A framework for energy-efficient design of 5G technologies

Alessio Zappone, Luca Sanguinetti, Giacomo Bacci, Eduard Jorswieck, Mérouane Debbah

A Framework for Energy-Efficient Design of 5G Technologies

Alessio Zappone,* Luca Sanguinetti,†§ Giacomo Bacci,§ Eduard Jorswieck,* and Mérouane Debbah‡
* Technische Universität Dresden, 5G Lab Germany, Dresden, Germany
† Large Systems and Networks Group (LANEAS), Supélec, Gif-sur-Yvette, France
§ Dipartimento di Ingegneria dell'Informazione, University of Pisa, Pisa, Italy
‡ Mathematical and Algorithmic Sciences Lab, Huawei France, Paris, France

***Abstract*—This paper considers the problem of energy efficiency maximization in the uplink of a cluster of multiple-antenna coordinated access points. A framework for energy efficiency optimization is developed in which the signal-to-interference-plus-noise ratio takes a more general expression than existing alternatives so as to encompass most 5G candidate technologies. Two energy efficiency optimization problems are formulated, also considering QoS constraints: 1) network global energy efficiency maximization; 2) worst-case energy-efficient design. These fractional, non-convex problems are tackled by means of fractional programming coupled with sequential convex optimization, and two low-complexity resource allocation algorithms are designed, which are guaranteed to converge to Karush-Kuhn-Tucker points of the non-convex problems. Numerical results show that the proposed algorithm can efficiently balance between the goals of maximizing the energy efficiency and meeting the QoS constraints. Moreover, it is shown that a small sum-rate reduction allows large energy savings.**

I. Introduction

Currently, the percentage of the global world CO_2 emissions due to the information and communications technology (ICT) is estimated to be 5% [1]. While this may seem a small percentage, it is rapidly increasing, and the situation will escalate in the near future with the advent of 5G cellular networks. Credited sources (e.g., [2], [3]) foresee the number of connected devices to reach 50 billions by 2020. If no countermeasure is taken, the energy demand to operate and serve this massive number of devices will become unmanageable, and the resulting greenhouse gas emissions and electromagnetic pollution will exceed safety thresholds. A promising answer to this issue lies in optimizing the energy efficiency (EE) of the system, i.e., in minimizing the amount of energy required to transmit data. Moreover, EE is of paramount importance for operators (e.g., to save on electricity bills), and for end-users (e.g., to prolong the lifetime of batteries).

This paper aims to develop a framework for energy-efficient resource allocation in cellular networks employing candidate 5G transmission technologies [4], [5]. Specifically, we consider the uplink transmission in a heterogeneous network consisting of a cluster of coordinated access points, possibly with the presence of an amplify-and-forward (AF) relay [6]. Each access point is equipped with multiple antennas, and orthogonal frequency division multiple access (OFDMA) with universal frequency reuse is assumed. A non-exhaustive list of recent works dealing with energy-efficient resource allocation with coordinating base stations includes [7]–[11]. Fractional programming theory is employed and different performance metrics are considered. In [7], the global energy efficiency (GEE) of the cluster is considered, whereas the sum and the minimum of the individual EEs are maximized in [8] and [9], respectively. In [10], the GEE as well as the sum and the product of the individual EEs are optimized. In [11], fractional programming is used to design beamforming techniques for maximizing the weighted sum EEs in single-stream multiple-input multiple-output (MIMO) systems.

We extend the aforementioned works in different directions.

1) A more general expression of the signal-to-interference-plus-noise-ratio (SINR) is considered so as to extend the developed framework to those 5G technologies that cannot be tackled with existing solutions. This includes heterogeneous, multi-hop networks [10], practical, hardware-impaired massive MIMO systems [12], full-duplex systems [13].
2) Within the above setup, we consider two different EE problems in which the maximization is performed with respect to subcarrier assignment and transmit power allocation. The first is focused on the network GEE while the second aims at maximizing the users' minimum EE so as to guarantee some fairness. Unlike most previous contributions, our formulation also considers quality-of-service (QoS) constraints in terms of minimum rate requirements.
3) Both formulations are shown to lead to non-convex fractional problems. We propose the joint use of fractional programming [14] and sequential convex optimization [15] to design low-complexity resource allocation algorithms that are guaranteed to converge to a point fulfilling the Karush-Kuhn-Tucker (KKT) conditions, by only solving a sequence of convex problems.

The work of Alessio Zappone has received funding from the German Research Foundation (DFG) project CEMRIN, under grant ZA 747/1-2. The research leading to these results has also received funding from the European Union's FP7 under REA Grant agreement PIEF-GA-2012-330731 Dense4Green, and is also supported by NEWCOM# (Grant agreement no. 318306). The work of Eduard Jorswieck is supported by the German Research Foundation (DFG) in the Collaborative Research Center 912 "Highly Adaptive Energy-Efficient Computing".

The rest of the paper is organized as follows. Section II introduces the signal model. Section III formulates the two EE problems and provides algorithms to solve both. Numerical results are shown in Section IV to validate the performance of the proposed algorithms, while some conclusions are drawn in Section V.

II. Signal model

Consider a wireless network with K transmitters, S receivers, and N available resource blocks. Assume that the SINR experienced by transmitter k at its intended receiver on resource block n takes the following general form:

$$\gamma_{k,n} = \frac{\alpha_{k,n} p_{k,n}}{1 + \phi_{k,n} p_{k,n} + \sum_{j \neq k} \omega_{j,n} p_{j,n}} \tag{1}$$

where $p_{k,n}$ is the k-th user's transmit power over resource block n, and $\alpha_{k,n}$, $\phi_{k,n}$, $\omega_{k,n}$ are positive quantities that do not depend on the users' transmit powers, but only on system parameters and propagation channels (that are assumed without loss of generality to be normalized to the noise power). By simply specializing these coefficients, the expression in (1) may account for the SINR experienced by users in several relevant instances of communication systems. More details on this will be provided in Section II-B and II-C.

