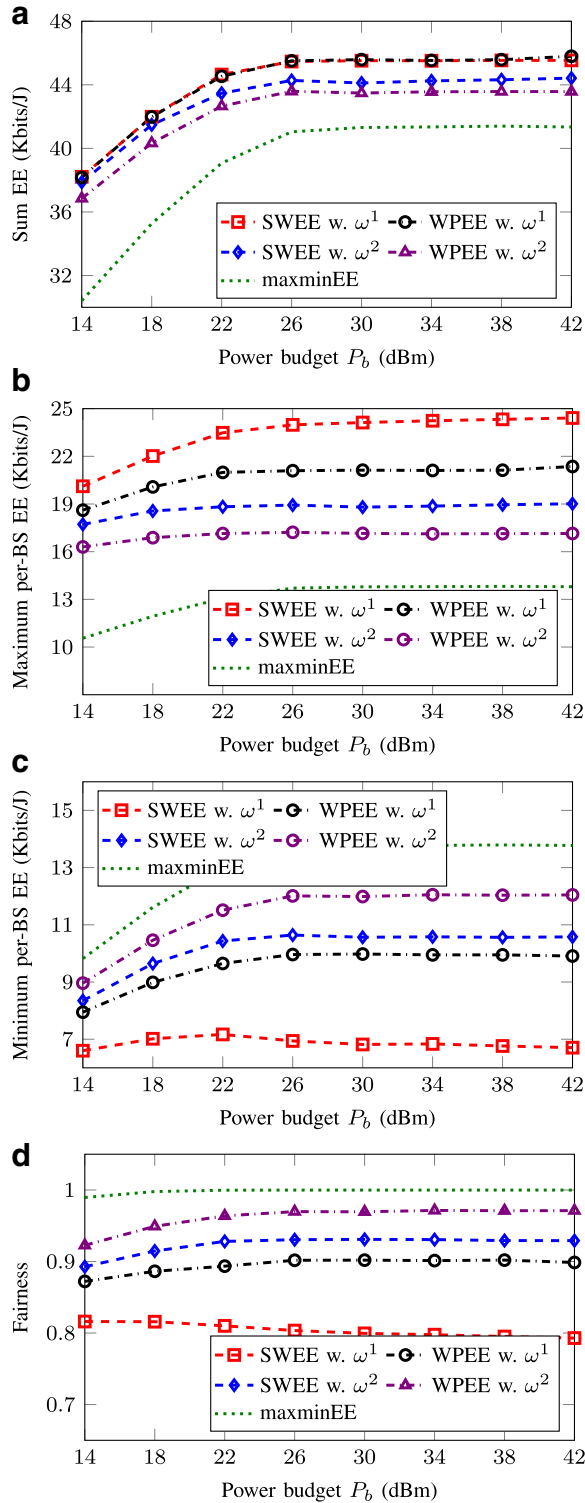### 4.4   Achieved per-BS EE performance
In Fig. 6, we compare the achieved fairness when using different fairness-oriented EE metrics, i.e., SWEE, WPEE, and maxminEE, as a function of the maximum transmit power budget P_b. Note that the fairness is considered in terms of energy efficiency and not rates. Specifically,

