A Novel Multiobjective Cell Switch-Off Framework for Cellular Networks

David González G., *Member, IEEE,* Jyri Hämäläinen, *Member, IEEE,*
Halim Yanikomeroglu, *Senior Member, IEEE,* Mario García-Lozano and Gamini Senarath

Abstract—**Cell Switch-Off (CSO) is recognized as a promising approach to reduce the energy consumption in next-generation cellular networks. However, CSO poses serious challenges not only from the resource allocation perspective but also from the implementation point of view. Indeed, CSO represents a difficult optimization problem due to its NP-complete nature. Moreover, there are a number of important practical limitations in the implementation of CSO schemes, such as the need for minimizing the real-time complexity and the number of on-off/off-on transitions and CSO-induced handovers. This article introduces a novel approach to CSO based on multiobjective optimization that makes use of the statistical description of the service demand (known by operators). In addition, downlink and uplink coverage criteria are included and a comparative analysis between different models to characterize intercell interference is also presented to shed light on their impact on CSO. The framework distinguishes itself from other proposals in two ways: 1) The number of on-off/off-on transitions as well as handovers are minimized, and 2) the computationally-heavy part of the algorithm is executed offline, which makes its implementation feasible. The results show that the proposed scheme achieves substantial energy savings in small cell deployments where service demand is not uniformly distributed, without compromising the Quality-of-Service (QoS) or requiring heavy real-time processing.**

Index Terms—**Cellular networks, energy efficiency, cell switch-off, CSO, multiobjective optimization, Pareto efficiency.**

I. INTRODUCTION

FUTURE hyper-dense small-cell deployments are expected to play a pivotal role in delivering high capacity and reliability by bringing the network closer to users [1]. However, in order to make hyper-dense deployments a reality, enhancements including effective interference management, self-organization, and energy efficiency are required [2]. Given that large-scale deployments composed of hundreds or thousands of network elements can increase the energy consumption substantially, the need for energy efficiency (*green communications*) has been recognized by the cellular communications industry as an important item in research projects and standardization activities [3, 4].

Initial attempts to improve the energy efficiency in cellular networks were oriented towards minimizing the power radiated through the air interface, which in turn reduces the electromagnetic pollution and its potential effects on human health.

David González G. and Jyri Hämäläinen are with the School of Electrical Engineering at Aalto University, Finland. Corresponding email: david.gonzalez.g@ieee.org.

Halim Yanikomeroglu is with the Department of Systems and Computer Engineering, Carleton University, Ottawa, Canada.

Mario García-Lozano is with the Department of Signal Theory and Communications, Universitat Politècnica de Catalunya, Barcelona, Spain.

Gamini Senarath is with Huawei Canada R&D Centre.

However, most of the energy consumption (between 50% to 80%) in the radio access network takes place in base stations (BSs) [4] and it is largely independent of the BSs' load. Since cellular networks are dimensioned to meet the service demand in *the busy hour* (i.e., peak demand), it is expected that, under non-uniform demand distributions (both in space and time), a substantial portion of the resources may end up being underutilized, thus incurring in an unnecessary expenditure of energy. The problem may become worse in many of the scenarios foreseen for 5G, presumably characterized by hyper-dense small-cell deployments, hierarchical architectures, and highly heterogeneous service demand conditions [5]. Therefore, the idea of switching off lightly loaded base stations has been considered recently as a promising method to reduce the energy consumption in cellular networks. This framework is referred to as Cell Switch-Off (CSO) and it is focused on determining the largest set of cells that can be switched off without compromising the Quality-of-Service (QoS) provided to users. Unfortunately, CSO is difficult to carry out due to the fact that it represents a highly challenging (combinatorial) optimization problem whose complexity grows exponentially with the number of BSs, and hence, finding optimal solutions is not possible in polynomial time. Moreover, the implementation of CSO requires coordination among neighbor cells and several other practical aspects, such as coverage provision and the need for minimizing the number of (induced) handovers and on-off/off-on transitions. In practice, optimizing the number of transitions, as well as the time required for them, is advisable because switching on/off BSs is far from being a simple procedure, and indeed, this process must be gradual and controlled [6, 7]. Moreover, a large number of transitions could result in a high number of handovers with a potentially negative impact on QoS [8].

Although CSO is a relatively young research topic, a significant amount of contributions has been made. Hence, an exhaustive survey is both out of the scope and not feasible herein. Instead, a literature review including, in the opinion of the authors, some of the most representative works is provided. Thus, in the comparative perspective shown in Table I, the following criteria have been considered:

- *CSO type / architecture*: CSO solutions can be classified as 'snapshot' or 'traffic profiling' CSO depending on the approach followed to take the on/off decisions. In snapshot CSO (e.g., [14, 16, 17, 19]), decisions involve the analysis of discrete realizations of users, i.e., whenever a CSO decision is required, information of every single user in the network needs to be available at a central unit where an heuristic or optimization procedure is performed. Given its nature, this

TABLE I
SUMMARY OF RELATED WORK.

Ref	Year	CSO type / architecture	Realistic ICI model	Coverage aspects	Complexity and feasibility analysis
[9]	2016	Snapshot / Semi-distributed	Full-load	Downlink	✓
[10]	2016	Traffic profiling / Centralized	Full-load	Downlink	Partially
[11]	2016	Traffic profiling / Semi-distributed	Load-coupling	Downlink	Partially
[12]	2015	Snapshot / Centralized	Full-load	Downlink	Partially
[13]	2015	Snapshot / Centralized	Full-load	Downlink	Partially
[14]	2014	Snapshot / Centralized	Constant	Downlink	×
[15]	2014	Traffic profiling / Semi-distributed	Full-load	Downlink	✓
[16]	2014	Snapshot / Centralized	Constant	Downlink	×
[17]	2013	Snapshot / Centralized	Full-load	Downlink	×
[18]	2013	Traffic profiling / Centralized	Load-coupling	Downlink	×
[19]	2013	Snapshot / Centralized	Constant	×	Partially
[20]	2013	Snapshot / Centralized	Full-load	Downlink	Partially
[21]	2013	Snapshot / Centralized	Full-load	Downlink	×
[22]	2012	Snapshot / Centralized	Constant	×	×
[23]	2012	Snapshot / Centralized	Full-load	Downlink	×
[24]	2011	Traffic profiling / Semi-distributed	Full-load	Downlink	Partially
[25]	2010	Snapshot / Centralized	Full-load	Downlink	×
[26]	2010	Snapshot / Centralized	Constant	Downlink	×
[27]	2010	Traffic profiling / Semi-distributed	Full-load	×	×
[28]	2009	Snapshot / Both	Full-load	×	×

type of solution provides the ultimate performance in terms of energy savings. However, feasibility becomes a serious issue as it will be shown later on. On the other hand, in traffic profiling CSO (e.g., [15, 24, 27]), a certain knowledge about the service demand behavior is assumed.

- *Realistic ICI model*: Switching off/on BSs modifies not only the levels of Intercell Interference (ICI), but also the resulting load coupling [29]. The load of each BS can be understood as the fraction of resources that are being used to satisfy a certain service demand. There are three different ICI models that can be used in CSO:

  1) *Constant*: The assumption is that ICI is always constant or it does not exist (e.g., [14, 19]). In such cases, ICI levels are calculated as if all cells were always active no matter their actual state (on or off).

  2) *Full-load* (FL): ICI is only created by active BSs, but assuming that they are fully loaded (e.g., [9, 15, 20]). This approach is reasonable in CSO since traffic is concentrated in a subset of BSs which tend to be highly loaded. Thus, active BSs always cause ICI to its neighbors while BSs in sleeping mode do not.

  3) *Load-coupling* (LC): The ICI created by each BS is proportional to its load (e.g., [11, 18]). This model is more realistic and accurate but involves more complexity.

- *Coverage*: It indicates whether coverage aspects (in downlink, uplink, or both) are considered. This is an important criterion in CSO to avoid coverage holes.

- *Complexity and feasibility*: It indicates whether a complexity and/or feasibility analysis is provided. For instance, in case of CSO, how much information is required to be exchanged among cells is an important metric. Moreover, the number of transitions and handovers in a dynamic environment must also be estimated.

As it can be seen from Table I, snapshot-based schemes require centralized operation, and due to the NP-complete nature of the problem, heuristics are the preferred way to deal with it. Representative examples of this type of solutions include [14, 17, 22, 26] where the main idea is, in general, to sequentially switch 1) on highly loaded cells (cells that can get more traffic), or 2) off lightly loaded ones (cells whose users can easily be migrated to neighbor BSs) until the service demand is fulfilled.

In general, this type of schemes requires employing basic models for ICI, e.g., [14, 22, 26]. As indicated, another approach to CSO is traffic profiling, the case of [15, 18, 24, 27], where CSO decisions are taken assuming a certain knowledge about the service demand [30, 31]. However, most of the solutions presented so far employ models and assumptions oriented to macrocellular deployments, such as [11, 18, 24], where the use of CSO is not so clear due to practical issues [32]. Other practical aspects, such as the number of transitions and handovers are often overlooked as well. Recently, CSO has also been studied in the context of infrastructure sharing [12] and by means of stochastic geometry [10].