A. Energy-efficiency formulation

Given (1), the bit/Joule EE of the k-th user is defined as [16]

$$\eta_k = \frac{\sum_{n=1}^{N} \log_2(1 + \gamma_{k,n})}{p_{c,k} + \sum_{n=1}^{N} p_{k,n}} \tag{2}$$

with $p_{c,k}$ being the circuit power dissipated to operate the k-th transmitter. While (2) is a user-centric EE function, two more relevant performance metrics (from a system perspective) are the network GEE ψ given by

$$\psi = \frac{\sum_{k=1}^{K} \sum_{n=1}^{N} \log_2 (1 + \gamma_{k,n})}{p_c + \sum_{k=1}^{K} \sum_{n=1}^{N} p_{k,n}} \tag{3}$$

with p_c denoting the total circuit power dissipated in the network, and the weighted minimum of the EEs defined as

$$\underline{\eta} = \min_{k=1,\ldots,K} w_k \eta_k. \tag{4}$$

The above two metrics represent the two extreme points in the trade-off between global performance and fairness. In particular, ψ can be seen as the benefit-cost ratio of the system, as it is defined as the ratio between the sum achievable rate and the total consumed power. However, it does not directly depend on the users' EEs, and therefore it does not allow to tune the individual EE according to user needs as might be the case in heterogeneous networks. On the other hand, maximizing the (weighted) minimum of the EEs allows us to achieve a more fair resource allocation policy. In particular, it is known that maximizing (4) yields a Pareto-efficient point where each quantity $w_k \eta_k$ is the same for all $k = 1, \ldots, K$. The whole Pareto-boundary can be simply achieved by varying the weights $\{w_k\}$. However, this usually comes at the price of a performance loss in terms of the benefit-cost ratio of the system.

Both problems will be addressed in this work, by providing algorithms that maximize either (3) or (4) when power and QoS constraints are imposed. In these circumstances, the GEE maximization problem can be mathematically formulated as:

$$\begin{aligned} \max_{\{p_{k,n} \geq 0\}_{k,n}} \quad & \psi \\ \text{s.t.} \quad & \sum_{n=1}^{N} p_{k,n} \leq \overline{P}_k \ \forall k \\ & \sum_{n=1}^{N} \log_2 (1 + \gamma_{k,n}) \geq \underline{\theta}_k \ \forall k \end{aligned} \tag{5}$$

whereas the minimum-EE maximization problem can be written as:

$$\begin{aligned} \max_{\{p_{k,n} \geq 0\}_{k,n}} \quad & \underline{\eta} \\ \text{s.t.} \quad & \sum_{n=1}^{N} p_{k,n} \leq \overline{P}_k \ \forall k \\ & \sum_{n=1}^{N} \log_2 (1 + \gamma_{k,n}) \geq \underline{\theta}_k \ \forall k \end{aligned} \tag{6}$$

where $\overline{P}_k$ and $\underline{\theta}_k$ denote user k's maximum power and minimum achievable throughput, respectively.

Due to the fractional nature of the performance metrics in both formulations, we will make use of fractional programming, which is a branch of optimization theory concerned with the properties and optimization tools of fractional functions [14]. For completeness, a brief background on fractional programming is provided in the Appendix with the attempt to describe the main tools and results that will be useful to solve the problems at hand.

Before turning to the analysis of the optimization problems, we should provide more details on the SINR expression in (1). As already mentioned, it is general enough to lend itself to several relevant applications for different expressions of the coefficients $\alpha_{k,n}$, $\phi_{k,n}$, $\omega_{k,n}$. A few illustrative examples are given in the sequel.

B. Two-hop multi-cell multiple-antenna OFDMA network

Consider the uplink channel of a two-hop multi-cell OFDMA network, with K single-antenna users, N available subcarriers, S base stations, each equipped with M antennas, where the users communicate to the base stations via a single-antenna AF relay . Denoting by $h_{k,n}$ the channel from user k to the relay on subcarrier n, by $w_{r,n}$ the relay thermal noise, the signal received at the relay on subcarrier n is $r_{k,n} = \sqrt{p_{k,n}} h_{k,n} b_{k,n} + \sum_{j \neq k} \sqrt{p_{j,n}} h_{j,n} b_{j,n} + w_{r,n}$. The power of the received signal on subcarrier n can be computed as

$$P_{t,n} = \sum_{j=1}^{K} p_{j,n} |h_{j,n}|^2 + \sigma_{r,n}^2 \tag{7}$$

with $\sigma_{r,n}^2$ being the relay noise power. In order to avoid amplifier saturation, the received signal needs to be normalized

by its received power $P_{t,n}$ before it can be amplified. Then, the resulting signal is amplified by $\sqrt{p_{r,n}}$, with $p_{r,n}$ denoting the available power on subcarrier n. The signal received at base station $a(k)$ is

$$\boldsymbol{y}_{k,a(k),n} = \boldsymbol{w}_{a(k),n} + \sqrt{\frac{p_{k,n}p_{r,n}}{P_{t,n}}} h_{k,n} b_{k,n} \boldsymbol{g}_{a(k),n} + \sum_{j\neq k} \sqrt{\frac{p_{j,n}p_{r,n}}{P_{t,n}}} h_{j,n} b_{j,n} \boldsymbol{g}_{a(k),n} + \sqrt{\frac{p_{r,n}}{P_{t,n}}} w_{r,n} \boldsymbol{g}_{a(k),n} \,, \tag{8}$$