we consider two settings of weighting vector, i.e., $\boldsymbol{\omega}^1 = [0.7, 0.5, 0.3]$ and $\boldsymbol{\omega}^2 = [0.3, 0.5, 0.7]$, where each value implies the priority weight for a corresponding BS. In this experiment, users in cell 1 are dropped in the radius of 200 m to its serving BS while for cell 3, users' locations are near the cell edge. The served users of BS 2 are randomly placed in its coverage region. The figure is plotted to see how the priority parameters alter the per-BS EE behaviors. The average sum-EE, maximum and minimum EE, and EE fairness measure[4] among all the BSs are plotted in Fig. 6a–d, respectively. Similar results as in Fig. 5 can be observed. Particularly, the achieved EE values remain unchanged when P_b is sufficiently large. Moreover, it is seen that the SWEE methods outperform the other schemes in terms of sum-EE and maximum per-BS EE values. This is clear since maximizing the sum of individual EEs is the objective of the SWEE methods. In terms of minimum EE among all nodes, it is obvious that the maxminEE scheme obtains the best performance followed by the WPEE and SWEE criteria (with same assigned priority). Another important observation is that by assigning different priority weights $\boldsymbol{\omega}$ for SWEE and WPEE metrics, we can adjust the achieved EE of each node. It is discovered that with $\boldsymbol{\omega}^1$, more priority is given to BS 1 leading to an improvement in the sum-EE performance for these two schemes. In addition, since BS 3 is more penalized, the gap between maximum and minimum per-BS EE values is enlarged and, thus, implying high EE unfairness among the BSs. On the contrary, since $\boldsymbol{\omega}^2$ prioritizes BS 3 and restricts BS 1, it reduces the sum-EE performance of the network but encourages the fairness among all parties. As a conclusion, the EE fairness measure in Fig. 6d shows that the SWEE and WPEE schemes can tune the EE fairness of the system by the priority parameters while the maxminEE can establish the absolute fairness among all the per-BS EEs. Also, with the same weighting vector, the WPEE metric outperforms the SWEE in terms of EE balancing. The WPEE achieves a better trade-off between fairness and EE performance compared to the two other schemes SWEE and maxminEE.
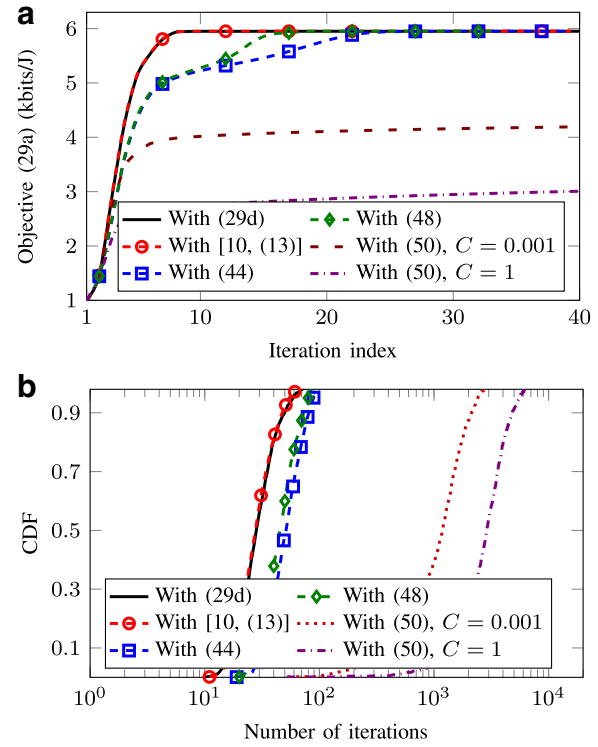
4.5   SCA with different conic approximations
We now illustrate the performances of the SOCP formulations provided in Section 3.4, by focusing on the SWEE metric. In Fig. 7a, we compare the convergence rate of the objective (29a) when the original logarithmic constraint (29d) and its SOC constraints ([10], (13)), (44), (48), and (50) are used in (32). It is seen that with ([10], (13)), the algorithm converges with the same rate as using (29d) and faster than the other SOC approximations. This is understandable because the set of conic constraints ([10], (13)) is in fact a tight approximation of logarithmic function up to a fixed accuracy level. On the other hand, the other ones are the upper bounds of the logarithmic function

Nguyen *et al. EURASIP Journal on Wireless Communications and Networking* (2018) 2018:207

