PEAK LOAD MINIMIZATION IN LOAD COUPLED INTERFERENCE NETWORKS

R. L. G. Cavalcante and S. Stańczak

Fraunhofer Heinrich Hertz Institute and Technical University of Berlin

ABSTRACT

We propose a novel power control algorithm for the minimization of the peak load in the widely used load coupled network model, which is an abstract model able to capture the behavior of current and possibly future wireless networks. We first prove that the solution to the optimization problem we pose here requires all base stations with the same load. This necessary condition for optimality gives rise to a solver based on a bisection algorithm that requires an oracle able to answer whether a probed load is greater than the optimal value. By exploiting known properties of concave mappings, we devise an iterative oracle that, with a very mild assumption, provably gives the correct answer with a finite number of iterations. Simulations in an ultra-dense network show that the proposed algorithm can decrease the peak load by around 40% when compared to the peak load induced by the common approach of fixing the power of every base station to the maximum value.

***Index Terms*—** Network utility optimization, power control, radio resource management, nonlinear systems

1. INTRODUCTION

Fifth-generation (5G) networks are envisioned to use multiple access schemes that divide the time-frequency grid into basic units called resource blocks [1], which are typically allocated to users based on different optimization criteria such as total throughput maximization, proportional fairness, and others. This approach is similar to that currently used in commercial networks based on the orthogonal frequency-division multiple access (OFDMA) technology. Therefore, we can reasonably expect that radio resource management (RRM) algorithms for future 5G networks to be strongly based on those proposed for OFDMA networks. This observation partially explains the growing body of literature on RRM algorithms for OFDMA systems. Unfortunately, in OFDMA networks, many problems addressed by RRM algorithms are known to be huge instances of NP-hard problems [2], so interference models able to capture the behavior of OFDMA-like networks, while giving rise to tractable mathematical problems, have been the focus of many recent studies.

In particular, algorithms based on the widely used load coupled interference model described in [3–9], which is the model considered in this study, have successfully addressed many network optimization tasks, including data offloading [7], load balancing [6], antenna tilt optimization [9], energy savings [4,5,8,10], and utility optimization [11], to cite a few. One of the main advantages of these algorithms is that they focus on the long-term optimization of wireless resources (e.g., power, rates, antenna tilts, etc.) by considering only the average resource block usage (i.e., the load) at the base stations. The influence of short-term mechanisms for the assignment of resource blocks to users is simplified, and as a result the dimensionality of the optimization problem is kept at reasonable levels.

With the assumptions of the load coupled model, a recent study [8] (see also [5]) has shown that the minimum sum (transmit) power able to support the rates demanded by users induces base stations transmitting at full load. Despite this good theoretical property of fully loaded networks, system engineers may prefer to leave the load at low levels for various reasons, including avoiding problems caused by the vagaries of the requested rates, or being able to accommodate new users without overly complex handover schemes.

Against this background, in this study we start by posing an optimization problem that has the objective of minimizing the peak load observed in the network by means of power control with fixed rate requirements. We show a simple condition that guarantees the existence of a unique solution to this min-max optimization problem. In addition, we prove that the solution is characterized by two main properties: (i) the load is the same at every base station and (ii) at least one base station transmits at full power. In particular, the former property gives rise to a simple solver based on a bisection algorithm that requires an oracle able to answer whether a given load is greater than the optimal value. We exploit known properties of concave mappings to develop an oracle that gives the correct answer with a finite number of iterations, unless the probed load is *exactly* the optimal load, which is unlikely to happen in practice. Nevertheless, even in this unlikely case, the oracle has a stopping criterion with a strong theoretical justification. We evaluate the resulting algorithm for load minimization in an ultra-dense network mimicking the stadium scenario proposed by the METIS project [12], a large European project involving key players in the wireless industry.

As mentioned above, the proposed algorithm is related to those in [5,8], but it differs in the following main aspect. One of the main applications of the algorithms in [5,8] is to compute the power allocation inducing the *maximum* load at every base station in order to save energy. In this sense, the load at base stations is not an optimization variable as considered here. The load is a given parameter that takes its maximum value. Algorithms for load optimization in load coupled networks have also been proposed in [6], but that study focuses on the user-base station assignment mechanisms, and no algorithm for power control is considered.

2. SYSTEM MODEL

We first establish some of the less common notation and definitions used in this paper. We denote by $\mathbb{R}_+$ and $\mathbb{R}_{++}$ the sets of non-negative and positive reals, respectively. For $(\boldsymbol{x},\boldsymbol{y}) \in \mathbb{R}^M \times \mathbb{R}^M$, vector inequalities such as $\boldsymbol{x} \geq \boldsymbol{y}$ should be understood as coordinate-wise inequalities. A mapping $T : \mathbb{R}^M_+ \to \mathbb{R}^M_{++}$ is said to be a positive concave mapping if, for every $(\boldsymbol{x},\boldsymbol{y}) \in \mathbb{R}^M_+ \times \mathbb{R}^M_+$ and every $\alpha \in\ ]0,1[$, we have $T(\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y}) \geq \alpha T(\boldsymbol{x}) + (1-\alpha)T(\boldsymbol{y})$.