where $\boldsymbol{g}_{a(k),n}$ is the M-dimensional channel vector from the relay to user k's base station on subcarrier n. After linear reception by the filter $\boldsymbol{c}_{a(k),n}$ and upon plugging (7) into (8) the resulting SINR takes the form (1) with $\alpha_{k,n} = p_{r,n}|h_{k,n}|^2 |\boldsymbol{c}_{a(k),n}^H \boldsymbol{g}_{a(k),k}|^2 / z_{k,n}$, $\phi_{k,n} = \sigma_{k,n}^2 |h_{k,n}|^2 / z_{k,n}$, $\omega_{k,n} = (p_{r,n}|\boldsymbol{c}_{a(k),n}^H \boldsymbol{g}_{a(k),k}|^2 + \sigma_{k,n}^2 \|\boldsymbol{c}_{a(k),n}\|^2)|h_{j,n}|^2 / z_{k,n}$, and $z_{k,n} = \sigma_{r,n}^2 (p_{r,n}|\boldsymbol{c}_{a(k),n}^H \boldsymbol{g}_{a(k),k}|^2 + \sigma_{k,n}^2 \|\boldsymbol{c}_{a(k),n}\|^2)$.

C. Hardware-impaired massive MIMO network

Consider a single-cell massive MIMO system wherein K transmitters communicate with an access point equipped with $M >> K$ antennas where low-power and low-cost circuitry is deployed[1]. This translates into hardware impairments that might largely impact the system performance especially for small values of SNR, if not accounted for. A possible received signal model for hardware-impaired massive MIMO links is provided in [12], where it is shown that the overall effect of hardware impairments can be modeled as an additional Gaussian interference term, whose power is proportional to the useful signal power. In particular, the signal vector $\boldsymbol{r}$ received at the access point can be written as:

$$\boldsymbol{r} = \sum_{k=1}^{K} \boldsymbol{h}_k x_k + \boldsymbol{w} + \boldsymbol{\eta} \,, \tag{9}$$

wherein $\boldsymbol{h}_k$ and x_k are the k-th user's channel to the access point and information symbol with power p_k, $\boldsymbol{w}$ is the white thermal noise at the receiver with covariance matrix $\sigma^2 \boldsymbol{I}_M$, while $\boldsymbol{\eta}$ is the hardware-impairment term, also modeled as a zero-mean Gaussian random vector, but with covariance matrix $\tau \sum_{j=1}^{K} p_j \boldsymbol{D}_j$, wherein $\boldsymbol{D}_j = \mathrm{diag}(\{|h_j(m)|^2\}_{m=1}^M)$ and τ is a proportionality coefficient related to the particular employed hardware [12]. The resulting SINR enjoyed by the k-th user, after linear reception by the filter $\boldsymbol{c}_k$, is thus written as in (1), with $\alpha_k = |\boldsymbol{c}_k^H \boldsymbol{h}_k|^2 / (\sigma^2 \|\boldsymbol{c}_k\|^2)$, $\phi_k = \tau \boldsymbol{c}_k^H \boldsymbol{D}_k \boldsymbol{c}_k / (\sigma^2 \|\boldsymbol{c}_k\|^2)$, and $\omega_k = (\sum_{j\neq k} |\boldsymbol{c}_k^H \boldsymbol{h}_j|^2 + \tau \boldsymbol{c}_k^H \boldsymbol{D}_j \boldsymbol{c}_k) / (\sigma^2 \|\boldsymbol{c}_k\|^2)$.

There exist other applications in which the SINR takes the form in (1). Among them, we mention ultra-wideband systems [17] and transmissions affected by inter-symbol interference, as well as other key technologies for 5G networks such as full-duplex devices [13] and two-hop MIMO systems [10].

[1]This is the typical operation condition envisioned for massive MIMO systems, due to the large amount of required hardware.

III. Proposed solution

As described in the Appendix, fractional programming provides efficient tools to maximize ratios in which the numerator is a concave function and the denominator is a convex one. Unfortunately, neither (3) nor (4) have a concave function at the numerator. For this reason, aiming at providing computationally efficient optimization algorithms, we will leverage the tool of sequential convex programming [15], [18], [19]. Roughly speaking, the idea is to solve (5) and (6) by solving a sequence of approximate easier problems whose solutions converge to a (possibly local) solution of the original problem. Formally speaking, assume $\mathcal{P}$ is a maximization problem with objective $f_0(\boldsymbol{x})$ and constraint functions $\{f_i(\boldsymbol{x})\}_{i=1}^I$. Let us then consider a sequence of Problems $\{\mathcal{P}_j\}_j$ with objective $\tilde{f}_{0,j}(\boldsymbol{x})$, constraints $\{\tilde{f}_{i,j}(\boldsymbol{x})\}_{i=1}^I$, and optimal solution $\tilde{\boldsymbol{x}}_j$. We require that, for any j and $i = 0, \ldots, I$, $\tilde{f}_{i,j}(\boldsymbol{x})$ enjoys the following properties.

P1) $\tilde{f}_{i,j}(\boldsymbol{x}) \leq f_i(\boldsymbol{x})$, for all $\boldsymbol{x}$
P2) $\tilde{f}_{i,j}(\tilde{\boldsymbol{x}}_{j-1}) = f_i(\tilde{\boldsymbol{x}}_{j-1})$
P3) $\nabla \tilde{f}_{i,j}(\tilde{\boldsymbol{x}}_{j-1}) = \nabla f_i(\tilde{\boldsymbol{x}}_{j-1})$

As we will see, these properties ensure that the sequence $\{\tilde{\boldsymbol{x}}_j\}_j$ of the solutions to the approximated problems $\{\mathcal{P}_j\}_j$ converges to a point $\tilde{\boldsymbol{x}}$, which fulfills the KKT optimality conditions of the original problem $\mathcal{P}$. This gives us a way to find local solutions to (5) and (6), provided we can find suitable approximations that fulfill properties P1-P3. The next two subsections show how to accomplish this for (3) and (4), respectively.