Page 14 of 18

Fig. 6 Averaged EE versus transmit power budget P_b. For SWEE and WPEE schemes, we use two sets of priority parameters, i.e., $\omega^1 = [0.7, 0.5, 0.3]$ and $\omega^2 = [0.3, 0.5, 0.7]$. **a** Sum EE. **b** Maximum EE. **c** Minimum EE. **d** Fairness index

Fig. 7 Convergence behavior of the SCA algorithm solving (32) with different conic approximations. We take $P_b = 40$ dBm. **a** Convergence rate for one random channel realization. **b** CDF of the number of required iterations over 1000 channel realizations

which are tight only in the fixed point at each iteration. Between (44) and (48), we can observe that (48) offers a better convergence rate. This may be understood as (48) directly approximates the bound of logarithmic function while (48) is derived from the bound of the equivalent transformation of (29d). For the SOC approximation based on (50), it is seen that the convergence behavior largely depends on *C*. When a large value of *C* is used, the objective slowly converges while the convergence rate significantly improves when using small value of *C*. The performance agrees with the analysis of (50) which argues that the smaller *C* provides a tighter approximation in (50) and, thus, can lead to a faster convergence.

In Fig. 7b, we depict the CDF of the number of iterations required for convergence with different SOCP formulations of (32). As expected, the result is consistent with that observed from Fig. 7a. Specifically, adopting the set of conics constraints ([10], (13)) to approximate the logarithmic function does not require more iterations for convergence compared to using the original constraint (29d). Also, the number of iterations of (32) with (48) is smaller than that of applying (44). In general, we can see that for 90% of the channel realization, the approximation methods ([10], (13)), (44), and (48) can provide a good
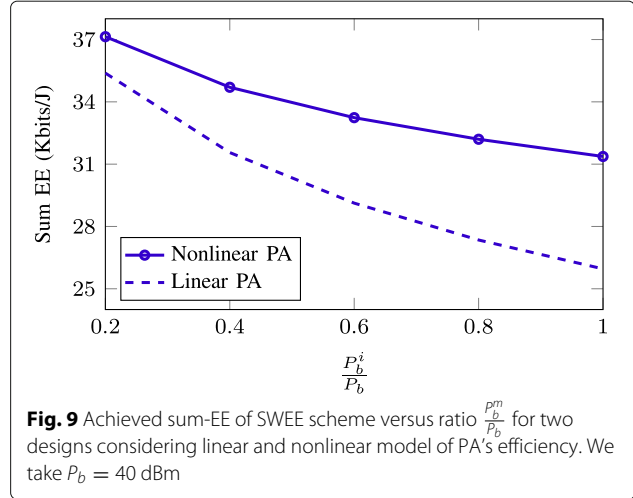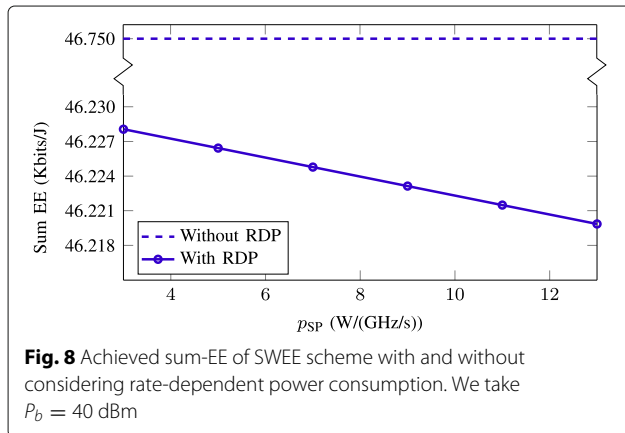
convergence rate which is smaller than 100 iterations. On the other hand, (50) results in slow convergence speed in the considered setting.