Thus, in the light of these observations, this paper presents a novel multiobjective framework[1] that 1) includes the strengths of previous proposals, 2) overcomes many of their drawbacks, and 3) extends the analysis to address aspects that have not fully been investigated.

Main contribution: a framework for multiobjective optimization for CSO that explicitly takes into account a statistical description of the service demand distribution when computing the performance metrics.

As it will be illustrated by the numerical results, this idea has the following advantages:

✓ The use of the spatial service demand distribution (repre-

---

[1]A preliminary version of this work was presented in [15]. From it, a US patent application has also been made: US patent application no: 14/334,134, application date: 17 July 2014.

sented by a spatial probability density function) allows to the proposed algorithm to rapidly identify network topologies, i.e., on/off patterns, providing higher capacity to areas where high service demand is more likely to appear, and hence, the *search space* and required computational effort is significantly reduced.

✓ Given that, in general, traffic profiles are *stable* in time scales of dozens of minutes, the (computationally-heavy) optimization can be done offline and required topologies can be applied as needed as these traffic profiles are recognized/observed during network operation. This feature makes the implementation of CSO feasible, given that the required BS coordination and real-time processing are significantly reduced. However, the conceptual idea will still be valid when new paradigms (cloud computing, software defined networking, and network virtualization, see [1]) allow faster computation and information exchange among network nodes. Thus, more dynamic traffic profile recognition and optimization will also be possible under the proposed framework.

✓ The proposed optimization formulation allows considering several downlink and uplink coverage criteria, such as minimum received power and Signal to Interference plus Noise Ratio (SINR).

✓ Given that there exist a correlation between the topologies that are specific for a given traffic profile, the number of handovers and transitions is minimized.

Finally, the following set of secondary/minor contributions are also presented in this work:

1) Analysis of the impact of the most extended interference models (FL and LC) on the performance of CSO.

2) A quantitative assessment of how CSO operation affects the critical uplink power consumption (on user equipment side). To the best of the authors' knowledge, this aspect/criterion has been overlooked in most of previous studies, and only recently in [13] it has been integrated within an optimization framework.

3) While the multiobjective problem formulation presented herein can be solved by means of standard stochastic optimization tools [33], an alternative iterative algorithm is also proposed for computing the important tradeoff between the number of active cells and network capacity. As it will be shown shortly, although its performance is slightly inferior to stochastic search, it is significantly faster, and hence, it could be used when real-time operation becomes feasible.

The rest of the paper is organized as follows: the next section presents the system model. The proposed Multiobjective Optimization (MO) framework (performance metrics and problem formulation) is explained in Section III. Section IV describes the evaluation setting and benchmarks used in simulations. Numerical results are also analyzed therein. Section V closes the article with conclusions and research directions.

II. SYSTEM MODEL

A. Definitions and notations

In this study, an OFDMA cellular network is considered. The system bandwidth is B and the network is composed of L BSs that can be independently switched off/on. The indices of the BSs are contained in the set $\mathcal{L} = \{0, 1, \cdots, L-1\}$. The set $\mathcal{A}$, composed of $A = |\mathcal{A}|$ small area elements, represents the spatial domain to which the network provides service. It is assumed that within each area elements $a \in \mathcal{A}$, the average received power is constant. The maximum transmit power per cell is $P_{\max}$. The network geometry is captured by the path loss matrix $\mathbf{G} \in \mathbb{R}^{A \times L}$ (distance dependent attenuation, antenna gains, and shadowing). The vectors $\mathbf{p}_{\mathrm{PS}}$ and $\mathbf{p}_{\mathrm{D}}$, both $\in \mathbb{R}^L$, indicate the transmit power at each cell in Pilot Signals (PS), and data channels, respectively. Cell selection is based on the average PS received power, which can be calculated by means of the following expression:

$$\mathbf{R}_{\mathrm{PS}} = \mathbf{G} \cdot \mathrm{diag}\left(\mathbf{p}_{\mathrm{PS}} \odot \mathbf{x}\right), \quad \mathbf{R}_{\mathrm{PS}} \in \mathbb{R}^{A \times L}. \quad (1)$$

The operator $\odot$ denotes Hadamard (pointwise) operations. The vector $\mathbf{x} \in \{0, 1\}^L$ indicates which cells are active and which ones are switched off. Hereafter, $\mathbf{x}$ is also referred to as 'network topology'. Thus, the matrix $\mathbf{R}_{\mathrm{PS}}$ in (1) contains the PS received power, i.e., $\mathbf{R}_{\mathrm{PS}}(a, l)$ indicates the received power from the lth BS in the ath area element. Of interest is the Number of Active Cells (NAC) in each network topology as energy consumption is related to it. The ath area element (the ath row in $\mathbf{R}_{\mathrm{PS}}$) is served by cell $l^{\star}$ if

$$l^{\star} = \underset{l \in \{0, 1, \ldots, L-1\}}{\mathrm{argmax}} \mathbf{R}_{\mathrm{PS}}(a, l). \quad (2)$$

The dependence of $l^{\star}$ on $\mathbf{x}$ has not been explicitly indicated for the sake of clarity. Based on (1) and (2), the binary coverage matrices $\mathbf{S}$ and $\mathbf{S}^{\mathrm{c}} \in \mathbb{R}^{A \times L}$ can be obtained. If the ath pixel is served by $l^{\star}$, then $\mathbf{S}(a, l^{\star}) = 1$. $\mathbf{S}^{\mathrm{c}}$ is the binary complement of $\mathbf{S}$. Hence, the coverage pattern, implicitly defined in $\mathbf{S}$, is a function of $\mathbf{x}$. The cell selection rule indicated by (2) can be regarded as a connectivity function $f_{\mathrm{c}} : \mathcal{A} \rightarrow \mathcal{L} \cup \{-1\}$. If $f_{\mathrm{c}}(a) = -1$, the ath area element is out of coverage, i.e.,

- the received power in $a \in \mathcal{A}$ ($\mathbf{R}_{\mathrm{PS}}(a, l^{\star})$) is smaller than $P_{\min}^{\mathrm{Rx}}$, i.e., $\mathbf{R}_{\mathrm{PS}}(a, l^{\star}) \leq P_{\min}^{\mathrm{Rx}}$,
- the SINR (ψ_a) in the area element $a \in \mathcal{A}$ is smaller than $\psi_{\min}$, i.e., $\psi_a \leq \psi_{\min}$, or
- the path loss $\mathbf{G}(a, l^{\star})$ between the area element and its server is greater than $G_{\max}^{\mathrm{UL}}$, i.e., $\mathbf{G}(a, l^{\star}) \geq G_{\max}^{\mathrm{UL}}$. In practice, $G_{\max}^{\mathrm{UL}}$ is the maximum path-loss obtained from the uplink link budget (a design criterion).

The cell $\mathcal{A}_l$ is the subset of $\mathcal{A}$ served by the lth BS. Thus, $\mathcal{A}_l \triangleq \{a \in \mathcal{A} : f_{\mathrm{c}}(a) = l\}$, where $\mathcal{A}_i \cap \mathcal{A}_j = \emptyset, \forall i \neq j$. The set $\mathcal{A}_{\mathrm{c}}$ is the subset of $\mathcal{A}$ that are associated to one BS. Thus, $\mathcal{A}_{\mathrm{c}} \triangleq \{a \in \mathcal{A} : f_{\mathrm{c}}(a) \neq -1\} = \bigcup_{l \in \mathcal{L}} \mathcal{A}_l$. The vector $\mathbf{\Gamma} \in \mathbb{R}^A$ corresponds to the spatial service demand distribution. Thus, $\mathbf{\Gamma}(a)$ indicates the probability, in the event of a new user, that the ath pixel has the user on it, and hence, $\mathbf{\Gamma}^{\mathrm{T}} \cdot \mathbf{1} = 1$. It should be noted that $\mathbf{\Gamma}$ is time-dependent, however it is reasonable to assume that $\mathbf{\Gamma}$ is constant during fixed intervals [30]. In order to represent the service demand volume, two parameters are considered: inter-arrival time (λ) and session time (μ). Both are modeled as exponentially distributed random variables. Thus, service demand's spatial distribution and volume are described by $\mathbf{\Gamma}$ and the first order statistics of λ and μ, i.e., $\mathbb{E}\{\lambda\}$ and $\mathbb{E}\{\mu\}$, respectively. It is assumed that the QoS of a

user is satisfied if the target rate ($r_{\min}$) is fulfilled. Hence, the total service demand volume (R) in $\mathcal{A}$ is given by

$$R = \sum_{a \in \mathcal{A}} r_a \quad \text{[bps]}, \tag{3}$$

where

$$r_a = \frac{\mathbb{E}\{\mu\}}{\mathbb{E}\{\lambda\}} \cdot \mathbf{\Gamma}(a) \cdot r_{\min} \quad \text{[bps]} \tag{4}$$

corresponds to the average demand in the ath area element. The previous model for the service demand can easily be extended to the general case of more than one service to account with the fact that service time, inter-arrival time, spatial distribution, and target rate can be *service-specific*. Assuming that there are N_{S} service classes, each of them with different characteristics, i.e., μ^c, λ^c, $\mathbf{\Gamma}^c$, and $r_{\min}^c$ for $c = 1, 2, \cdots, N_{\text{S}}$, (4) can be rewritten as follows:

$$r_a^{\text{S}} = \sum_{c=1}^{N_{\text{S}}} \left(\frac{\mathbb{E}\{\mu^c\}}{\mathbb{E}\{\lambda^c\}} \cdot \mathbf{\Gamma}^c(a) \cdot r_{\min}^c \right) \quad \text{[bps]}. \tag{5}$$