A. GEE maximization (5)

Let us consider the following bound of the logarithm function [20]. Specifically, for any $\gamma, \tilde{\gamma} \geq 0$ we have

$$\log_2(1+\gamma) \geq a \log_2 \gamma + b \,, \tag{10}$$

$$\text{with} \quad a = \tfrac{\tilde{\gamma}}{1+\tilde{\gamma}} \quad \text{and} \quad b = \log_2(1+\tilde{\gamma}) - \tfrac{\tilde{\gamma}}{1+\tilde{\gamma}} \log_2 \tilde{\gamma} \tag{11}$$

The right-hand side (RHS) and left-hand side (LHS) of (10) are equal at $\gamma = \tilde{\gamma}$, and the same holds for their derivatives with respect to γ evaluated at $\gamma = \tilde{\gamma}$. Then we have

$$\begin{aligned} \psi &\geq \frac{\sum_{k=1}^K \sum_{n=1}^N [a_{k,n} \log_2(\gamma_{k,n}) + b_{k,n}]}{p_c + \sum_{k=1}^K \sum_{n=1}^N p_{k,n}} \\ &= \frac{\sum_{k=1}^K \sum_{n=1}^N [b_{k,n} + a_{k,n} \log_2(\alpha_{k,n} p_{k,n})]}{p_c + \sum_{k=1}^K \sum_{n=1}^N p_{k,n}} \\ &\quad - \frac{\sum_{k=1}^K \sum_{n=1}^N \left[a_{k,n} \log_2\left(1 + \phi_{k,n} p_{k,n} + \sum_{j\neq k} \omega_{j,n} p_{j,n}\right)\right]}{p_c + \sum_{k=1}^K \sum_{n=1}^N p_{k,n}} \\ &= \tilde{\psi} \end{aligned} \tag{12}$$

Next, using the variable change $p_{k,n} = 2^{q_{k,n}}$, for all $k = 1, \ldots, K$ and $n = 1, \ldots, N$, (12) becomes

$$\tilde{\psi} = \frac{\sum_{k=1}^{K}\sum_{n=1}^{N}\left[b_{k,n} + a_{k,n}\log_2\left(\alpha_{k,n}\right) + a_{k,n}q_{k,n}\right]}{p_c + \sum_{k=1}^{K}\sum_{n=1}^{N} 2^{q_{k,n}}} - \frac{\sum_{k=1}^{K}\sum_{n=1}^{N}\left[a_{k,n}\log_2\left(1+\phi_{k,n}2^{q_{k,n}}+\sum_{j\neq k}\omega_{j,n}2^{q_{j,n}}\right)\right]}{p_c + \sum_{k=1}^{K}\sum_{n=1}^{N} 2^{q_{k,n}}} \tag{13}$$

A similar approach can be used to approximate the QoS constraint in (5), which yields, for all $k = 1, \ldots, K$,

$$\tilde{R}_k = \sum_{n=1}^{N}\left[b_{k,n} + a_{k,n}\log_2\left(\alpha_{k,n}\right) + a_{k,n}q_{k,n}\right] - \sum_{n=1}^{N}\left[a_{k,n}\log_2\left(1+\phi_{k,n}2^{q_{k,n}}+\sum_{j\neq k}\omega_{j,n}2^{q_{j,n}}\right)\right] \geq \underline{\theta}_k \tag{14}$$

Then, we can approximate (5) as

$$\begin{aligned} \max_{\{q_{k,n}\}_{k,n}} \ & \tilde{\psi} \\ \text{s.t.} \ & \sum_{n=1}^{N} 2^{q_{k,n}} \leq \overline{P}_k \ , \forall k \\ & \tilde{R}_k \geq \underline{\theta}_k \ , \ \forall k \ . \end{aligned} \tag{15}$$

Note that for any fixed $\{a_{k,n}\}_{k,n}$ and $\{b_{k,n}\}_{k,n}$, the numerator and the denominator of (13) are both differentiable, and respectively concave[2] and convex in $\{q_{k,n}\}_{k,n}$, whereas the QoS constraint is concave. Moreover, both the denominator and the numerator are non-negative.[3] As a consequence, Problem (15) is a fractional problem with a pseudo-concave objective function and convex constraints, and therefore can be globally and efficiently solved by means of Dinkelbach's algorithm [16], [21]. Finally, a general resource allocation algorithm can be formulated as Algorithm 1.

Algorithm 1 GEE maximization

Initialize $p_{k,n}$ to feasible power values; Set $j = 0$;
For all $k = 1, \ldots, K$ and $n = 1, \ldots, N$, set $\tilde{x}_{k,n}^{(j)} = \gamma_{k,n}^{(j)}(\{p_{k,n}^{(j)}\}_{k,n})$, and compute $a_{k,n}^{(j)}$ and $b_{k,n}^{(j)}$ as in (11);
repeat
  $j = j + 1$;
  Solve (15) by Dinkelbach's algorithm and set $\{p_{k,n}^{(j)}\}_{k,n}$ as its solution;
  For all $k = 1, \ldots, K$ and $n = 1, \ldots, N$, set $\tilde{x}_{k,n}^{(j)} = \gamma_{k,n}(\{p_{k,n}^{(j)}\}_{k,n})$;
  Update $a_{k,n}^{(j)}$ and $b_{k,n}^{(j)}$ as in (11);
until convergence

Proposition 1: Algorithm 1 monotonically increases the GEE value and converges to a point fulfilling the KKT conditions of the original non-convex problem (5).

[2]The log-sum-exp function is convex.

[3]The numerator is ensured to be non-negative on the feasible set due to the QoS constraints.