4.6 Achieved performance with general power consumption model

In this numerical experiment, we provide insights to the impact of rate-dependent signal processing power and the nonlinear model of PA's efficiency on the achieved EE performance. Figure 8 compares the achieved sum-EE of the SWEE schemes without and with considering the rate-dependent power (RDP) consumption, which are labeled as "Without RDP" and "With RDP," respectively. The curve "Without RDP" is obtained by solving problem (29) with $p_{SP} = 0$. Then, we recalculate the EE values for "With RDP" scheme with the given p_{SP} in the horizontal axis. Our observation is that as p_{SP} increases, the achieved EE monotonically decreases. This is understandable because, for a fixed achieved data rate, higher p_{SP} increases the total power consumption and thus degrades the EE. This result suggests that RDP may be included when optimizing the EE performance of a wireless network.

Next, in Fig. 9, we evaluate the impact of nonlinear PA's efficiency on the sum-EE achieved by the SWEE method. For this purpose, we perform the EE optimization based on the nonlinear PA's efficiency model (9). As PA's efficiency does not depend on P_b but P_b^m, we fix $P_b = 40$ dBm and plot the achieved sum EE as a function of the ratio $\frac{P_b^m}{P_b}$. The two following schemes are compared:

- "Nonlinear PA"—problem (28) is solved using the nonlinear PA's efficiency model (9) with $\epsilon_{max} = 0.35$.
- "Linear PA"—problem (28) is solved using the PA's power consumption model (8) with fixed PA's efficiency $\epsilon = 0.35$. The resulting beamforming solution is used to compute the actual EE performance following the PA's efficiency model (9).

Fig. 8 Achieved sum-EE of SWEE scheme with and without considering rate-dependent power consumption. We take $P_b = 40$ dBm

Fig. 9 Achieved sum-EE of SWEE scheme versus ratio $\frac{P_b^m}{P_b}$ for two designs considering linear and nonlinear model of PA's efficiency. We take $P_b = 40$ dBm

As can be seen, "Linear PA" scheme is inferior to "Nonlinear PA" one which clearly shows that the power modeling has a remarkable influence on the achieved EEs. More specifically, the EE maximization based on the assumption that the PA's efficiency is the same regardless of the output power potentially degrades the EE performance in practical implementation, where the PA's efficiency actually depends on the desired output power [21–24]. Another observation is that the achieved EEs of both schemes decrease against the increase of $\frac{P_b^m}{P_b}$. The result can be explained as follows. Recall that the effective PA's efficiency depends on P_b^m and the actual transmit power (ATP) on the antenna, that is, with increasing P_b^m, the efficiency slope of that PA is changed so that the efficiency is worse in the lower ATP regime (see (9)). Thus, the decreased PA efficiency simply deteriorates the achieved EE.

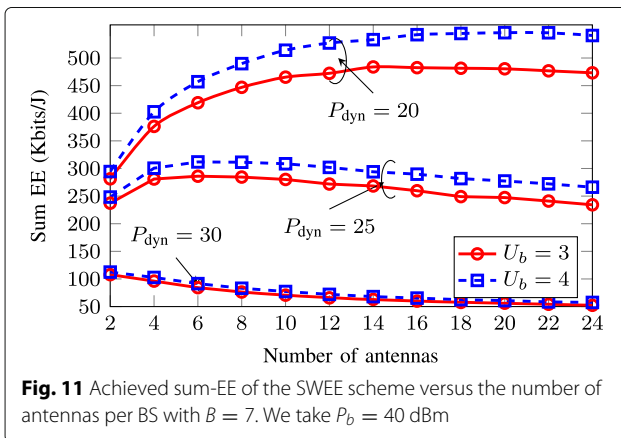### 4.7 Achieved EE in large-scale network settings

In the final set of numerical experiments, we illustrate how the achieved EE behaves in a larger network. A seven-cell network model which consists of $B = 7$ BSs is considered. For simplicity, we adopt the conventional power consumption model with (8) and $p_{SP} = 0$ and simulate only the SWEE scheme. Figure 10 shows the achieved EE versus the number of users per cell U_b with different values of transmit power P_b when the number of per-BS antennas is fixed to $M = 4$. It is seen that the EE values increase with the increasing number of served users. This is because the sum rate is an increasing function of the number of users and, thus, increased when more users are involved in the transmission. We can also observe that for fixed U_b, the EE grows if the power budget is larger. However, when P_b is large enough, further increasing P_b does not bring significant improvement in EE. This result is consistent with that observed in Figs. 5 and 6.