The resulting spatial service demand distribution (that is required to compute the performance metrics introduced later on) can be obtained by considering the resulting demand as follows:

$$\mathbf{\Gamma}^{\text{S}}(a) = \frac{r_a^{\text{S}}}{\sum_{a \in \mathcal{A}} r_a^{\text{S}}}. \tag{6}$$

Hereafter, one single service class (possibly the result of a mix of many others) is assumed for the sake of clarity, and hence, one single set of parameters (μ, λ, $\mathbf{\Gamma}$, and $r_{\min}$) are considered.

Definition II.1 (Cell load). The load of the lth BS ($\alpha_l(t)$), at any given time t, is defined as the fraction of the available resources (bandwidth) that are being used.

The average load of the lth BS is $\bar{\alpha}_l \triangleq \mathbb{E}\{\alpha_l(t)\}$. Thus, the vector $\bar{\boldsymbol{\alpha}} = [\ \bar{\alpha}_0\ \bar{\alpha}_1\ \cdots\ \bar{\alpha}_{L-1}\ ]$ indicates the load conditions in the network, on average. Note that if $\mathbf{x}(l) = 0$, then $\bar{\boldsymbol{\alpha}}(l) = 0$. As the reader can easily infer, as long as $\bar{\boldsymbol{\alpha}} \leq \mathbf{1}$, the network topology ($\mathbf{x}$) is able to satisfy, on average, the service demand given by $\mathbf{\Gamma}$, $\mathbb{E}\{\lambda\}$, and $\mathbb{E}\{\mu\}$, and hence, it can be said that $\mathbf{x}$ is *adequate*.

Definition II.2 (Network capacity). The network capacity (V_{Cap}) is defined as the maximum service demand volume such that $\bar{\boldsymbol{\alpha}} \leq \mathbf{1}$. Thus, $V_{\text{Cap}} \triangleq \max V : \bar{\boldsymbol{\alpha}} \leq \mathbf{1}$.

Definition II.3 (Saturation point). The saturation point (V_{Sat}) is the minimum service demand volume such that $\bar{\boldsymbol{\alpha}} \geq \mathbf{1}$. Thus, $V_{\text{Sat}} \triangleq \min V : \bar{\boldsymbol{\alpha}} \geq \mathbf{1}$.

As indicated, different models can be used for modeling ICI. In this work, two models are considered: '*full load*' and '*load coupling*'. Recall that in full load, active cells are assumed to have full load, i.e., $\bar{\alpha}_l = 1$, if $\mathbf{x}(l) = 1$, and $\bar{\alpha}_l = 0$, if $\mathbf{x}(l) = 0$. In case of load coupling, the ICI created by each cell is proportional to its load. An iterative algorithm to estimate the cell load coupling is provided in Appendix A. Thus, the vector $\boldsymbol{\Psi} \in \mathbb{R}^A$ representing the average SINR at each area element is given by

$$\boldsymbol{\Psi} = [(\mathbf{S} \odot \mathbf{G}) \cdot (\mathbf{p}_{\text{D}} \odot \mathbf{x})] \oslash [[(\mathbf{S}^{\text{c}} \odot \mathbf{G}) \cdot (\mathbf{p}_{\text{D}} \odot \mathbf{x} \odot \bar{\boldsymbol{\alpha}})] \oplus \sigma^2]. \tag{7}$$

The operators $\oslash$ and $\oplus$ denote Hadamard (pointwise) operations. By means of (7), average SINR figures as function of the network topology ($\mathbf{x}$) are obtained. Since load levels also depend on SINR values, the load coupling generates a system of non-linear equations which have a unique non-negative ($\bar{\boldsymbol{\alpha}} \geq \mathbf{0}$) solution [29]. In order to estimate $\bar{\boldsymbol{\alpha}}$, let's look at the average SINR at area element level. The average SINR at $a \in \mathcal{A}_l$ can be expressed as

$$\psi(a) = \frac{\mathbf{p}_{\text{D}}(l) \cdot \mathbf{G}(a, l)}{\left(\sum_{j \in \mathcal{L} \setminus \{l\}} \bar{\alpha}_j \cdot \mathbf{p}_{\text{D}}(j) \cdot \mathbf{G}(a, j) \right) + \sigma^2}. \tag{8}$$

In (8), the ICI coming from neighbor BSs is proportional to their average loads ($\bar{\alpha}_j$'s). It is customary to define link performance in terms of $\psi(a)$ by means of a concave (e.g., logarithmic) function (f_{LP}) of it, such that $\gamma_a = f_{\text{LP}}(\psi(\mathbf{a}))$ [bps/Hz]. The bandwidth requirement of a single user in $a \in \mathcal{A}_l$ to satisfy the QoS can be obtained as

$$b_{\text{u}}(a) = \frac{r_{\min}}{f_{\text{LP}}(\psi(a))} \quad \text{[Hz]}, \tag{9}$$

and the average load ($\bar{\alpha}_l$) in the lth BS would be given by

$$\bar{\alpha}_l = \frac{1}{B_{\text{sys}}} \cdot N_{\text{u}}^l \cdot b_l, \tag{10}$$

where

$$N_{\text{u}}^l = \left(\sum_{a \in \mathcal{A}_l} \mathbf{\Gamma}(a) \right) \frac{\mathbb{E}\{\mu\}}{\mathbb{E}\{\lambda\}}, \tag{11}$$

and

$$b_l = \sum_{a \in \mathcal{A}_l} \left(\frac{\mathbf{\Gamma}(a)}{\sum_{k \in \mathcal{A}_l} \mathbf{\Gamma}(k)} \right) b_{\text{u}}(a) \quad \text{[Hz]}. \tag{12}$$

In (10), N_{u}^l and b_l are the average number of users and bandwidth consumption in BS l, respectively.

In order to take into account the coverage criteria and penalize solutions with coverage holes, i.e., a significant number of area elements without coverage, the spectral efficiency of the ath area element is stored in the vector $\mathbf{H} \in \mathbb{R}^A$ and it is computed according to the following rule: $h_a = \mathbf{v}(a) \cdot f_{\text{LP}}(\psi_a)$. The binary vector $\mathbf{v} \in \{0, 1\}^A$ indicates if the ath is out of coverage. Therefore, if the ath area element is in outage, $\mathbf{v}(a) = 1$, and 0 otherwise.

Finally, the list of symbols is provided in Table II.

III. METRICS, PROBLEM FORMULATION, AND SOLUTION

A. Multiobjective optimization: basics

In order to study the tradeoffs in CSO, the use of multiobjective optimization has been considered. Multiobjective optimization is the discipline that focuses on the resolution of the problems involving the simultaneous optimization of several conflicting objectives, and hence, it is a convenient tool to investigate CSO, where the two fundamental metrics, energy consumption and network capacity, are in conflict. The target is to find a subset of *good* solutions $\mathcal{X}^\star$ from a set $\mathcal{X}$ according to a set of criteria $\mathcal{F} = \{f_1, f_2, \cdots, f_{|\mathcal{F}|}\}$, with cardinality $|\mathcal{F}|$ greater than one. In general, the objectives are

TABLE II
BASIC NOTATION.