Proof: For the sake of brevity, here we give a sketch of the proof. We have already shown that $\psi \geq \tilde{\psi}$ for all $\{p_{k,n}\}_{k,n}$. It is also easy to verify that for any given $\{\tilde{p}_{k,n}\}_{k,n}$, the bound and its derivative can be made tight at $\{\tilde{p}_{k,n}\}_{k,n}$ if $a_{k,n}$ and $b_{k,n}$ are set according to (11), for all $k = 1, \ldots, K$ and $n = 1, \ldots, N$. Then, at the j-th iteration of Algorithm 1 we have

$$\begin{aligned} \psi\left(\{p_{k,n}^{(j)}\}_{k,n}\right) &\geq \tilde{\psi}_j\left(\{p_{k,n}^{(j)}\}_{k,n}\right) \geq \tilde{\psi}_j\left(\{p_{k,n}^{(j-1)}\}_{k,n}\right) \\ &= \psi\left(\{p_{k,n}^{(j-1)}\}_{k,n}\right) \ , \end{aligned} \tag{16}$$

wherein the first inequality follows by virtue of property P1, the second inequality holds because $\{p_{k,n}^{(j)}\}_{k,n}$ is the solution of Problem P_j whose objective is $\tilde{\psi}_j$, and the equality follows from property P2. Finally, by a similar reasoning and exploiting property P3, we can show that upon convergence the KKT conditions of the original problem are fulfilled. ■

Before closing this section, it should be remarked that the proposed framework can be straightforwardly modified to optimize the sum-rate instead of the GEE. Indeed, the system sum-rate is the numerator of the GEE, which can be turned into a concave function by means of the same approach used for GEE. In Section IV, the performance of GEE and sum-rate optimization will be compared.

B. Min-EE maximization (6)

The main challenge of Problem (6) with respect to Problem (5) is that it involves more than a single fractional function. In particular, (6) falls within the framework of generalized fractional programming. In this section we will show how the generalized Dinkelbach's procedure described in the Appendix together with sequential convex optimization, can be successfully applied to tackle (6).

Exploiting the fact that the $\min(\cdot)$ function is increasing, we can use the bound (10) and the substitution $p_{k,n} = 2^{q_{k,n}}$, to lower-bound $\underline{\eta}$ as

$$\underline{\eta} \geq \min_{k=1,\ldots,K} w_k\tilde{\eta}_k = \tilde{\underline{\eta}} \tag{17}$$

where

$$\tilde{\eta}_k = \frac{\sum_{n=1}^{N}\left[b_{k,n} + a_{k,n}\log_2\left(\alpha_{k,n}\right) + a_{k,n}q_{k,n}\right]}{p_{c,k} + \sum_{n=1}^{N} 2^{q_{k,n}}} - \frac{\sum_{n=1}^{N}\left[a_{k,n}\log_2\left(1+\phi_{k,n}2^{q_{k,n}}+\sum_{j\neq k}\omega_{j,n}2^{q_{j,n}}\right)\right]}{p_{c,k} + \sum_{n=1}^{N} 2^{q_{k,n}}}. \tag{18}$$

We see that each ratio in (18) has a concave numerator and a convex denominator. Then, Problem (6) can be approximated by the problem

$$\begin{aligned} \max_{\{q_{k,n}\}_{k,n}} \ & \tilde{\underline{\eta}} \\ \text{s.t.} \ & \sum_{n=1}^{N} 2^{q_{k,n}} \leq \overline{P}_k \ , \forall k \\ & \tilde{R}_k \geq \underline{\theta}_k \ , \ \forall k \ , \end{aligned} \tag{19}$$

with $\tilde{R}_k$ defined in (14), which can be solved by means of the generalized Dinkelbach's algorithm. Finally, we can formulate a resource allocation procedure as Algorithm 2.

Algorithm 2 Min-EE maximization

Initialize $\{p_{k,n}^{(j)}\}_{k,n}$ to feasible power values; Set $j = 0$;
For all $k = 1, \ldots, K$ and $n = 1, \ldots, N$, set $\tilde{x}_{k,n}^{(j)} = \gamma_{k,n}^{(j)}(\{p_{k,n}^{(j)}\}_{k,n})$, and compute $a_{k,n}^{(j)}$ and $b_{k,n}^{(j)}$ as in (11);
repeat
  $j = j + 1$;
  Solve (19) by the generalized Dinkelbach's algorithm and set $\{p_{k,n}^{(j)}\}_{k,n}$ as its solution;
  For all $k = 1, \ldots, K$ and $n = 1, \ldots, N$, set $\tilde{x}_{k,n}^{(j)} = \gamma_{k,n}^{(j)}(\{p_{k,n}^{(j)}\}_{k,n})$;
  Update $a_{k,n}^{(j)}$ and $b_{k,n}^{(j)}$ as in (11);
until convergence

Proposition 2: Algorithm 2 monotonically increases the value of $\underline{\eta}$ and converges.

Proof: We have already shown that $\underline{\eta} \geq \tilde{\eta}$. The rest of the proof follows along the same lines of Proposition 1. However, note that in this case we do not claim convergence to a point fulfilling the KKT conditions of the original problem (6), since the objective of (6) is not differentiable and therefore we can not consider KKT conditions. For this reason, property P3 is not needed in this case. ■

Since for all users the numerator of the EE coincides with the achievable rate, Algorithm 2 can be easily adapted to maximize the minimum of the users' achievable rates, just as Algorithm 1 could be adapted to maximize the system sum-rate.

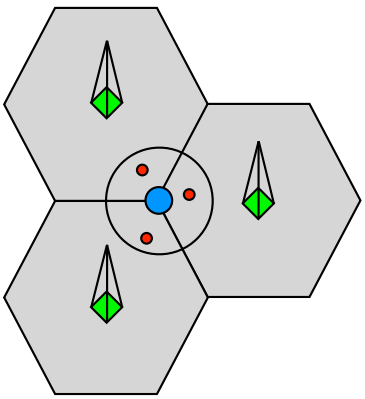

Figure 1. MIMO-OFDMA, two-hop, cellular scheme.