Fig. 10 Achieved sum-EE of the SWEE scheme versus the number of users per cell U_b with $B = 7, M = 4$
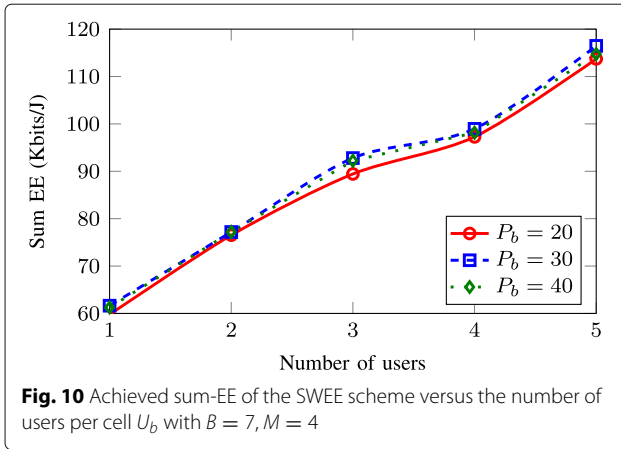
In Fig. 11, we show the achieved EE versus the number of per-BS antennas M for different values of P_{dyn} and U_b. As can be seen, for small and moderate values of P_{dyn}, the achieved EE first increases then decreases as M keeps increasing, while for the large value of P_{dyn}, it monotonically decreases. The reason can be explained as follows. Recall the fact that additional antennas provide more degree of freedom which leads to the improvement in the achievable data rate. However, since the total power consumption linearly scales with the transmit antennas, adding more antennas consumes more circuit power. Thus, as long as the the benefit offered from transmitting with additional antennas is beyond the cost of the power consumption, the achieved EE increases. Otherwise, increasing the number of transmit antennas degrades the achieved EE. Another observation, which agrees with the result in Fig. 10, is that adding more users improves the achieved EE.

5 Conclusions

We have provided a summary and performance comparison of various algorithms for the problems of EE

Fig. 11 Achieved sum-EE of the SWEE scheme versus the number of antennas per BS with $B = 7$. We take $P_b = 40$ dBm

optimization in multi-cell multiuser MISO downlink, under four energy efficiency metrics. We have reviewed and presented the SCA framework to provide efficient solutions for the energy efficiency optimization. The algorithms have been numerically evaluated and compared with different fractional programming solutions for the same problems. The SCA-based algorithms have been shown to outperform the existing FP ones in terms of convergence speed. This paper can be viewed as a guideline for the application of the SCA in solving the energy-efficient beamforming designs in particular and the nonconvex problems in wireless communications in general.

The EE optimization will be important for the sustainability of the future digital society. Several important problems still remain. For example, acquiring accurate CSI is challenging in practice, and the transmission designs taking into account the effect of imperfect CSI are an important topic to be explored. The impact of data sharing over (wireless) power and bandwidth-limited backhaul in the CoMP joint processing transmission is also an important topic. The hybrid analog/digital beamforming transceiver architecture-based EE optimization for the evolving millimeter wave wireless communications is an important item for the evolving 5G system design. The cloud radio access network (CRAN) architecture with more processing options either close to the antenna or at computing cloud requires also the EE-based design and analysis. Finally, the power consumption in user devices is much more difficult to model and control than that in the base stations or cloud but constitutes a significant portion of the overall network power consumption.