We now briefly review the downlink load coupled network model studied in [4–9]. Given a unit of time, we divide the time and frequency grid into $K \in \mathbb{N}$ units called resource blocks. Users belonging to the same base station do not share resource blocks, but different base stations can transmit data at the same resource block. This transmission mechanism causes the well known phenomenon of intercell interference. The sets $\mathcal{N} := \{1, \ldots, N\}$ and $\mathcal{M} := \{1, \ldots, M\}$ represent, respectively, the set of N users and M base stations in the network. The set $\mathcal{N}_i \subset \mathcal{N}$, assumed nonempty, denotes the set of users connected to base station $i \in \mathcal{M}$. The pathloss between base station $i \in \mathcal{M}$ and user $j \in \mathcal{N}$ is given by $g_{i,j} \in \mathbb{R}_{++}$. The power vector and the load vector are given by, respectively, $\boldsymbol{p} = [p_1, \ldots, p_M] \in \mathbb{R}^M_{++}$ and $\boldsymbol{\rho} = [\rho_1, \ldots, \rho_M] \in \mathbb{R}^M_{++}$, where the ith coordinate of these vectors correspond to the power per resource block or the load of base station $i \in \mathcal{M}$. In the load coupled model, load is the long-term fraction of resource blocks used for data transmission. This model assumes uniform transmit power per resource block, and it also assumes that all resource blocks experience the same (long-term) pathloss. Therefore, the achievable rate of each resource block used for the link connecting base station $i \in \mathcal{M}$ to user $j \in \mathcal{N}$ is given by [3–9]:

$$\omega_{i,j}(\boldsymbol{\rho}, \boldsymbol{p}) = B \log_2 \left(1 + \frac{p_i g_{i,j}}{\sum_{k \in \mathcal{M}\backslash\{i\}} \rho_k p_k g_{k,j} + \sigma^2}\right),$$

where $\sigma^2 \in \mathbb{R}_{++}$ is the noise per resource block and $B \in \mathbb{R}_{++}$ is the bandwidth of each resource block. Denoting by $d_j \in \mathbb{R}_{++}$ the data rate requested by user $j \in \mathcal{N}$, for a given power allocation $\boldsymbol{p} \in \mathbb{R}^M_{++}$, we obtain the load at the base stations by solving the following system of nonlinear equations [3, 4, 9, 13]:

$$\begin{aligned} \rho_1 &= f_1(\boldsymbol{\rho}, \boldsymbol{p}) \\ &\vdots \\ \rho_M &= f_M(\boldsymbol{\rho}, \boldsymbol{p}) \end{aligned} \tag{1}$$

where each $f_i : \mathbb{R}^M_{+} \times \mathbb{R}^M_{++} \to \mathbb{R}_{++}, i \in \mathcal{M}$, is the continuous function of two vectors given by

$$f_i(\boldsymbol{\rho}, \boldsymbol{p}) := \sum_{j \in \mathcal{N}_i} \frac{d_j}{K \omega_{i,j}(\boldsymbol{\rho}, \boldsymbol{p})}. \tag{2}$$

Intuitively, each term of the sum in (2) is the fraction of resource blocks that user j requests from base station i to achieve the rate d_j. Given $\boldsymbol{p} \in \mathbb{R}^M_{++}$, the load $\boldsymbol{\rho}^\star \in \mathbb{R}^M_{++}$ solving the system in (1) (if a solution exists) is the fixed point of the positive concave mapping given by $T_{\boldsymbol{p}} : \mathbb{R}^M_{+} \to \mathbb{R}^M_{++} : \boldsymbol{\rho} \mapsto [f_1(\boldsymbol{\rho}, \boldsymbol{p}), \ldots, f_M(\boldsymbol{\rho}, \boldsymbol{p})]$ [4, 9, 14]; i.e., $\boldsymbol{\rho}^\star \in \mathrm{Fix}(T_{\boldsymbol{p}}) := \{\boldsymbol{\rho} \in \mathbb{R}^M_{+} \mid \boldsymbol{\rho} = T_{\boldsymbol{p}}(\boldsymbol{\rho})\}$. We recall that the set $\mathrm{Fix}(T_{\boldsymbol{p}})$ is a singleton if not empty [4]. [1]

Instead of computing the load for a given power allocation, the algorithm for peak load reduction proposed in this study requires an efficient method to compute the power allocation inducing a given load. As proved in [5, Proposition 1], the power allocation $\boldsymbol{p}^\star \in \mathbb{R}^M_{++}$ inducing the load $\boldsymbol{\rho} \in \mathbb{R}^M_{++}$ is the fixed point of the positive concave mapping given by $P_{\boldsymbol{\rho}} : \mathbb{R}^M_{+} \to \mathbb{R}^M_{++} : \boldsymbol{p} \mapsto [P_{\boldsymbol{\rho},1}(\boldsymbol{p}), \ldots, P_{\boldsymbol{\rho},M}(\boldsymbol{p})]$, where

$$P_{\boldsymbol{\rho},i}(\boldsymbol{p}) := \begin{cases} \dfrac{p_i}{\rho_i} \sum_{j \in \mathcal{N}_i} \dfrac{d_j}{K \omega_{i,j}(\boldsymbol{\rho}, \boldsymbol{p})}, & \text{if } p_i \neq 0 \\ \displaystyle\sum_{j \in \mathcal{N}_i} \frac{d_j \ln 2}{K B g_{i,j} \rho_i} \left(\sum_{k \in \mathcal{M}\backslash\{i\}} \rho_k p_k g_{k,j} + \sigma^2 \right), & \text{otherwise;} \end{cases}$$

i.e., $\boldsymbol{p}^\star \in \mathrm{Fix}(P_{\boldsymbol{\rho}}) := \{\boldsymbol{p} \in \mathbb{R}^M_{+} \mid \boldsymbol{p} = P_{\boldsymbol{\rho}}(\boldsymbol{p})\}$. We note that the set $\mathrm{Fix}(P_{\boldsymbol{\rho}})$ is also a singleton is not empty [5, 14]. Furthermore, from the definitions of the mappings, we verify that $\boldsymbol{\rho}^\star \in \mathrm{Fix}(T_{\boldsymbol{p}^\star})$ if and only if $\boldsymbol{p}^\star \in \mathrm{Fix}(P_{\boldsymbol{\rho}^\star})$.