IV. Numerical results

The two-hop multi-cellular OFDMA system shown in Figure 1 is considered for numerical simulations. Three cells coordinate their transmissions over $N = 16$ subcarriers with bandwidth $B = 180\,\text{kHz}$, and are equipped with $M = 3$ antennas each. An AF relay is placed at the intersection of the cells, to serve $K = 3$ cell-edge users, which are placed at a distance from the relay randomly generated in the interval $[100; 300]$ m. All mobiles have the same maximum feasible power $\overline{P}$ and hardware-dissipated power $p_c = 10\,\text{dBmW}$, while the relay has a maximum feasible power of $\overline{P}_r = 0\,\text{dBW}$ on each subcarrier. The noise power at each receiver and at the relay is $\sigma^2 = FB\mathcal{N}_0$, with $F = 3\,\text{dB}$ being the receiver noise figure, and $\mathcal{N}_0 = -174\,\text{dBW/Hz}$ being the noise power spectral density. The minimum rate to be granted to each user is $\underline{\theta}_k = N\theta$, with $\theta = 2.5$ bit/s/Hz being the minimum average rate per sub-carrier. For all $k = 1, \ldots, K$ and all $n = 1, \ldots, N$, the channel from mobile k to the relay on subcarrier n is generated as $h_{k,n} = \sqrt{L_k}\delta_{k,n}$, where $\delta_{k,n}$ is a realization of a zero-mean, unit-variance complex Gaussian random variable accounting for Rayleigh fading, and $L_k = L_0\,(d_0/d_k)^4$, accounts for the path loss, wherein d_k is the user-relay distance, $d_0 = 100\,\text{m}$, and L_0 is the free-space attenuation at d_0 [22]. A similar channel model is used to generate each entry of the vector channel $\boldsymbol{g}_{j,n}$ from the relay to receiver j on subcarrier n. The presented results have been obtained averaging over 10^3 independent system scenarios.

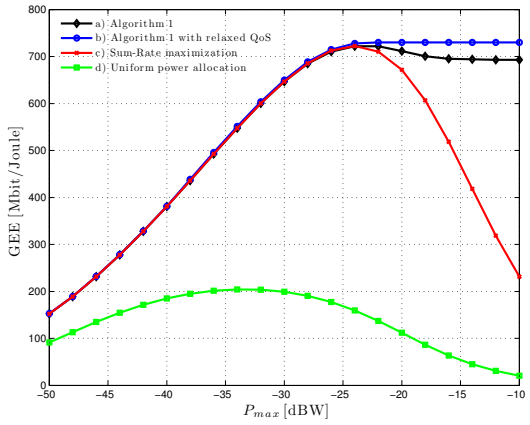

Figure 2. $N = 16$; $K = 3$; $M = 3$. Achieved GEE versus P_{max} for: a) Algorithm 1 for GEE maximization; b) Algorithm 1 for GEE maximization with relaxed QoS constraints; c) sum-rate maximization by adapted Algorithm 1; d) uniform power allocation over each subcarrier.

In Figure 2, the achieved GEE versus $\overline{P}$ is illustrated for the following resource allocation algorithms:

a) Algorithm 1 for GEE maximization.
b) Algorithm 1 for GEE maximization, in which the QoS constraints have been relaxed;
c) Sum-rate maximization by adapting Algorithm 1 as explained in Section III-A;
d) as a baseline scheme, we consider uniform power allocation, i.e. $p_{k,n} = \overline{P}/N$ for all k, n.

It should be mentioned that, for a given minimum rate to be guaranteed, in general the GEE resource allocation problem might turn out to be unfeasible, depending on the particular channel realizations and especially on the maximum feasible power $\overline{P}$. When this happens in Algorithm 1, we relax the constraints and adopt the solution obtained without QoS constraints.

Figure 2 indicates that, when $\overline{P} \leq -24\,\text{dBW}$, allocations a),

b), and c) all result in the same power allocation policy. This shows that, for $\overline{P} \leq -24$ dBW, the sum-rate at the numerator of the GEE is more significant than the corresponding consumed power at the denominator, and maximizing the GEE is equivalent to maximizing its numerator. Instead, the situation changes for higher values of $\overline{P}$. If the GEE is maximized without enforcing any QoS constraint, i.e. allocation b), the achieved GEE remains constant, since a maximum feasible power of $\overline{P} = -24$ dBW already allows to attain the maximum GEE value and further increasing the transmit power would only decrease the achieved GEE. Indeed, we see that the GEE achieved by sum-rate maximization degrades, since sum-rate optimization requires a larger transmit power than GEE maximization. The performance of Algorithm 1, i.e. allocation a), is between these two extremes. For $\overline{P} > -24$ dBW, the GEE decreases a little, because more transmit power is used in order to increase the rate and meet the QoS constraints. However, once the QoS constraints are met, further increasing the transmit power is not necessary and the achieved GEE remains constant for higher $\overline{P}$.

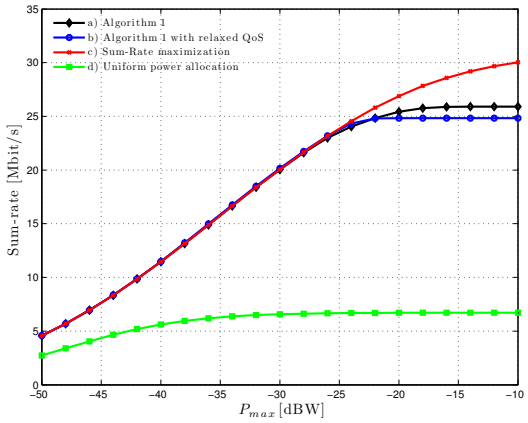

Figure 3. $N = 16; K = 3; M = 3$. Achieved sum-rate versus P_{max} for: a) Algorithm 1 for GEE maximization; b) Algorithm 1 for GEE maximization with relaxed QoS constraints; c) sum-rate maximization by adapted Algorithm 1; d) uniform power allocation over each subcarrier.