Appendix
Proof of Lemma 1:

For proving the lemma, we show that constraints (24b)–(24e) are active at the optimality by the contradiction. Let $(\mathbf{v}^*, \eta^*, z^*, t^*, \{g^*_{b_u}\})$ be an optimal solution of (24) and suppose that (24e) is not active at the optimum, i.e., $\frac{|\mathbf{h}_{b,b_u} \mathbf{v}^*_{b_u}|^2}{G_{b_u}(\mathbf{v}^*) + \sigma^2_{b_u}} > g^*_{b_u}$ for some b_u. Then, we can scale down the transmit power for user b_u and achieve a new beamformer $\|\hat{\mathbf{v}}_{b_u}\|^2_2$ such that $\|\hat{\mathbf{v}}_{b_u}\|^2_2 = \tau \|\mathbf{v}^*_{b_u}\|^2_2 < \|\mathbf{v}^*_{b_u}\|^2_2$ for $\tau \in (0, 1)$ while keeping the others remaining unchanged, i.e., $\hat{\mathbf{v}}_{b_k} = \mathbf{v}^*_{b_k}$ for all $b_k \neq b_u$. By this way, we can achieve $\frac{|\mathbf{h}_{b,b_u} \hat{\mathbf{v}}_{b_u}|^2}{G_{b_u}(\hat{\mathbf{v}}) + \sigma^2_{b_u}} > g^*_{b_u}$ for all b_u since interference power at all users has reduced. In addition, we have $\hat{t} = P_{\text{total}}(\hat{\mathbf{v}}) < t^* = P_{\text{total}}(\mathbf{v}^*)$. Consequently, we can find $\hat{\eta} = \frac{(z^*)^2}{\hat{t}} > \eta^*$. This contrasts to the fact that $\left(\mathbf{v}^*, \eta^*, z^*, t^*, \left\{g^*_{b_u}\right\}\right)$ is the optimal solution. The same spirit is applied to the other constraints. This completes the proof.

Endnotes

[1] An equal number of users in each cell are assumed merely for lightening up the notations.

[2] The algorithm frameworks provided in this paper can be straightforwardly extended to the case of joint transmission as well.

[3] By nontractable, we mean that it cannot be reformulated as an equivalent convex program, or such a transformation is not known in the literature.

[4] The EE fairness measure is calculated following the index defined in ([41], (1)). Particularly, let us denote $\eta^* = [\eta_1^*, \ldots \eta_B^*]$ as the achieved per-BS EEs after solving, e.g., WSEE problems (28), WPEE problem (33), or maxminEE (36). The fairness index is fairness $= \frac{\left(\sum_{b=1}^{B} \eta_b^*\right)^2}{B \sum_{b=1}^{B} (\eta_b^*)^2}$.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1] Centre for Wireless Communications, University of Oulu, P.O. Box 4500, FI-90014 Oulu, Finland. [2] School of Electrical and Electronic Engineering, University College Dublin, Dublin, Ireland.