We end this section with two technical results that are used in the proofs of the main results in the next section:

Fact 1. [8, Theorem 2] (see [5] for an alternative proof) Assume that $\boldsymbol{\rho}' \in \mathrm{Fix}(T_{\boldsymbol{p}'}) \neq \emptyset$ for some $\boldsymbol{p}' \in \mathbb{R}^M_{++}$. Then, for every $\boldsymbol{\rho}'' \geq \boldsymbol{\rho}'$ with $\boldsymbol{\rho}' \neq \boldsymbol{\rho}''$, the mapping $P_{\boldsymbol{\rho}''}$ has a unique fixed point $\boldsymbol{p}'' \in \mathbb{R}^M_{++}$ satisfying $\boldsymbol{p}'' < \boldsymbol{p}'$.

Fact 2. [7] If there exists at least one $\boldsymbol{p} \in \mathbb{R}^M_{++}$ for which $\mathrm{Fix}(T_{\boldsymbol{p}}) \neq \emptyset$, then $\mathrm{Fix}(T_{\boldsymbol{p}}) \neq \emptyset$ for every $\boldsymbol{p} \in \mathbb{R}^M_{++}$.

3. THE PROPOSED ALGORITHM

The objective of the proposed algorithm is to obtain the power allocation minimizing the maximum observed load in the network (without any changes in the rate demands or user-base station assignments). Formally, the optimization problem is given by:

Problem 1.

$$\begin{array}{ll} \min._{(\boldsymbol{\rho}, \boldsymbol{p}) \in \mathbb{R}^M_{++} \times \mathbb{R}^M_{++}} & \|\boldsymbol{\rho}\|_\infty \\ \text{s.t.} & \boldsymbol{p} \in \mathrm{Fix}(P_{\boldsymbol{\rho}}) \\ & \|\boldsymbol{p}\|_\infty \leq p_{\max}, \end{array} \tag{3}$$

where $\|\cdot\|_\infty$ denotes the standard l_∞ norm, and $p_{\max} \in \mathbb{R}_{++}$ is the maximum allowed transmit power.

Note that the first constraint in (3) simply states that the power allocation should induce the optimal load. To derive a simple algorithm able to solve Problem 1, we start with the following result:

Proposition 1. *If $(\boldsymbol{\rho}^\star, \boldsymbol{p}^\star) \in \mathbb{R}^M_{++} \times \mathbb{R}^M_{++}$ solves Problem 1, then both conditions hold:*

(i) $\|\boldsymbol{p}\|_\infty = p_{\max}$; and

(ii) there exists $c^\star \in \mathbb{R}_{++}$ such that $\boldsymbol{\rho}^\star = c^\star \mathbf{1}$.

Proof. We prove (i) and (ii) by obtaining a contradiction.

(i) Assume that the tuple $(\boldsymbol{\rho}', \boldsymbol{p}') \in \mathbb{R}^M_{++} \times \mathbb{R}^M_{++}$ solves Problem 1 and that $\|\boldsymbol{p}'\|_\infty < p_{\max}$. For $\alpha := p_{\max}/\|\boldsymbol{p}'\|_\infty > 1$, we have $\|\alpha \boldsymbol{p}'\|_\infty = p_{\max}$. Therefore, we can verify from the definition of the mapping $T_{\boldsymbol{p}}$ that $T_{\alpha\boldsymbol{p}}(\boldsymbol{\rho}) < T_{\boldsymbol{p}}(\boldsymbol{\rho})$ for every $(\boldsymbol{p}, \boldsymbol{\rho}) \in \mathbb{R}^M_{++} \times \mathbb{R}^M_{++}$. In particular, we have $T_{\alpha\boldsymbol{p}'}(\boldsymbol{\rho}') < T_{\boldsymbol{p}'}(\boldsymbol{\rho}') = \boldsymbol{\rho}'$. Now use [4, Fact 3.2 and 3.3] to verify that $\boldsymbol{\rho}'' \in \mathrm{Fix}(T_{\alpha\boldsymbol{p}'}) \neq \emptyset$ and that $\boldsymbol{\rho}'' < \boldsymbol{\rho}'$. In summary, we have just proved that $(\boldsymbol{\rho}'', \alpha\boldsymbol{p}')$ satisfies all constraints of Problem 1 and that $\|\boldsymbol{\rho}''\|_\infty < \|\boldsymbol{\rho}'\|_\infty$. These observations contradict optimality of $(\boldsymbol{\rho}', \boldsymbol{p}')$, and the proof is complete.