Symbol	Description	Symbol	Description
B	System bandwidth	$P_{\max}$	Maximum transmit power per cell
L	Number of BSs	$\mathcal{L}$	Set with the base stations' indexes
A	Number of area elements	$\mathcal{A}$	Set of area elements in the target area
$\mathbf{G}$	Path-loss matrix	$\mathbf{p}_{PS}, \mathbf{p}_{D}$	Power vectors: pilots and data channels
$\mathbf{R}_{PS}$	Received power matrix	$\mathbf{x}$	Network topology
$\mathcal{A}_{c}$	Coverage area ($\mathcal{A}_{c} \subseteq \mathcal{A}$)	$\mathcal{A}_{l}$	Coverage of the lth BS
$\mathbf{R}_{PS}$	Received power matrix	$\mathbf{S}, \mathbf{S}^{c}$	Coverage matrices, i.e., coverage of each cell
f_{c}	Connectivity function (cell selection)	f_{LP}	Link performance model
$P_{\min}$	Minimum received power	$\psi_{\min}$	Minimum SINR
$G_{\max}^{UL}$	Maximum path-loss	$r_{\min}$	Minimum target rate (QoS criterion)
λ	Inter-arrival time	μ	Session time
$\mathbf{\Gamma}$	Spatial demand distribution	$\bar{\boldsymbol{\alpha}}$	Average load vector
V	Service demand volume	$\mathbf{\Psi}$	Average SINR vector
$\mathbf{H}$	Spectral efficiency vector	κ_{UL}	Uplink fractional compensation
$\mathbf{v}$	Coverage vector	κ_{COV}	Coverage threshold
$\mathbf{n}$	Inverse of cell's size		

in conflict, and so, improving one of them implies worsening another. Consequently, it makes no sense to talk about a single global optimum, and hence, the notion of an optimum set $\mathcal{X}^{\star}$ becomes very important. A central concept in multiobjective optimization is Pareto efficiency. A solution $\mathbf{x}^{\star} \in \mathcal{X}$ has Pareto efficiency if and only if there does not exist a solution $\mathbf{x} \in \mathcal{X}$, such that $\mathbf{x}$ dominates $\mathbf{x}^{\star}$. A solution $\mathbf{x}_{1}$ is preferred to (dominates) another solution $\mathbf{x}_{2}$, ($\mathbf{x}_{1} \succ \mathbf{x}_{2}$), if $\mathbf{x}_{1}$ is better than $\mathbf{x}_{2}$ in at least one criterion and not worse than any of the remaining ones. The set $\mathcal{X}^{\star}$ of Pareto efficient solutions is called optimal nondominated set and its image is known as the Optimal Pareto Front (OPF). In multiobjective optimization, it is unusual to obtain the OPF due to problem complexity; instead, a near-optimal or estimated Pareto front (PF) is found. Readers are referred to [33] for an in-depth discussion.

B. Performance metrics

The following performance metrics have been considered[2]:

- *The number of active cells* (f_1). Under the full-load assumption, energy consumption is proportional to the number of active cells [22, 26]:

$$f_1 = \mathbf{x} \cdot \mathbf{1}. \tag{13}$$

- *Average network capacity* (f_2). This metric is based on the expected value of the spectral efficiency at area element level. Hence, the effect of the spatial service demand distribution ($\mathbf{\Gamma}$) must be considered. The metric is defined as follows:

$$f_2 = (B \cdot A) \left[\left[(\mathbf{H} \odot \mathbf{\Gamma})^{\text{T}} \cdot \mathbf{S} \right] \odot \mathbf{n} \right] \cdot \mathbf{1}. \tag{14}$$

The vector $\mathbf{H} \odot \mathbf{\Gamma}$ corresponds to the *weighted* spectral efficiency of each area element. The idea is to give more importance to the network topologies ($\mathbf{x}$'s) that provide better aggregate capacity (f_2) to the areas with higher service demand. In (14), A (the number of area elements) is used to normalize the obtained capacity to the uniform distribution

---

[2]In the definition of some metrics, the dependence with $\mathbf{x}$ is not explicit, however, it is important to note that all of them depend on $\mathbf{x}$, i.e., the network topology.

case, i.e., $\mathbf{\Gamma}(a) = 1/A$, $\forall a \in \mathcal{A}$. The vector $\mathbf{n} \in \mathbb{R}^{L}$ contains the inverse of the sum of each column in $\mathbf{S}$, i.e., the number of pixels served by each cell. It is assumed that each user is served by one cell at a time. This vector is used to distribute the capacity of each cell evenly over its coverage area, i.e., the bandwidth is shared equally by the area elements belonging to each cell. This improves the fairness in the long run similar to the proportional fairness policy that tends to share the resources equally among users as time passes. This fairness notion results in decreasing the individual rates as the number of users increases. This effect is also captured by $\mathbf{n}$ as the bandwidth per area element is inversely proportional to the size of the cell.

- *Cell edge performance* (f_3). The 5th percentile of the pixel rate Cumulative Distribution Function (CDF) is commonly used to provide an indicator for cell edge performance [34]. A vector with the weighted average rate at area element level can be obtained as follows:

$$\mathbf{r} = A \cdot (\mathbf{H} \odot \mathbf{\Gamma}) \odot \left[\mathbf{S} \cdot \left(\mathbf{n}^{\text{T}} \cdot \text{diag}(B) \right)^{\text{T}} \right]. \tag{15}$$

Then, the percentile 5 is given by

$$f_3 = \mathbf{r}'(0.05 \cdot A). \tag{16}$$

The vector $\mathbf{r}'$ is a sorted (ascending order) version of $\mathbf{r}$.

- *Uplink power consumption* (f_4). In order to provide an estimate of the uplink power consumption of any network topology, a fractional compensation similar to the Open Loop Power Control (OLPC) used in Long Term Evolution (LTE) is considered [35]. It is given by

$$f_4 = \frac{1}{\sum_{k \in \mathcal{A}_c} \mathbf{\Gamma}(k)} \cdot \sum_{l=0}^{L-1} \sum_{a \in \mathcal{A}_l} \mathbf{\Gamma}(a) \cdot (P_0 + \kappa_{UL} \cdot \mathbf{G}(a,l)), \tag{17}$$

where P_0 is a design parameter that depends on the allocated bandwidth and target Signal-to-Noise Ratio (SNR) and $\kappa_{UL} \in [0, 1]$ is the (network controlled) fractional compensation factor.

- *Load dependent power consumption* (f_5). In order to estimate the network power consumption under the load

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2625743, IEEE Access
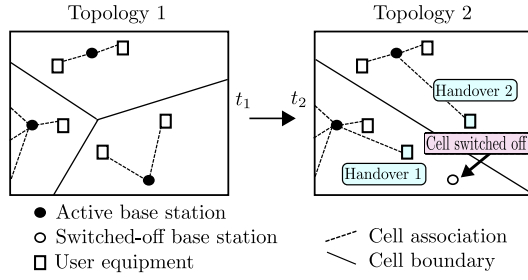
6

Fig. 1. Illustration of network-initiated handover due to CSO operation.

coupling assumption, the parameterized BS power model proposed in [36] has been used. Thus,

$$f_5 = \sum_{l=0}^{L-1} f_{\mathrm{PC}}^l(\bar{\alpha}_l), \qquad (18)$$

where $f_{\mathrm{PC}}^l(\bar{\alpha}_l)$ is a function that gives the power consumption in the lth BS as function of its load. Essentially, in this model, there is a fixed power consumption (P_0) that is independent of the load but that can be further reduced (till P_{CSO}) if the base station is switched-off. Moreover, there is a part that grows linearly with the load till a maximum power consumption ($P_{\max}$) that obviously contains the transmitted power over the air interface ($P_{\max}^{\mathrm{Tx}}$).

- *Load dispersion* (f_6). As it will be shown, load dispersion in load coupling conditions is an important parameter because it measures how well distributed the service demand is. In order to quantify this value, the Coefficient of Variation is considered. Thus,

$$f_6 = \frac{\mathrm{std}\{\bar{\boldsymbol{\alpha}}\}}{\mathrm{mean}\{\bar{\boldsymbol{\alpha}}\}}. \qquad (19)$$

- *Handovers.* In the context of CSO, handovers are a quite important concern [8]. Handovers are produced when users need to be associated to another base station because their serving cells are switched-off. In practice, handovers are mainly produced due to users mobility, but independently of the type, either user- or network-triggered, handovers require a certain time and signaling, both at the air interface and core network. Thus, the CSO operation should, as much as possible, minimize the number of handover, i.e., the transition from one topology to another should be done with the minimal impact and/or cost. A pictorial representation of the aforementioned situation is shown in Figure 1. Thus, handovers are considered herein as an important performance metric.

C. Multiobjective problem formulation

The multiobjective optimization problem considered herein can be formulated as follows:

$$\text{optimize } \mathbf{f}(\mathbf{x}) = [\, f_i(\mathbf{x})\,,\, f_j(\mathbf{x})\,], \qquad (20a)$$
$$\text{subject to:}$$
$$\left(A^{-1} \cdot \left(\mathbf{v}^{\mathrm{T}} \cdot \mathbf{1} \right) \right) \leq \kappa_{\mathrm{COV}}, \qquad (20b)$$
$$\mathbf{x} \in \{0,1\}^L, \ \mathbf{x} \neq \mathbf{0}. \qquad (20c)$$

Problem (20) proposes the simultaneous optimization of two of the previously introduced performance metrics as follows:

- Full load: if full load is assumed as model for intercell interference, $i = 1$ and $j \in \{2, 3, 4\}$ in (20a).
- Load coupling: if load coupling is assumed as model for intercell interference, $i = 5$ and $j = 6$ in (20a).