References
1. Qualcomm, The 1000x data challenge. Tech. rep. [Online]. Available: http://www.qualcomm.com/1000x
2. D Feng, C Jiang, G Lim, J Cimini, J L, G Feng, G Li, A survey of energy-efficient wireless communication. IEEE Commun. Surv. Tutor. **15**(1), 167–178 (2013)
3. GY Li, Z Xu, C Xiong, C Yang, S Zhang, Y Chen, S Xu, Energy-efficient wireless communications: tutorial, survey, and open issues. IEEE Wirel. Commun. **18**(6), 28–35 (2011)
4. C Isheden, Z Chong, E Jorswieck, G Fettweis, Framework for link-level energy efficiency optimization with informed transmitter. IEEE Trans. Wirel. Commun. **11**(8), 2946–2957 (2012)
5. A Zappone, E Jorswieck, Energy efficiency in wireless networks via fractional programming theory. Found. Trends Commun. Inf. Theory. **11**(3-4), 185–396 (2015)
6. O Tervo, LN Tran, H Pennanen, S Chatzinotas, M Juntti, B Ottersten, in *Proceedings of the 2017 IEEE International Conference on Communications Workshops (ICC Workshops)*. Energy-efficient coordinated multi-cell multi-group multicast beamforming with antenna selection (IEEE, 2017), pp. 1209–1214
7. O Tervo, L-N Tran, M Juntti, Optimal energy-efficient transmit beamforming for multi-user MISO downlink. IEEE Trans. Signal Process. **63**(20), 5574–5588 (2015)
8. DWK Ng, ES Lo, R Schober, Energy-efficient resource allocation in multi-cell OFDMA systems with limited backhaul capacity. IEEE Trans. Wirel. Commun. **11**(10), 3618–3631 (2012)
9. S He, Y Huang, L Yang, B Ottersten, Coordinated multicell multiuser precoding for maximizing weighted sum energy efficiency. IEEE Trans. Signal Process. **62**(3), 1005–1017 (2014)
10. K-G Nguyen, L-N Tran, O Tervo, Q-D Vu, M Juntti, Achieving energy efficiency fairness in multicell multiuser MISO downlink. IEEE Commun. Lett. **19**(8), 1426–1429 (2015)
11. KG Nguyen, QD Vu, M Juntti, LN Tran, Distributed solutions for energy efficiency fairness in multicell MISO downlink. IEEE Trans. Wirel. Commun. **16**(9), 6232–6247 (2017)
12. KG Nguyen, O Tervo, QD Vu, M Juntti, LN Tran, in *Proceedings of the 2017 European Conference on Networks and Communications (EuCNC)*. Energy-efficient transmission strategies for multiantenna downlink (IEEE, 2017), pp. 1–5
13. D Lee, H Seo, B Clerckx, E Hardouin, D Mazzarese, S Nagata, K Sayana, Coordinated multipoint transmission and reception in LTE-advanced: deployment scenarios and operational challenges. IEEE Commun. Mag. **50**(2), 148–155 (2012)
14. O Arnold, F Richter, G Fettweis, O Blume, in *Proc.19th Future Network & Mobile Summit (ICT Summit'10)*. Power consumption modeling of different base station types in heterogeneous cellular networks (IEEE, Florence, 2010), pp. 1–8
15. G Auer, V Giannini, C Desset, I Godor, P Skillermark, M Olsson, MA Imran, D Sabella, MJ Gonzalez, O Blume, A Fehske, How much energy is needed to run a wireless network? IEEE Wirel. Commun. **18**(5), 40–49 (2011)
16. S He, Y Huang, S Jin, L Yang, Coordinated beamforming for energy efficient transmission in multicell multiuser systems. IEEE Trans. Commun. **61**(12), 4961–4971 (2013)
17. S He, Y Huang, S Jin, F Yu, L Yang, Max-min energy efficient beamforming for multicell multiuser joint transmission systems. IEEE Commun. Lett. **17**(10), 1956–1959 (2013)
18. C Isheden, GP Fettweis, in *2010 IEEE Global Telecommunications Conference GLOBECOM 2010*. Energy-efficient multi-carrier link adaptation with sum rate-dependent circuit power (IEEE, 2010), pp. 1–6
19. C Xiong, GY Li, S Zhang, Y Chen, S Xu, Energy-efficient resource allocation in OFDMA networks. IEEE Trans. Wirel. Commun. **60**(12), 3767–3778 (2012)
20. T Wang, L Vandendorpe, On the optimum energy efficiency for flat-fading channels with rate-dependent circuit power. IEEE Trans. Commun. **61**(12), 4910–4921 (2013)
21. S Mikami, T Takeuchi, H Kawaguchi, C Ohta, M Yoshimoto, in *2007 IEEE Radio and Wireless Symposium*. An efficiency degradation model of power amplifier and the impact against transmission power control for wireless sensor networks (IEEE, 2007), pp. 447–450
22. E Björnemo, Energy constrained wireless sensor networks: communication principles and sensing aspects (2009). PhD thesis, Institutionen för teknikvetenskaper
23. D Persson, T Eriksson, EG Larsson, Amplifier-aware multiple-input multiple-output power allocation. IEEE Commun. Lett. **17**(6), 1112–1115 (2013)
24. O Tervo, LN Tran, M Juntti, in *Proceedings of the 2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. Energy-efficient joint transmit beamforming and subarray selection with nonlinear power amplifier efficiency (IEEE, 2016), pp. 763–767