(ii) Denote by $(\boldsymbol{\rho}', \boldsymbol{p}') \in \mathbb{R}^M_{++} \times \mathbb{R}^M_{++}$ the solution to Problem 1, and define $[\rho'_1, \ldots, \rho'_M] := \boldsymbol{\rho}'$. Now assume that there exists $i \in \mathcal{M}$ such that $\rho'_i < \|\boldsymbol{\rho}'\|_\infty$. For this base station $i \in \mathcal{M}$, as a result of Fact 1, there exists a power allocation $\boldsymbol{p}'' \in \mathbb{R}^M_{++}$ satisfying $\boldsymbol{p}'' < \boldsymbol{p}' \leq p_{\max}\mathbf{1}$ that increases the load ρ'_i to the value $(\rho'_i + \|\boldsymbol{\rho}\|_\infty)/2 > \rho'_i$ while keeping the same load at every other base station. Denote this new load by $\boldsymbol{\rho}'' \in \mathrm{Fix}(T_{\boldsymbol{p}''})$. Note that $\|\boldsymbol{\rho}''\|_\infty = \|\boldsymbol{\rho}'\|_\infty$ by construction, so $(\boldsymbol{\rho}'', \boldsymbol{p}'')$, which satisfies all constraints of the problem, is also optimal because this tuple achieves the minimum of the cost function in (3). However, the inequality $\|\boldsymbol{p}''\|_\infty < p_{\max}$ contradicts (i), so the value $\|\boldsymbol{\rho}''\|_\infty = \|\boldsymbol{\rho}'\|_\infty$ cannot be the minimum value achieved by the cost function of the optimization problem. This result contradicts optimality of $(\boldsymbol{\rho}', \boldsymbol{p}') \in \mathbb{R}^M_{++} \times \mathbb{R}^M_{++}$. □

Now, consider the following optimization problem:

[1] If (1) does not have a solution, or if a solution $\boldsymbol{\rho}^\star$ satisfies $\|\boldsymbol{\rho}^\star\|_\infty > 1$, then we obtain the useful information that the traffic demand cannot be satisfied by the network configuration. In particular, with $\|\boldsymbol{\rho}^\star\|_\infty > 1$, we can rank base stations according to their nonserved traffic demand [13].

Problem 2.

$$\begin{array}{lll} \min._{(c,\boldsymbol{p})\in\mathbb{R}_{++}\times\mathbb{R}^M_{++}} & c & \\ \text{s.t.} & \boldsymbol{p}\in \mathrm{Fix}(P_{c\mathbf{1}}) & (4) \\ & \|\boldsymbol{p}\|_\infty \le p_{\max}, & \end{array}$$

where $\|\cdot\|_\infty$ denotes the standard l_∞ norm, and $p_{\max}$ is the maximum allowed transmit power.

Problem 1 and Problem 2 are formally equivalent in the sense that, by Proposition 1(ii), the tuple $(c^\star, \boldsymbol{p}^\star) \in \mathbb{R}_{++}\times\mathbb{R}^M_{++}$ solves Problem 2 if and only if the tuple $(c^\star\mathbf{1}, \boldsymbol{p}^\star) \in \mathbb{R}^M_{++}\times\mathbb{R}^M_{++}$ solves Problem 1. For ease of reference, we say that a tuple $(c, \boldsymbol{p}) \in \mathbb{R}_{++}\times\mathbb{R}^M_{++}$ is feasible to Problem 2 if and only if $\boldsymbol{p}\in\mathrm{Fix}(P_{c\mathbf{1}})$ and $\|\boldsymbol{p}\|_\infty \le p_{\max}$.

The proposed algorithm to solve Problem 2 is based on the following key observation:

Proposition 2. *(i) If $(c', \boldsymbol{p}') \in \mathbb{R}_{++}\times\mathbb{R}^M_{++}$ is feasible to Problem 2, then, for every $c'' \ge c'$, there exists $\boldsymbol{p}''\in\mathbb{R}^M_{++}$ such that $(c'', \boldsymbol{p}'') \in \mathbb{R}_{++}\times\mathbb{R}^M_{++}$ is also feasible.*

(ii) If Problem 2 has a solution, then the solution is unique.

(iii) Let $F \subset \mathbb{R}_{++}\times\mathbb{R}^M_{++}$ be the set of feasible tuples to Problem 2. If there exists $\boldsymbol{p}\in\mathbb{R}^M_{++}$ such that $\mathrm{Fix}(T_{\boldsymbol{p}}) \neq \emptyset$, then $F\neq\emptyset$.

(iv) Problem 2 has a solution if and only if there exists $\boldsymbol{p}\in\mathbb{R}^M_{++}$ for which $\mathrm{Fix}(T_{\boldsymbol{p}}) \neq \emptyset$.

Proof. (i) Immediate from Fact 1.

(ii) If $c^\star \in \mathbb{R}^M_{++}$ is the optimal load to Problem 2, and both $(c^\star, \boldsymbol{p}_1)$ and $(c^\star, \boldsymbol{p}_2)$ solve Problem 2 with $\boldsymbol{p}_1 \neq \boldsymbol{p}_2$, then $\{\boldsymbol{p}_1, \boldsymbol{p}_2\} \subset \mathrm{Fix}(P_{c^\star\mathbf{1}})$, and this relation contradicts that the set $\mathrm{Fix}(P_{c^\star\mathbf{1}})$ is a singleton if not empty.

(iii) Let $\boldsymbol{p}\in\mathbb{R}^M_{++}$ be such that $\mathrm{Fix}(T_{\boldsymbol{p}}) \neq \emptyset$. By Fact 2, we know that $\mathrm{Fix}(T_{\boldsymbol{p}}) \neq \emptyset$ for every $\boldsymbol{p} \in C := \{\boldsymbol{p}\in\mathbb{R}^M_{++} \mid \|\boldsymbol{p}\|_\infty = p_{\max}\} \neq \emptyset$. Choose $\boldsymbol{p}' \in C$ arbitrarily, and let $\boldsymbol{\rho}' \in \mathrm{Fix}(T_{\boldsymbol{p}'}) \neq \emptyset$. For any scalar c'' satisfying $c'' > \|\boldsymbol{\rho}'\|_\infty$, we have as a consequence of Fact 1 that $\boldsymbol{p}'' \in \mathrm{Fix}(P_{c''\mathbf{1}}) \neq \emptyset$ and that $\mathbf{0} < \boldsymbol{p}'' < \boldsymbol{p}' \le p_{\max}\mathbf{1}$. In other words, we have just proved that $(c'', \boldsymbol{p}'') \in F$, so $F \neq \emptyset$.