The previous optimization scheme allows to study and characterized the tradeoffs between conflicting metrics (see Section III-B) in a deployment-specific manner. Constraints (20b) and (20c) correspond to the coverage criterion and feasible set definition, respectively.

In general, solving multiobjective problems such as (20) is very difficult [33]. Indeed, (20) is a combinatorial problem that belongs to the class NP-complete, and hence, optimal solutions cannot be found in polynomial time. The domain (search space) defined by the optimization variable ($\mathbf{x}$, the on/off pattern) is a set of size $2^L - 1$, where L is the number of BSs. The objective space (or *image*) is defined by the objective functions, and due to their mathematical structure, it is highly non-linear, non-convex, and full of discontinuities and local optima [37]. Certain algorithms such as Simplex [38] are susceptible to be trapped in local optima, while other optimization techniques, such as Sequential Quadratic Programming [39], require convexity to guarantee convergence. Moreover, traditional constrained optimization, in which only one objective function is optimized subject to a set of constraints on the remaining ones, has the drawback of limiting the visibility of the whole objective space. For this reason, *heuristic*-based algorithms are popular approaches in CSO as it was seen in Section I, but unfortunately, by means of this type of solutions it is very difficult to address multiobjective optimization problems. In order to overcome this difficulty, the use of Multiobjective evolutionary algorithms (MOEAs) [40] is proposed herein as described next.

D. Multiobjective evolutionary algorithms

As it was mentioned, heuristic solutions are usually problem-specific and typically used for single-objective optimization. Thus, the so-called 'metaheuristics' have become an active research field [37]. Metaheuristics can be used to solve very general kind of multiobjective optimization problems, such as the CSO formulation presented herein. Indeed, (20) requires a tool able to 1) find good solutions by efficiently exploring the search space, and 2) operate efficiently with multiple criteria and a large number of design variables. In addition, it should not have strong requirements, such as convexity or continuity. Multiobjective evolutionary algorithms (MOEAs) [40] fulfill the previous goals, and hence, their use is proposed to deal with the CSO framework presented herein. MOEAs are population-based metaheuristics that simulate the process of natural evolution and they are convenient due to their *black-box* nature that requires no assumption on the objective functions.

Thus, the Nondominated Sorting Genetic Algorithm II (NSGA-II) [41] is employed herein to solve (20). NSGA-II is accepted and well-recognized as a reference in the field of evolutionary optimization as it has desirable features, such as *elitism* (the ability to preserve good solutions), and mechanisms to flexibly improve *convergence* and *distribution*.

Further details can be found in [40]. One key insight for selecting evolutionary (genetic) algorithms is that, in CSO, a certain correlation is expected among network topologies that are suitable for a given spatial service demand distribution, i.e., they are expected to be similar, with more cells where the traffic is concentrated. The operation in evolutionary algorithms precisely does that, i.e., once a good (Pareto efficient) solution (network topology) is found, the algorithm iteratively try to improve it by 1) combining it with other good solutions (*crossover* mechanism), and 2) adding random minor variations to them (*mutation* mechanism). The complete description of NSGA-II can be found in [41]. As it will be shown, the use of MOEAs provides a quite convenient approach to CSO. However, depending on the scale of the problem, convergence can be slow, especially if computational resources are limited. Thus, based on the insight previously indicated, and in order to provide additional possibilities, Algorithm 1 is also proposed for solving (finding the set $\mathcal{X}^{\star}$ of Pareto efficient solutions) a particular, yet important, case of (20); when the number of active cells (f_1) and the average network capacity (f_2) need to be jointly optimized. Given that the need for minimizing the number of transitions is very important from a practical point of view, Algorithm 1 aims at finding a collection of network topologies, all with different number of active BSs, featuring 1) the minimum distance property, and 2) acceptable performance. In this context, the word *distance* refers to the Hamming distance (d_{H}), i.e., the number of positions in which the corresponding symbols in two different solutions are different. In this manner, for two solutions $\mathbf{x}_i$ and $\mathbf{x}_j$ in a set $\mathcal{X}^{\star}_{\mathrm{MD}}$ featuring the minimum distance property, $d_{\mathrm{H}}(\mathbf{x}_i, \mathbf{x}_j) = 1 \Rightarrow |(\mathbf{x}_i \cdot \mathbf{1}) - (\mathbf{x}_j \cdot \mathbf{1})| = 1$ always holds. Initially, Algorithm 1 determines the best topology with 1 active BS ($\mathbf{x}_1$) in line 2. Then, in lines 4-14, for each successive number of active cells (NAC = 2, ..., L), the algorithm sequentially finds the BSs that should be activated (resulting in the solution $\mathbf{x}_j$), such that 1) the Hamming distance with the previous solution $\mathbf{x}_{j-1}$ is one, and 2) the function f_2 is maximized. Thus, each solution added to $\mathcal{X}^{\star}_{\mathrm{MD}}$ provides the biggest increment in terms of f_2 with respect to the one previously added, and only one off/on transition is required. It should be noted that, although not explicitly indicated, Algorithm 1 indeed optimizes not only the number of active base stations and the network capacity, but also the number of transitions when moving from one topology to another. Thus, more than two objetives are jointly considered. The same applies for (20), i.e., more than two metrics could be considered, at expense of an increase in complexity, but it should be taken into account that in the context of CSO, a Pareto Front in more than two dimensions could complicate the implementation.

E. Conceptual design and implementation

Figure 2 illustrates the conceptual design of the proposed multiobjective framework. The framework relies on having a statistical description of the behavior of the service demand (in time and space). Thus, by means of different traffic distributions ($\mathbf{\Gamma}_x$), the spatial component of the traffic at

---

Algorithm 1: Minimum Distance Algorithm (MDA).

input : $\mathcal{X}_1$: $\mathcal{X}_1 = \{\mathbf{x} \in \mathcal{X} \mid \mathbf{x} \cdot \mathbf{1} = 1\}$, $|\mathcal{X}_1| = L$.
output : $\mathcal{X}^{\star}_{\mathrm{MDA}}$: A set of L network topologies.

1 $C^{\star} \leftarrow 0$; $\mathcal{X}^{\star}_{\mathrm{MD}} \leftarrow \emptyset$;
2 $\mathbf{x}_1 \leftarrow \texttt{BestBS}(\mathcal{X}_1)$;
3 $\mathcal{X}^{\star}_{\mathrm{MD}} \leftarrow \mathcal{X}^{\star}_{\mathrm{MD}} \cup \{\mathbf{x}_1\}$;
4 **for each** $j = 2 : L$ **do**
5     $C^{\star} \leftarrow 0$;
6   **for each** $\mathbf{x} \in \mathcal{X}_j \mid d_H(\mathbf{x}, \mathbf{x}_{j-1}) = 1$ **do**
7        $C_{\mathbf{x}} \leftarrow f_2(\mathbf{x})$;
8   **if** $C_{\mathbf{x}} > C^{\star}$ **then**
9           $C^{\star} \leftarrow C_{\mathbf{x}}$;
10           $\mathbf{x}_j \leftarrow \mathbf{x}$;
11   **end**
12   **end**
13     $\mathcal{X}^{\star}_{\mathrm{MD}} \leftarrow \mathcal{X}^{\star}_{\mathrm{MD}} \cup \{\mathbf{x}_j\}$;
14 **end**
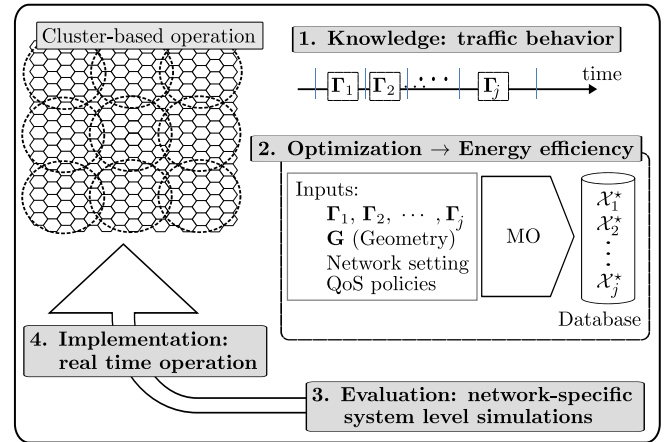15 **return** $\mathcal{X}^{\star}$;

Fig. 2. Conceptual design of the MO framework for CSO.