25. O Tervo, A Tölli, M Juntti, LN Tran, Energy-efficient beam coordination strategies with rate-dependent processing power. IEEE Trans. Signal Process. **65**(22), 6097–6112 (2017)
26. L Venturino, A Zappone, C Risi, S Buzzi, Energy-efficient scheduling and power allocation in downlink OFDMA networks with base station coordination. IEEE Trans. Wirel. Commun. **14**(1), 1–14 (2015)
27. W Dinkelbach, On nonlinear fractional programming. Manag. Sci. **13**(7), 492–498 (1967)
28. S Schaible, Fractional programming. i, duality. Manag. Sci. **22**(8), 858–867 (1976)
29. Q Wu, W Chen, DWK Ng, J Li, R Schober, User-centric energy efficiency maximization for wireless powered communications. IEEE Trans. Wirel. Commun. **15**(10), 6898–6912 (2016)
30. G Yu, Q Chen, R Yin, H Zhang, GY Li, Joint downlink and uplink resource allocation for energy-efficient carrier aggregation. IEEE Trans. Wirel. Commun. **14**(6), 3207–3218 (2015)
31. BR Marks, GP Wright, A general inner approximation algorithm for nonconvex mathematical programs. Oper. Res. **26**(4), 681–683 (1978)
32. S Boyd, L Vandenberghe, *Convex Optimization*. (Cambridge University Press, 2004)
33. A Ben-Tal, A Nemirovski, *Lectures on Modern Convex Optimization*. (MPS-SIAM Series on Optimization, SIAM, Philadelphia, 2001)
34. A Ben-Tal, A Nemirovski, On the polyhedral approximations of the second-order cone. Math. Oper. Res. **26**(2), 193–205 (2001)
35. QD Vu, LN Tran, R Farrell, EK Hong, Energy-efficient zero-forcing precoding design for small-cell networks. IEEE Trans. Commun. **64**(2), 790–804 (2016)
36. Q-D Vu, L-N Tran, M Juntti, E-K Hong, Energy-efficient bandwidth and power allocation for multi-homing networks. IEEE Signal Proc. Lett. **63**(7), 1684–1699 (2015)
37. O Tervo, LN Tran, M Juntti, in *Proceedings of the 2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. Decentralized coordinated beamforming for weighted sum energy efficiency maximization in multi-cell MISO downlink (IEEE, 2015), pp. 1387–1391
38. KG Nguyen, LN Tran, QD Vu, M Juntti, in *Proceedings on the 2016 IEEE International Conference on Communications (ICC)*. Distributed energy efficiency fairness optimization by ADMM in multicell MISO downlink, (2016), pp. 1–6
39. S Boyd, N Parikh, E Chu, B Peleato, J Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach. Learn. **3**(1), 1–122 (2011)
40. D Nguyen, L-N Tran, P Pirinen, M Latva-aho, Precoding for full duplex multiuser MIMO systems: spectral and energy efficiency maximization. IEEE Trans. Signal Process. **61**(16), 4038–3050 (2013)
41. R Jain, D-M Chiu, WR Hawe, *A quantitative measure of fairness and discrimination for resource allocation in shared computer system*. (Eastern Research Laboratory, Digital Equipment Corporation, Hudson, 1984)