(iv) If $(c^\star, \boldsymbol{p}^\star) \in \mathbb{R}_{++}\times\mathbb{R}^M_{++}$ solves Problem 2, then $\boldsymbol{p}^\star \in \mathrm{Fix}(P_{c^\star\mathbf{1}}) \Leftrightarrow c^\star\mathbf{1} \in \mathrm{Fix}(T_{\boldsymbol{p}^\star})$, which shows that $\mathrm{Fix}(T_{\boldsymbol{p}^\star}) \neq \emptyset$, and one direction of the proof is complete. Conversely, assume that $\boldsymbol{\rho}' \in \mathrm{Fix}(T_{\boldsymbol{p}'}) \neq \emptyset$. By (iii), we know that the set $F \subset \mathbb{R}_{++}\times\mathbb{R}^M_{++}$ of feasible tuples is nonempty. Denote by $L := \{c \in \mathbb{R}_{++} \mid (c, \boldsymbol{p}) \in F\} \neq \emptyset$ the set of feasible loads, and let $c^\star := \inf L \ge 0$. We can construct with the elements of the set L a monotonically non-increasing sequence $\{c_n\}_{n\in\mathbb{N}} \subset L$ satisfying $\lim_{n\to\infty} c_n = c^\star$. Now consider the vector sequence $\{\boldsymbol{p}_n \in \mathrm{Fix}(P_{c_n\mathbf{1}})\}_{n\in\mathbb{N}} \subset \mathbb{R}^M_{++}$, and note that $\mathrm{Fix}(P_{c_n\mathbf{1}}) \neq \emptyset$ and that $\boldsymbol{p}_n \in B := \{\boldsymbol{p} \in \mathbb{R}^M_{+} \mid \|\boldsymbol{p}\|_\infty \le p_{\max}\}$ for every $n \in \mathbb{N}$ because $(c_n, \boldsymbol{p}_n) \in F$. Furthermore, by Fact 1 and monotonicity of $\{c_n\}_{n\in\mathbb{N}}$, the bounded vector sequence $\{\boldsymbol{p}_n\}_{n\in\mathbb{N}} \subset B\cap\mathbb{R}^M_{++}$ is monotonically nondecreasing in each coordinate, so it converges to a vector $\boldsymbol{p}^\star \in \mathbb{R}^M_{++}$. Moreover, since $B\subset\mathbb{R}^M$ is a closed set and $\{\boldsymbol{p}_n\}_{n\in\mathbb{N}} \subset B$, we also have $\boldsymbol{p}^\star \in B$. Now, by the definition of load, we have $c_n\mathbf{1} = T_{\boldsymbol{p}_n}(c_n\mathbf{1}) = [f_1(c_n\mathbf{1}, \boldsymbol{p}_n), \ldots, f_M(c_n\mathbf{1}, \boldsymbol{p}_n)]$, where f_i, $i\in\mathcal{M}$, is the continuous function with domain $\mathbb{R}^M_{+}\times\mathbb{R}^M_{++}$ defined in (2). Therefore,

$$\begin{aligned} c^\star\mathbf{1} &= \lim_{n\to\infty} c_n\mathbf{1} = \lim_{n\to\infty} [f_1(c_n\mathbf{1}, \boldsymbol{p}_n), \ldots, f_M(c_n\mathbf{1}, \boldsymbol{p}_n)] \\ &= [f_1(c^\star\mathbf{1}, \boldsymbol{p}^\star), \ldots, f_M(c^\star\mathbf{1}, \boldsymbol{p}^\star)] = T_{\boldsymbol{p}^\star}(c^\star\mathbf{1}) > \mathbf{0}, \end{aligned}$$

which implies that $c^\star > 0$ and that $\boldsymbol{p}^\star \in \mathrm{Fix}(P_{c^\star\mathbf{1}}) \subset B\cap\mathbb{R}^M_{++}$. This result proves that the infimum value $c^\star$ for the cost function in Problem 2 is achieved for the feasible tuple $(c^\star, \boldsymbol{p}^\star) \in \mathbb{R}_{++}\times\mathbb{R}^M_{++}$ constructed above, and the proof is complete. □

In words, Proposition 2(iv) simply states that, if a network is able to support the current rate requirements, then Problem 1 has a solution. In addition, Proposition 2(i) suggests the use of the bisection algorithm to obtain the minimum feasible uniform load $c^\star\mathbf{1}$. The idea is to devise an oracle able to answer whether a given uniform load value $c\in\mathbb{R}_{++}$ is greater than the optimal load $c^\star$. By Proposition 2(i)-(ii), we have that $c \ge c^\star$ if and only if there exists a vector $\boldsymbol{p}_c \in \mathrm{Fix}(P_{c\mathbf{1}}) \neq \emptyset$ with $\|\boldsymbol{p}_c\|_\infty \le p_{\max}$. The oracle described later verifies the existence of such a vector without necessarily computing it. For the moment, assume knowledge of an oracle able to give the correct answer and knowledge of scalars $(c_{\min}, c_{\max}) \in \mathbb{R}_{+}\times\mathbb{R}_{++}$ satisfying $c^\star \in [c_{\min}, c_{\max}]$. [2] Then Problem 2 can be solved as follows. Starting with the interval $[c_{\min}, c_{\max}]$, we use the oracle to determine whether $\bar{c} := (c_{\max} + c_{\min})/2 \ge c^\star$. If the answer is positive, then we update the upper end $c_{\max}$ of the interval according to $c_{\max} \leftarrow \bar{c}$. Otherwise, we update the lower end of the interval $c_{\min} \leftarrow \bar{c}$. These steps are repeated until the length $c_{\max} - c_{\min}$ of the interval $[c_{\min}, c_{\max}] \ni c^\star$ is sufficiently small. Once the optimal load is estimated with a sufficient precision, we recover the optimal power allocation by computing the fixed point of $P_{c^\star\mathbf{1}}$ by using any existing method (see [5]). Algorithm 1 summarizes the proposed mechanism to minimize the maximum load.