In Figure 3 and Figure 4, the same allocations as in Figure 2 are compared, but the shown performance metrics are the achieved sum-rate in bits/s and the average transmit power per user, respectively. Similarly to Figure 2, all allocations perform similarly for $\overline{P} \leq -24$ dBW. For larger values of $\overline{P}$, it is seen that sum-rate optimization requires a larger transmit power, thus achieving a higher sum-rate. Instead, Algorithm 1 requires to increase the transmit power only up to the point when the QoS can be met. Finally, if no QoS is enforced, the transmit power remains constant for $\overline{P} > -24$ dBW, but this also means that the sum-rate does not increase further and remains below that achieved by Algorithm 1. Thus, Algorithm 1 is able to balance between the need for a high rate and of saving power. Moreover, comparing Figure 2-4, we observe that in general a moderate sum-rate reduction allows saving a large amount of transmit power, thus granting a much higher energy efficiency compared to the maximization of the sum-rate.

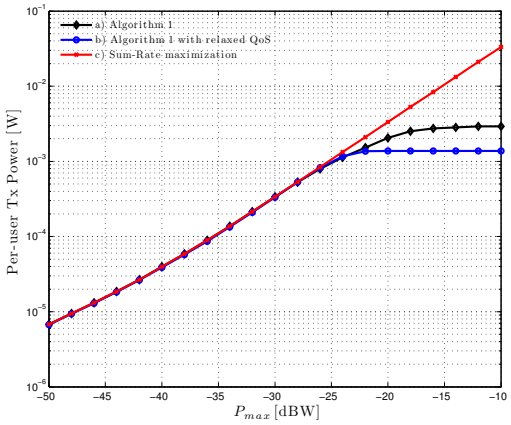

Figure 4. $N = 16; K = 3; M = 3$. Required transmit power versus P_{max} for: a) Algorithm 1 for GEE maximization; b) Algorithm 1 for GEE maximization with relaxed QoS constraints; c) sum-rate maximization by adapted Algorithm 1.

Table I
PROBABILITY OF UNFEASIBILITY $\mathbb{P}_u$ OF ALGORITHM 1 WITH QOS CONSTRAINTS, AND AVERAGE NUMBER OF ITERATIONS N_{it} NEEDED TO REACH CONVERGENCE, VERSUS THE NUMBER OF ACTIVE USERS.

$\overline{P}$ [dBW]	-50	-40	-30	-25	-20	-15	-10
$\mathbb{P}_u$	1	1	0.96	0.63	0.36	0.02	0
N_{it}	6.7	8.3	9.4	11.5	13.6	14.1	14.3

In Table I we report the number of iterations N_{it} required for Algorithm 1 to reach convergence, and the probability $\mathbb{P}_u$ that the GEE maximization problem with QoS is unfeasible, for different values of $\overline{P}$. We can see that Algorithm 1 converges in an affordable number of iterations, which increases as $\overline{P}$ increases. This is expected since increasing $\overline{P}$ increases the dimension of the feasible set of the optimization problems. Moreover, as expected, the GEE maximization problem is likely to be unfeasible for low values of $\overline{P}$, whereas $\mathbb{P}_u$ approaches 0 as $\overline{P}$ increases.

Here we have focused on the performance of Algorithm 1 for GEE maximization. Similar numerical simulations have been carried out with reference to the performance of Algorithm 2 for Min-EE maximization and its comparison with the maximization of the minimum of the users' rate. The results are not shown here due to space constraints, but the curves show a similar general behavior as for Figs. 2-4, providing similar insights as in the remarks made for Algorithm 1.

V. CONCLUSION

We have considered the problem of EE maximization in the uplink of a cooperative multi-point (CoMP) system, considering a more general SINR expression which models important candidate 5G technologies, such as heterogeneous, small-cells networks, hardware-impaired massive MIMO, and full duplex transmission. Fractional programming and sequential convex optimization have been used to tackle the non-convex

problems of GEE and max-min EE optimization, also considering QoS constraints. We proposed two resource allocation algorithms that, despite requiring only the solution of convex problems, are guaranteed to converge to points fulfilling the KKT conditions of the original non-convex problems.

APPENDIX

Definition 1 (Fractional program): Let $\mathcal{C} \subseteq \mathbb{R}^n$ be a convex set, and consider the functions $f : \mathcal{C} \to \mathbb{R}_0^+$ and $g : \mathcal{C} \to \mathbb{R}^+$. A fractional program is the optimization problem

$$\max_{\boldsymbol{x} \in \mathcal{C}} \frac{f(\boldsymbol{x})}{g(\boldsymbol{x})} \tag{20}$$

The following result relates the solution of (20) to the auxiliary function $F(\lambda) = \max_{\boldsymbol{x} \in \mathcal{C}} \{f(\boldsymbol{x}) - \lambda g(\boldsymbol{x})\}$.

Proposition 3 ([21], [23]): An $\boldsymbol{x}^* \in \mathcal{C}$ solves (20) if and only if $\boldsymbol{x}^* = \arg\max_{\boldsymbol{x} \in \mathcal{C}} \{f(\boldsymbol{x}) - \lambda^* g(\boldsymbol{x})\}$, with λ^* being the unique zero of $F(\lambda)$.

This result allows us to solve (20) by finding the zero of $F(\lambda)$. An efficient algorithm to do so is the Dinkelbach's algorithm [21], reported in Algorithm 3 for the reader's convenience.