different moments of the day can be captured. These patterns can be considered fairly constant during time intervals of *small* duration (tens of minutes or few hours) [24, 27]. Starting from the knowledge of a given $\mathbf{\Gamma}$, network analysis and optimization based on (20) is done offline. The main idea is that, for different demand conditions (spatial distribution and volume), different sets of Pareto efficient network topologies can be obtained, i.e., for each $\mathbf{\Gamma}_x$, there is a corresponding $\mathcal{X}^{\star}_x$. These sets of near-optimal solutions ($\mathcal{X}^{\star}_x$'s) can be evaluated by means of system level simulations (in which several QoS criteria, scheduling policies, and ICI models can be considered independently) in order to determine which network topologies ($\mathbf{x} \in \mathcal{X}^{\star}_x$) provide the desired level of QoS. Obviously, the network operator may act rather conservatively in this selection process as it will be explained in Section IV. Moreover, in order to allow for semi-distributed implementation, a cluster-based operation is encouraged. The benefit of doing so is twofold. First, the demand in relatively small areas covered by small cells (e.g., pico-cells in a university campus) can be characterized easily. Second, the amount of intercell coordination is reduced compared with the schemes aiming at operating in large urban areas. Since demand profiles are stored and indexed at coordinating points in each cluster, the amount of data that need to be exchanged (from time to time) is negligible.
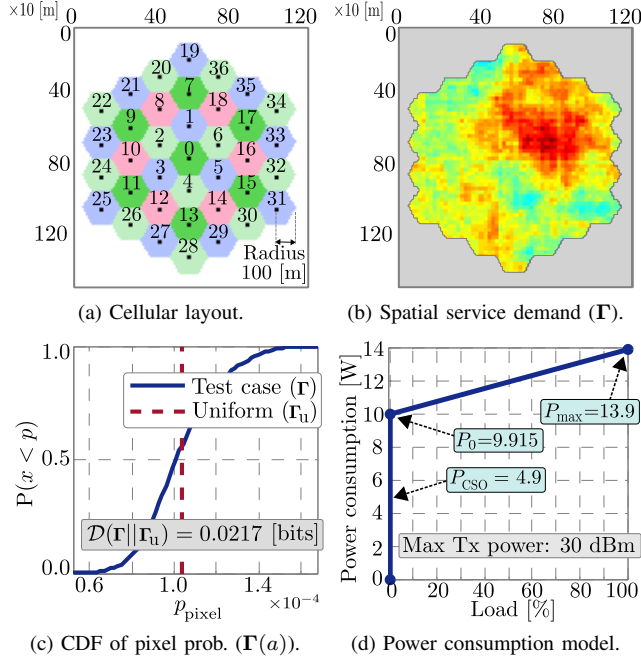
(a) Cellular layout.

(b) Spatial service demand ($\mathbf{\Gamma}$).

(c) CDF of pixel prob. ($\mathbf{\Gamma}(a)$).

(d) Power consumption model.

Fig. 3. Test case scenario.

(a) Coverage maps.

(b) Coverage vs. Tx power.

(c) Detectable cells.

(d) Candidate servers.

Fig. 4. Coverage aspects: impact of transmit power.

Instead, different clusters (a certain amount of overlapping can be allowed) can also share information in longer time scales, so that better decisions can be made in boundary cells. In any case, the idea of identifying traffic profiles and applying multiobjective-optimized on/off patterns, is compatible with novel paradigms that 1) are being considered for 5G (*cloud-networking* and *virtualization* [2, 5]) and 2) would allow for more dynamic and centralized operation. In addition, several research contributions in the increasingly research field of service demand modeling and pattern recognition [42–44] (a research problem out of our scope) are appearing, and hence, the method proposed herein can extensively be benefit from that activity.

IV. PERFORMANCE EVALUATION

A. Simulation conditions and parameters

The simulation setup is based on the assumptions for evaluating the IMT-Advanced systems [45]. The urban micro-cell (UMi) downlink scenario was chosen. Fig. 3a shows the corresponding cellular layout. As it can be seen, the network is composed of 37 small cells (radius = 100 m, network area $\approx 1\,\mathrm{km}^2$). Fig. 3b corresponds to the (irregular) spatial service demand distribution ($\mathbf{\Gamma}$) used in the numerical examples. The Kullback-Leiber distance $\mathcal{D}$ with respect to the uniform distribution ($\mathbf{\Gamma}_\mathrm{u}$) can be used as a measure of the non-uniformity of the spatial service demand distribution as it is shown in Figure 3c. This setting is perfectly valid to study CSO as in this context gains are obtained from the mismatch between demand and supply. Indeed, CSO is about finding the smallest network topology that is *compatible* [46] enough with the service demand to provide the required QoS and coverage.

In this study, a Single-Input Single-Output (SISO) is considered. This assumption does not imply any loss of generality as long as all cells use the same scheme. The load-dependent
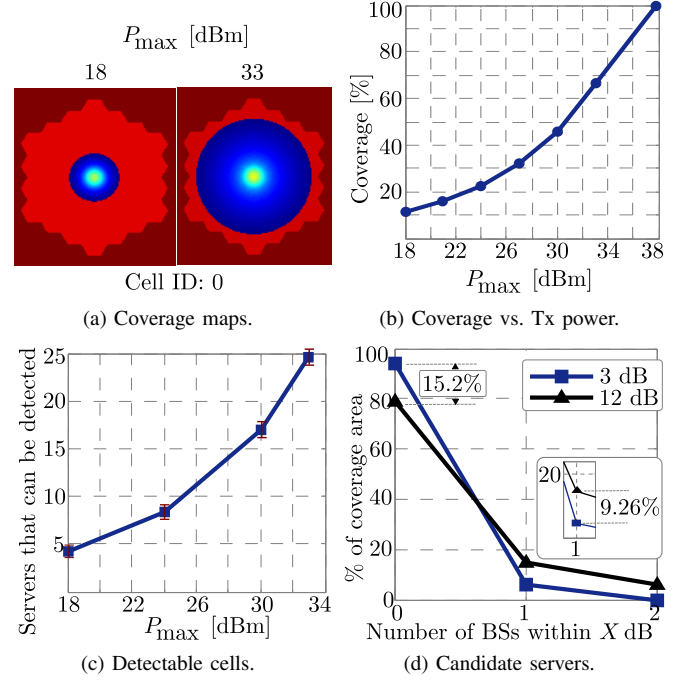
power consumption model, based on the parameters given in [36], for pico-BSs assuming an operating bandwidth of 5 MHz and a maximum transmission power of 30 dBm is shown in Fig. 3d.

Dynamic system level simulations are carried based on Monte Carlo experiments. The results compile statistics taken from 100 independent experiments each of which has a duration of 5400 s. At each cell, the scheduler assigns each user with a bandwidth such that the target rate (r_min) is satisfied. If the percentage of users that obtain a rate equal to r_min is greater or equal to the operator-specific target QoS (Q), then the QoS policy is said to be fulfilled. Thus, in order to satisfy the maximum number of users, users are sorted based on their spectral efficiency and served accordingly. When there is not enough bandwidth to satisfy a user, the resource allocation ends. The set of parameters used in simulations is provided in Table III. Calibration and complexity aspects of NSGA-II are briefly discussed in Section IV-F, and additional guidelines can be found in [34, 41]. The experimentally obtained setting is also shown in Table III.

B. Coverage aspects

The first part of this section is devoted to illustrate some coverage aspects and provides insights into the potential impact of the transmit power on the performance of CSO. Fig. 4a provides a qualitative perspective. The figure shows the size of the maximum coverage (points in which the received PS power is greater than P_min) for the central BSs ($l = 0$) for two different transmit powers ($P_\mathrm{max} = 18\,\mathrm{dBm}$ and $P_\mathrm{max} = 33\,\mathrm{dBm}$). For the sake of clarity, shadowing is not considered. A quantitative description is shown in Fig. 4b which indicates the percentage of the target area ($\mathcal{A}$) that can be covered with different values of P_max. Note for instance

TABLE III
EVALUATION SETTING AND PARAMETERS.

General setting					
Number of cells (L)	37 (wraparound, omni)	Carrier freq.	2.140 GHz	Bandwidth (B)	5 MHz
Max. BS transmit power (P_{max}^{Tx})	30 dBm	Pixels' resolution	5×5 m^2	Path loss	M.2135 Umi [45]
BS's height	15 m	Noise power (σ^2)	-174 dBm/Hz	Shadowing	$\mathcal{N}(0.4)$ [dB]
Cell selection (f_c)	Highest Rx. power	Rx. power (P_{min})	-123 dBm	Small scale fad.	As in [45]
Link performance (f_{LP})	Shannon's formula	Frac. comp. (κ_{UL})	1.00	Cov. (κ_{COV})	0.02
Max. path loss (G_{max}^{UL})	163 dB	Min. rate (r_{min})	400 kbps	Target QoS (Q)	97.5%
User distribution	According to $\mathbf{\Gamma}$	Traffic model	Full buffers	SINR (ψ_{min})	-7.0 dB
QoS checking interval	1 s				
Calibration of NSGA-II					
Population size	100	Crossover. prob.	1.00	Type of var.	Discrete
Mutation prob.	$1/L$	Termination crit.	Hypervolume $< 0.001\%$, [33]		

that, starting from 18 dBm ($15\,\%$ of coverage), P_{max} need to be increased more than eight times (up to 30 dBm) to double the coverage (up to $30\,\%$), while reaching $60\,\%$ of coverage requires less than four times the power required for $30\,\%$ of coverage. Obviously, this depends on the propagation model, but the message is that this analysis should be taken into account during the design phase of any CSO strategy in order to determine appropriate values for P_{max}. In the results shown in Figs. 4c and 4d, all the cells are active and transmit at the same P_{max}. Fig. 4c indicates the average number of BS that can be *detected* as a function of P_{max} (the average is taken over the whole coverage area). Fig. 4d shows the percentage of the coverage area in which x BSs (servers) are *heard* with a quality (SINR) within X dB below the one of the best server. From these results, it becomes clear that the choice of P_{max} has a big influence on the size of the feasible set in (20), i.e., the set of $\mathbf{x}$'s for which Constraint 20b is fulfilled. Hence, the impact of P_{max} is significant, mainly in low load conditions.