Algorithm 1: Bisection algorithm for load minimization.

Data: Power mapping $P_{\boldsymbol{\rho}}$; maximum interval length $\epsilon > 0$; $(c_{\min}, c_{\max}) \in \mathbb{R}_{+}\times\mathbb{R}_{++}$ satisfying $c_{\min} < c^\star \le c_{\max}$, where $(c^\star, \boldsymbol{p}^\star) \in \mathbb{R}_{++}\times\mathbb{R}^M_{++}$ is the solution to Problem 2, assumed to exist;

Result: Solution $(c^\star, \boldsymbol{p}^\star) \in \mathbb{R}_{++}\times\mathbb{R}^M_{++}$ to Problem 2.

while $c_{\max} - c_{\min} > \epsilon$ **do**

  $\bar{c} \leftarrow \dfrac{c_{\max} + c_{\min}}{2}$;

  if $\bar{c} \ge c^\star$ **then**

    $c_{\max} \leftarrow \bar{c}$;

  else

    $c_{\min} \leftarrow \bar{c}$;

      ▷ Use Algorithm 2 to check the inequality;

$\boldsymbol{p}^\star \leftarrow \boldsymbol{p} \in \mathrm{Fix}(P_{c_{\max}\mathbf{1}})$;

  ▷ See [5, 14] for simple fixed point algorithms;

return $(c_{\max},\ \boldsymbol{p}^\star)$

We now turn our attention to the oracle, which is a direct application of the following result (the proof can be obtained with little effort by using [15, Facts 2 and 3], but it is omitted owing to the space limitation):

Proposition 3. *For a given (uniform) load value $c\in\mathbb{R}_{++}$, consider the sequences $\{\boldsymbol{p}'_n\}_{n\in\mathbb{N}}$ and $\{\boldsymbol{p}''_n\}_{n\in\mathbb{N}}$ generated by $\boldsymbol{p}'_{n+1} = P_{c\mathbf{1}}(\boldsymbol{p}_n)$ and $\boldsymbol{p}''_{n+1} = \min\{P_{c\mathbf{1}}(\boldsymbol{p}''_n), p_{\max}\mathbf{1}\}$, where $\boldsymbol{p}'_1 = \mathbf{0}$, $\boldsymbol{p}''_1 = p_{\max}\mathbf{1}$, and $\min$ denotes the coordinate-wise minimum operator. Denote by $c^\star \in \mathbb{R}_{++}\cup\{\infty\}$ the optimal value attained by the cost function in Problem 2 (i.e., the minimum load), where we use the convention that $c^\star = \infty$ if Problem 2 has no solution. Then each of the following holds:*

---

[2]If the current network configuration is able to support the rate requirement of users with load $\boldsymbol{\rho}$, then Proposition 2 reveals that Problem 1 has a solution, and the current load $c_{\max} = \|\boldsymbol{\rho}\|$ is clearly an upper bound of the optimal uniform load $c^\star$. For $c_{\min}$, we can use the trivial bound $c_{\min} = 0$.

(i) The sequences $\{\|\boldsymbol{p}'_n\|_\infty\}_{n\in\mathbb{N}}$ and $\{\|\boldsymbol{p}''_n\|_\infty\}_{n\in\mathbb{N}}$ are monotonically nondecreasing and non-increasing, respectively.

(ii) $c < c^\star$ if and only if there exists $n \in \mathbb{N}$ such that $\|\boldsymbol{p}'_n\|_\infty > \|\boldsymbol{p}''_n\|_\infty$.

(iii) $c > c^\star$ if and only if there exists $n \in \mathbb{N}$ such that $\|\boldsymbol{p}''_n\|_\infty < p_{\max}$.

(iv) If $c \geq c^\star$, then, for every $n \in \mathbb{N}$, we have $0 \leq \max\{\|\boldsymbol{p}'_n - \boldsymbol{p}^\star\|_\infty, \|\boldsymbol{p}''_n - \boldsymbol{p}^\star\|_\infty\} \leq \|\boldsymbol{p}''_n - \boldsymbol{p}'_n\|_\infty =: e_n$ and $\lim_{n\to\infty} e_n = 0$, where $\boldsymbol{p}^\star$ denotes the optimal power allocation.