Algorithm 3 Dinkelbach's algorithm

Set $\varepsilon > 0$; $\lambda = 0$;
repeat
  $\boldsymbol{x}^* = \arg\max_{\boldsymbol{x} \in \mathcal{C}} \{f(\boldsymbol{x}) - \lambda g(\boldsymbol{x})\}$
  $F = f(\boldsymbol{x}^*) - \lambda g(\boldsymbol{x}^*)$;
  $\lambda = f(\boldsymbol{x}^*) / g(\boldsymbol{x}^*)$;
until $F \leq \varepsilon$

If $f(\boldsymbol{x})$ and $g(\boldsymbol{x})$ are concave and convex, respectively, then the Dinkelbach's algorithm requires to solve one convex problem in each iteration. Moreover, the convergence rate of Dinkelbach's algorithm is known to be super-linear [21].

A considerable extension of (20) is to consider more than a single fractional function. The case we are interested in here, is the maximization of the minimum of a set of ratios $f_i(\boldsymbol{x})/g_i(\boldsymbol{x})$, with $i = 1, \ldots, I$. This problem is usually referred to as *generalized* fractional programming, and has been first studied in [24]. There, an optimization procedure is provided, based on a modification of Dinkelbach's algorithm. Specifically, the auxiliary function to be considered is $F(\lambda) = \min_{1 \leq i \leq I} \{f_i(\boldsymbol{x}) - \lambda g_i(\boldsymbol{x})\}$, and the algorithm works as shown in Algorithm 4.

Algorithm 4 Generalized Dinkelbach's algorithm

Set $\varepsilon > 0$; $\lambda = 0$;
repeat
  $\boldsymbol{x}^* = \arg\max_{\boldsymbol{x} \in \mathcal{C}} \{\min_{1 \leq i \leq I} [f_i(\boldsymbol{x}) - \lambda g_i(\boldsymbol{x})]\}$;
  $F = \min_{1 \leq i \leq I} \{f_i(\boldsymbol{x}^*) - \lambda g_i(\boldsymbol{x}^*)\}$;
  $\lambda = \min_{1 \leq i \leq I} f_i(\boldsymbol{x}^*) / g_i(\boldsymbol{x}^*)$;
until $F < \varepsilon$

Similarly to the original procedure described by Dinkelbach, if each ratio has a concave numerator and a convex denominator, then we can solve the generalized fractional problem by solving a sequence of convex problems.[4] In this case the convergence rate is linear [24].

[4]Recall that the minimum of concave functions is also concave.

REFERENCES

[1] A. Fehske, J. Malmodin, G. Biczók, and G. Fettweis, "The global footprint of mobile communications: The ecological and economic perspective," *IEEE Commun. Mag.*, pp. 55–62, Aug. 2011.

[2] Ericsson White Paper, "More than 50 billion connected devices," Ericsson, Tech. Rep. 284 23-3149 Uen, Feb. 2011.

[3] UMTS Forum, "Mobile traffic forecasts 2010-2020 report," UMTS Forum, Tech. Rep. 44, Jan. 2011.

[4] J. Hoydis, M. Kobayashi, and M. Debbah, "Green small-cell networks," *IEEE Veh. Technol. Mag.*, vol. 6, no. 1, pp. 37–43, Mar. 2011.

[5] J. G. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[6] L. Sanguinetti, A. D'Amico, and Y. Rong, "A tutorial on the optimization of amplify-and-forward MIMO relay systems," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 8, pp. 1331–1346, Sept. 2012.

[7] D. W. K. Ng, E. S. 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.

[8] S. He, Y. Huang, S. Jin, and L. Yang, "Coordinated beamforming for energy efficient transmission in multicell multiuser systems," *IEEE Trans. Commun.*, vol. 61, no. 12, pp. 4961–4971, Dec. 2013.

[9] B. Du, C. Pan, W. Zhang, and M. Chen, "Distributed energy-efficient power optimization for CoMP systems with max-min fairness," *IEEE Commun. Lett.*, vol. 18, no. 6, pp. 999–1002, Jun. 2014.

[10] A. Zappone, Z. Chong, E. Jorswieck, and S. Buzzi, "Energy-aware competitive power control in relay-assisted interference wireless networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 4, pp. 1860–1871, Apr. 2013.

[11] 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.

[12] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Trans. Information Theory*, 2014, submitted. [Online]. Available: http://arxiv.org/abs/1307.2584

[13] B. Bangerter, S. Talwar, R. Arefi, and K. Stewart, "Networks and devices for the 5G era," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 90–96, Feb. 2014.

[14] S. Schaible, "Fractional programming," *Zeitschrift für Operations Research*, vol. 27, no. 1, pp. 39–54, 1983.

[15] 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, July-Aug. 1978.

[16] 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, Nov. 2012.

[17] G. Bacci, M. Luise, H. V. Poor, and A. M. Tulino, "Energy efficient power control in impulse radio UWB wireless networks," *IEEE J. Sel. Topics. Signal Process.*, vol. 1, no. 3, pp. 508–520, Oct. 2007.

[18] M. Chiang, C. Wei, D. P. Palomar, D. O'Neill, and D. Julian, "Power control by geometric programming," *IEEE Trans. Wireless Commun.*, vol. 6, no. 7, pp. 2640–2651, Jul. 2007.

[19] 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.*, 2014, to appear.

[20] J. Papandriopoulos and J. Evans, "Low-complexity distributed algorithms for spectrum balancing in multi-user DSL networks," in *Proc. IEEE Intl. Conf. Commun.*, Istanbul, Turkey, Jun. 2006, pp. 3270–3275.

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

[22] A. Goldmith, *Wireless Communications*. Cambridge, UK: Cambridge Univ. Press, 2005.

[23] R. Jagannathan, "On some properties of programming problems in parametric form pertaining to fractional programming," *Management Science*, vol. 12, no. 7, pp. 609–615, Mar. 1966.

[24] J.-P. Crouzeix and J. A. Ferland, "Algorithms for generalized fractional programming," *Mathematical Programming*, vol. 52, no. 1-3, pp. 191–207, 1991.