C. Estimation of network topologies

First, the results regarding the solution of (20) for the objectives functions introduced in Subsection III-B (f_1, f_2, f_3, and f_4) are provided. Fig. 5a shows the resulting Pareto Front by solving (20), when $i = 1$ and $j = 2$ in (20a), i.e., the joint optimization of the number of active BS (f_1) and the average network capacity (f_2), by means of MOEAs (algorithm NSGA-II) and Algorithm 1. As expected, the use of evolutionary optimization provides *better* solutions than Algorithm 1, i.e., greater values of f_2 for the same value of f_1. However, it is important to recall that the solutions obtained through Algorithm 1 feature the minimum distance property (see Section III-C), and that, Algorithm 1 ($\mathcal{O}(L^2)$) is, in case of small-to-moderate cluster size, less complex than NSGA-II ($\mathcal{O}(N^2 \cdot |\mathcal{F}|)$, N: population size). A quantitative perspective of such performance gap is shown in Fig. 5b. The blue/circle pattern corresponds to the gain in terms of f_2 for each value of f_1 indicated in the left vertical axis as 'Average capacity gain'. As a result of the combinatorial nature of NSGA-II, the gains are higher when network topologies are composed of less BSs, i.e., small values of f_1. The red/square pattern shows the capacity gain per cell, indicated in the right vertical axis. It can be seen that the gain of using MOEA is around 1 Mbps/cell in topologies with less than 20 active BSs ($f_1 \leq 20$). Hence, the use of MOEAs implies

(a) Capacity: MOEA vs. MDA.

(b) Gains: MOEA vs. MDA.

(c) Cell edge performance.
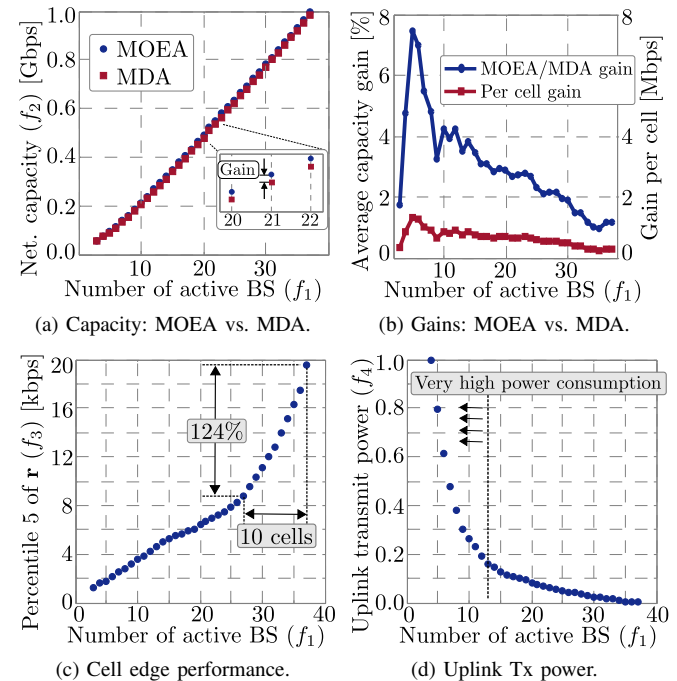

(d) Uplink Tx power.

Fig. 5. Multiobjective optimization results.

better network topologies in cases where the computational complexity can be afforded. The resulting Pareto Front by solving (20), for ($i = 1$, $j = 3$) and ($i = 1$, $j = 4$) in (20a), are shown in Figs. 5c and 5d, respectively. The first case illustrates the impact of CSO on cell edge performance. Note that while Fig. 5a shows a fairly linear growth of the average network capacity with the number of active cells, Fig. 5c indicates that cell edge performance (represented by f_3) is substantially improved only by network topologies featuring a higher number of active cells ($f_1 \geq 27$). This result clearly suggests that mechanisms for Intercell Interference Coordination (ICIC) should be applied together with CSO in cases of low load conditions to improve the QoS of cell edge users. Fig. 5d illustrates the impact of CSO on the power consumption of users (uplink). As it was mentioned, the goal is not to determine exact uplink power consumption figures, but to create means for comparison among network topologies with different number of active BSs. Thus, a normalized version of f_4 (see 17) is considered. As it can be seen, it turns out that the relationship between the number of

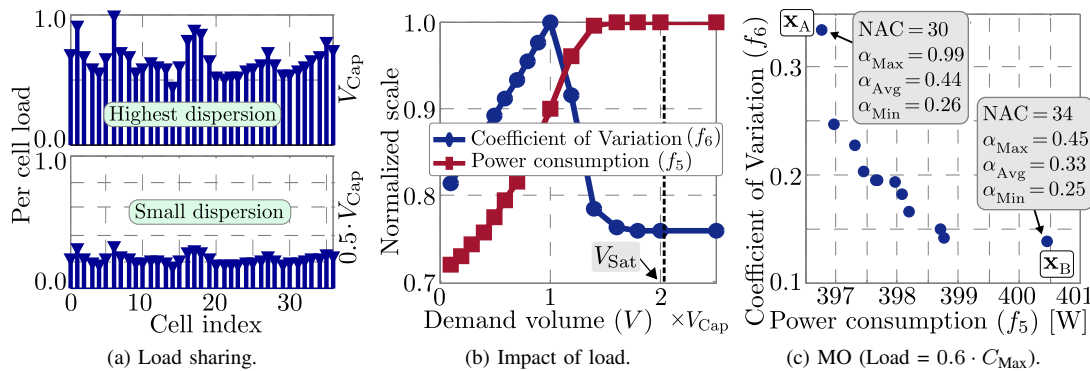(a) Load sharing.　　　　(b) Impact of load.　　　　(c) MO (Load $= 0.6 \cdot C_{Max}$).

Fig. 6.　Analysis considering cell coupling and load-dependent power consumption.

active BSs and the resulting uplink (open-loop-based) power consumption is highly nonlinear, being the energy expenditure considerably high in sparse network topologies ($f_1 < 15$). Hence, in scenarios where the lifetime of devices should be maximized (sensor networks), the use of CSO is not clear. Recall that uplink link budget is also considered as a coverage criterion.

To close this subsection, Fig. 6 shows the results corresponding to the solution of (20) for the objective functions introduced in Section III-B (f_5 and f_6). According to Definition II.2, and given the spatial demand distribution $\mathbf{\Gamma}$ (see Fig. 3b), $\mathbb{E}\{\lambda\} = 115.0$ ms and $\mathbb{E}\{\mu\} = 119.2$ s yield a demand volume (V) equal to V_{Cap}. The resulting load sharing patterns (obtained by means of Algorithm 2) for $V = V_{Cap}$ and $V = 0.5 \cdot V_{Cap}$ are shown in Fig. 6a. Note that increasing V results in higher load dispersion. To quantify this, Fig. 6b shows the impact of V on the Coefficient of Variation (CV) of the loads (f_6). The associated load-dependent power consumption (f_5) is also indicated. Note that f_5 and f_6 are maximized when $V = V_{Sat}$ and $V = V_{Cap}$, respectively. As expected, the load dependent power consumption (f_5) is maximized when $\bar{\alpha} \geq 1$, i.e., $V \geq V_{Sat}$. The dependence of f_6 on V is explained by the strong nonlinearity of (10) and the fact that, from the load-coupling point of view, $\bar{\alpha} \leq 1$, and hence, no change is expected after $V = V_{Sat}$. The results shown in Figs. 6a and 6b are obtained for $\mathbf{x} = \mathbf{1}$, i.e., when all the BSs are active. The joint optimization of f_5 and f_6 is shown in Fig. 6c. As it can be seen, there is a conflicting relationship between them. The attributes of the *extreme* solutions ($\mathbf{x}_A$ and $\mathbf{x}_B$) in the Pareto Front are indicated. There is also a certain correlation between the objectives (f_5 and f_6) and the number of active cells (NAC). The topology with the lowest energy consumption (f_5) requires less active BSs but it has the highest load dispersion (f_6). Note the difference between the highest and lowest loaded BS in $\mathbf{x}_A$. In contrast, the best load balancing ($\mathbf{x}_B$) involves more active BSs, and hence, worst values of f_5. A comparison among solutions obtained through each ICI model, FL and LC, is provided next.