In practice, Proposition 3(ii)-(iii) enables us to verify whether a given load c is strictly greater or smaller than the optimal load $c^\star$ in a *finite number of steps*. In the unlikely case that the probed load is *exactly* the optimal load $c^\star$, then the results in Proposition 3(ii)-(iii) are unable to produce a certificate that $c \geq c^\star$. However, if after a sufficiently large number $n \in \mathbb{N}$ of iterations we observe negligible changes in the sequences $\{\boldsymbol{p}'_n\}$ and $\{\boldsymbol{p}''_n\}$, and $\|\boldsymbol{p}''_n - \boldsymbol{p}'_n\|_\infty$ is small, then Proposition 3(iv) provides us with some confidence that the probed load is likely to be close to the optimal load and that the optimal power allocation $\boldsymbol{p}^\star$ is close to both $\boldsymbol{p}'_n$ and $\boldsymbol{p}''_n$ with respect to the metric $d : \mathbb{R}^M \times \mathbb{R}^M \to \mathbb{R}_+ : (\boldsymbol{x}, \boldsymbol{y}) \mapsto \|\boldsymbol{x} - \boldsymbol{y}\|_\infty$. Algorithm 2 summarizes the proposed oracle based on Proposition 3.

Algorithm 2: Oracle for Algorithm 1.

Data: Maximum transmit power $p_{\max}$; minimum step $\epsilon > 0$ per iteration; load $c \in \mathbb{R}_{++}$ to be probed;
Result: Answer whether $c < c^\star$ or $c \geq c^\star$ ($c^\star$ denotes the optimal load to Problem 2);
$\boldsymbol{p}'_1 \leftarrow \boldsymbol{0}$; $\boldsymbol{p}''_1 \leftarrow p_{\max}\boldsymbol{1}$; $\boldsymbol{p}'_2 = P_{c\boldsymbol{1}}(\boldsymbol{0})$;
$\boldsymbol{p}''_2 = \min\{P_{c\boldsymbol{1}}(p_{\max}\boldsymbol{1}), p_{\max}\boldsymbol{1}\}$; $n \leftarrow 2$;
while *True* **do**
  $\boldsymbol{p}'_{n+1} = P_{c\boldsymbol{1}}(\boldsymbol{p}'_n)$; $\boldsymbol{p}''_{n+1} = \min\{P_{c\boldsymbol{1}}(\boldsymbol{p}''_n), p_{\max}\boldsymbol{1}\}$;
  $n \leftarrow n + 1$;
  if $\|\boldsymbol{p}'_n\|_\infty > \|\boldsymbol{p}''_n\|_\infty$ **then**
    return *Certificate that* $c < c^\star$;
  else if $\|\boldsymbol{p}''_n\|_\infty < p_{\max}$ **then**
    return *Certificate that* $c > c^\star$;
  else if $\|\boldsymbol{p}'_n - \boldsymbol{p}'_{n-1}\|_\infty < \epsilon$ *and* $\|\boldsymbol{p}''_n - \boldsymbol{p}''_{n-1}\|_\infty < \epsilon$ *and* $\|\boldsymbol{p}''_n - \boldsymbol{p}'_n\|_\infty < \epsilon$ **then**
    return *The probed value c is likely to be close to* $c^\star$;
    ▷ Stop the algorithm because no progress can be made;

4. SIMULATIONS

We now solve the load optimization problem in a scenario mimicking the stadium test case proposed by the European METIS project [12]. In more detail, we place 142 micro base stations under the roof of a stadium with a height of 33 m. The roof covers all stands. In the direction of the rows, base stations are separated by 10 m, and in the perpendicular direction they have a separation of 15 m. The network operates at 2.6 GHz with a bandwidth of $B = 180$ MHz. We assume that there are $K = 900$ resource blocks in the system, and the maximum total transmit power per resource block is $1/900$ W. To compute the path loss between users and base stations, we use the ITU urban micro line-of-sight model (ITU UMi LOS) [16, Table A1-2]. The main parameters for the configuration of the antennas at the base stations is set as follows: antenna tilt $\phi_{\text{tilt}} = 90°$, maximum attenuation

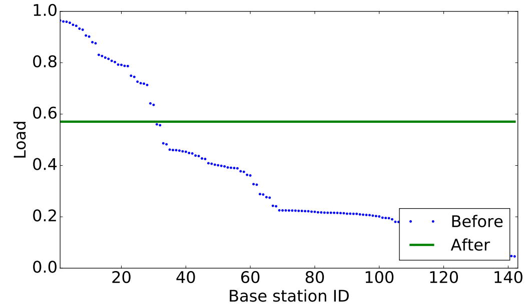

Fig. 1: Load before and after optimization.

$A_{\text{m}} = 30$ dB, 3-dB beamwidth $\theta_{3\text{dB}} = 15°$, 3-dB elevation $\phi_{3\text{dB}} = 20°$, and base station antenna gain $G = 17$ dBi.

Stands have an inclination of 30° with respect to the ground. There are 43,800 simultaneously active users demanding 384 kbps of traffic. These users are uniformly distributed in the stands, and they are connected to the base stations with the strongest received signal when all base stations transmit at full power. The antenna of the users are placed 1 m above the ground.

With the above configuration, if all base stations transmit at full power per resource block $p_{\max} = 1/900$ W, the maximum observed load in the network is around 0.96, which can be seen as too close to its physical limit. We then use Algorithm 1 ($\epsilon = 10^{-5}, c_{\min} = 0, c_{\max} = 1$) with the oracle in Algorithm 2 ($\epsilon = 10^{-6}$) to minimize the maximum observed load in the network. Results for the load at each base station before and after load optimization are shown in Fig. 1. For visual clarity, with the initial power configuration, we label base stations according to their load in descending order.

It is clear from Fig. 1 that the peak load has been greatly reduced at the expense of increased load at base stations that were originally lightly loaded. However, after optimization, all base stations have more than 40% of the resource blocks available for data transmission. Furthermore, as discussed in [8] (see also [5]), increasing the load in lightly loaded base stations is not necessarily a bad feature because the transmit power decreases. In fact, in this scenario, although saving energy is not an objective of the proposed algorithm, the total transmit power of the network (which is proportional to $\boldsymbol{p}^t\boldsymbol{\rho}$, where $\boldsymbol{p}$ is the power inducing the load $\boldsymbol{\rho}$) after optimization is roughly 40% of the total transmit power before optimization. Similar results have been obtained in different scenarios proposed by the METIS project, but we omit these results owing to the space limitation.

5. SUMMARY AND CONCLUSIONS

We have shown that a simple power control scheme can decrease the peak load in ultra-dense networks. The proposed scheme can be especially useful when few base stations have high load and resource blocks in these base stations have to the released to, for example, accommodate new users in the system. Simulations show that the total transmit power can also be reduced by a large factor if most base stations in the network are lightly loaded (but we emphasize that the total transmit power may not be always reduced because reducing the total transmit power is not the objective of the proposed approach).

Acknowledgement: This work was partially supported by the Deutsche Forschungsgemeinschaft (DFG) under Grant STA 864/9-1.

6. REFERENCES

[1] Gerhard Wunder, Peter Jung, Martin Kasparick, Thorsten Wild, Frank Schaich, Yejian Chen, Stephan Ten Brink, Ivan Gaspar, Nicola Michailow, Andreas Festag, et al., "5GNOW: non-orthogonal, asynchronous waveforms for future mobile applications," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 97–105, 2014.

[2] Ian C Wong, Zukang Shen, Brian L Evans, and Jeffrey G Andrews, "A low complexity algorithm for proportional resource allocation in OFDMA systems," in *Signal Processing Systems, 2004. SIPS 2004. IEEE Workshop on*. IEEE, 2004, pp. 1–6.

[3] K. Majewski and M. Koonert, "Conservative cell load approximation for radio networks with Shannon channels and its application to LTE network planning," in *Telecommunications (AICT), 2010 Sixth Advanced International Conference on*, May 2010, pp. 219 –225.

[4] R. L. G. Cavalcante, S. Stańczak, M. Schubert, A. Eisenbläter, and U. Türke, "Toward energy-efficient 5G wireless communication technologies," *IEEE Signal Processing Mag.*, vol. 31, no. 6, pp. 24–34, Nov. 2014.

[5] R. L. G. Cavalcante, S. Stańczak, J. Zhang, and H. Zhuang, "Low complexity iterative algorithms for power estimation in ultra-dense load coupled networks," *IEEE Trans. Signal Processing*, vol. 7, no. 22, pp. 6058–6070, Nov. 2016.

[6] Iana Siomina and Di Yuan, "Load balancing in heterogeneous LTE: Range optimization via cell offset and load-coupling characterization," in *Communications (ICC), 2012 IEEE International Conference on*. IEEE, 2012, pp. 1357–1361.

[7] C Ho, Di Yuan, and Sumei Sun, "Data offloading in load coupled networks: A utility maximization framework," *IEEE Trans. Wireless Commun.*, vol. 13, no. 4, pp. 1921–1931, April 2014.

[8] Chin Keong Ho, Di Yuan, Lei Lei, and Sumei Sun, "On power and load coupling in cellular networks for energy optimization," *IEEE Trans. Wireless Commun.*, vol. 14, no. 1, pp. 500–519, Jan. 2015.

[9] Albrecht Fehske, Henrik Klessig, Jens Voigt, and Gerhard Fettweis, "Concurrent load-aware adjustment of user association and antenna tilts in self-organizing radio networks," *IEEE Trans. Veh. Technol.*, vol. 62, no. 5, pp. 1974–1988, June 2013.

[10] E. Pollakis, R. L. G. Cavalcante, and S. Stanczak, "Base station selection for energy efficient network operation with the majorization-minimization algorithm," in *Signal Processing Advances in Wireless Communications (SPAWC), 2012 IEEE 13th International Workshop on*, June 2012.

[11] R. L. G. Cavalcante, M. Kasparick, and S. Stańczak, "Max-min utility optimization in load coupled interference networks," *IEEE Trans. Wireless Commun.*, accepted for publication.

[12] Patrick Agyapong, Volker Braun, Mikael Fallgren, Alexandre Gouraud, Martin Hessler, Sebastian Jeux, Andreas Klein, Ji Lianghai, David Martn-Sacristn, Michal Maternia, Martti Moisio, Jose F. Monserrat, Krystian Pawlak, Hugo Tullberg, and Andreas Weber, "Deliverable D6.1 - simulation guidelines," Tech. Rep., METIS, Oct. 2013.

[13] Ioana Siomina and Di Yuan, "Analysis of cell load coupling for LTE network planning and optimization," *IEEE Trans. Wireless Commun.*, vol. 11, no. 6, pp. 2287–2297, June 2012.

[14] Renato L. G. Cavalcante, Yuxiang Shen, and Slawomir Stańczak, "Elementary properties of positive concave mappings with applications to network planning and optimization," *IEEE Trans. Signal Processing*, vol. 64, no. 7, pp. 1774–1873, April 2016.

[15] R. L. G. Cavalcante, E. Pollakis, and S. Stanczak, "Power estimation in LTE systems with the general framework of standard interference mappings," in *IEEE Global Conference on Signal and Information Processing (GlobalSIP' 14)*, Dec. 2014.

[16] ITU-R, "ITU-R M. 2135: Guidelines for evaluation of radio interface technologies for IMT-advanced," Tech. Rep., 2008.