D. System level simulations

As indicated earlier, solving (20) results in a set of Pareto efficient (nondominated) network topologies that are specific for either a spatial service demand distribution ($\mathcal{X}_{FL}$: full-load) or a service demand conditions, i.e., spatial demand distribution plus volume ($\mathcal{X}_{LC}$: load-coupling). Recall that $\mathcal{X}_{FL}$ is obtained by joint optimizing f_1 and f_2 in (20) for a given spatial demand distribution ($\mathbf{\Gamma}$), while obtaining $\mathcal{X}_{LC}$ involves the joint optimization of f_5 and f_6 in (20) for a given $\mathbf{\Gamma}$ and V (volume). Note that, the 'full-load' analysis is volume-independent, and hence, it does not require specify V (full load is assumed for the active cells). Thus, in order to evaluate these solutions by means of system level simulations, it is initially assumed that at each QoS checking interval (evaluation parameters are shown in Table III), the (nondominated) network topologies of each set ($\mathcal{X}_{FL}$ and $\mathcal{X}_{LC}$) are all applied and evaluated. The goal is to create QoS statistics for each network topology and load condition. Then, the network topology that is able to provide the desired QoS ($Q\%$ of users are satisfied $Q\%$ of time) is selected and applied (as indicated in Subsection III-E). The comparative assessment is shown in Fig. 7, where the legends indicate the set the applied network topology belongs to ($\mathcal{X}_{FL}$ or $\mathcal{X}_{LC}$) and the ICI model (FL or LC) used in the system level trials. Fig. 7a shows the load-dependent power consumption of each network topology. Clearly, from the CSO point of view, the topologies in $\mathcal{X}_{LC}$ result in lower power consumption as they feature less active BSs (NAC is indicated in green boxes) given that the load-coupling model predicts better SINR than full-load (see Fig. 9a), and hence, network capacity is favored. However, as V increases, both models become somehow equivalent as the loads tend to 1; as a result, the energy consumption is quite similar. Figs. 7b and 7c show the QoS level (in terms of the number of satisfied users) that is obtained with the selected solution of each set for $V = 0.2 \cdot V_{Cap}$ and $V = 0.6 \cdot V_{Cap}$, respectively. The results make evident that the performance of the network topologies in $\mathcal{X}_{LC}$ is severely degraded if the ICI levels become higher than the ones from which they were calculated for, see $\mathcal{X}_{LC}$@FL (full ICI). Indeed, the performance of these solutions is sensitive to variations from the mean values (that happens when considering snapshots) in moderate-to-high load conditions, even when the load-coupling based ICI is considered, as seen in Fig. 7c for $\mathcal{X}_{LC}$@LC. On the other hand, the network topologies in $\mathcal{X}_{FL}$ provide consistent performance when they are evaluated under
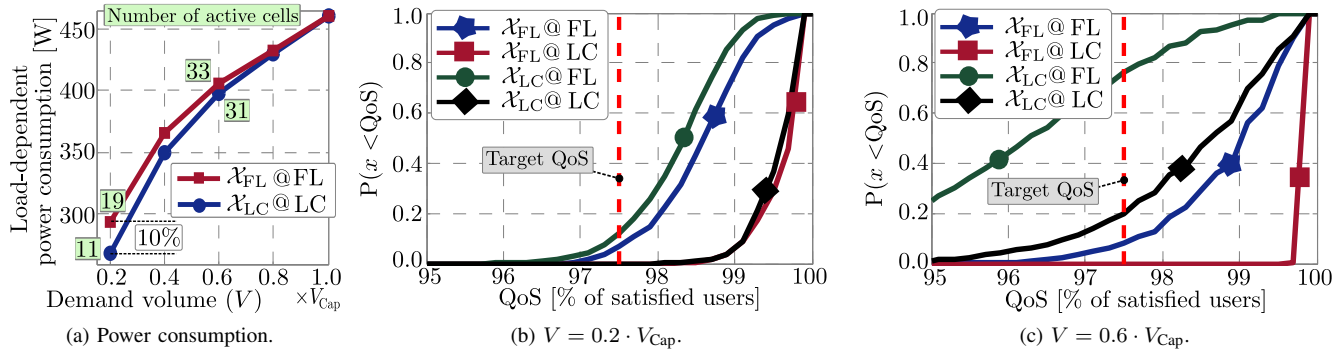
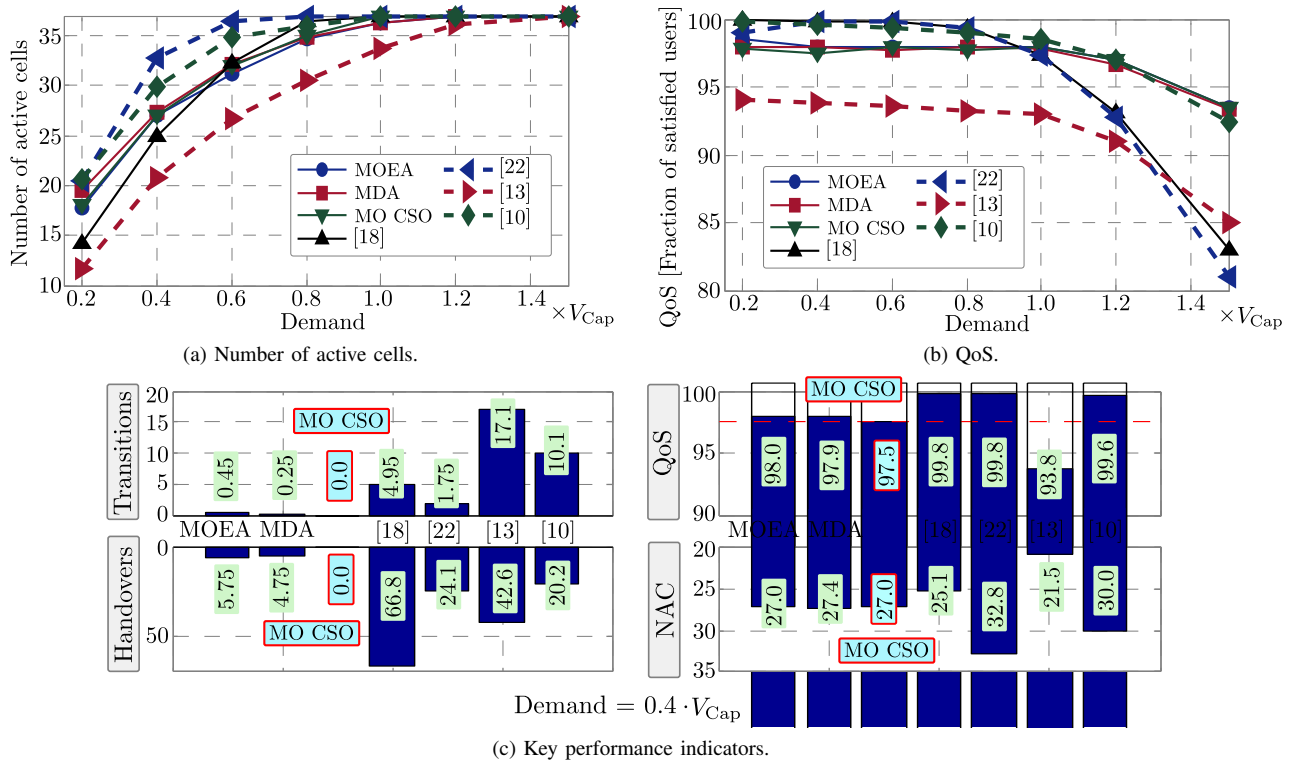Fig. 7. Comparative analysis: Load-coupling vs. Full ICIC MO optimization.

Fig. 8. Performance comparisons.

full load ($\mathcal{X}_{\mathrm{FL}}$@FL), and obviously, provide an even better performance under load-coupling ($\mathcal{X}_{\mathrm{FL}}$@LC) for both demand volumes. Hence, given that the energy consumption gain is in the order of 10% in the best case, it can be concluded that the full load model provides a competitive and somehow *safer* energy-saving vs. QoS tradeoff in the context of CSO. The proposed CSO scheme can use either approach. Summarizing:

✓ The MO for FL, i.e., f_1 and f_2 in (20), is volume-independent; offline system level simulations are required for each load condition (V), and energy saving is smaller in comparison to LC.

✓ In MO for LC, i.e., f_5 and f_6 in (20), is volume dependent; different offline optimization procedures are required for each load condition (V), and energy saving is larger in comparison to FC.

E. Performance comparison

In order to provide a wide perspective of the merit of the CSO framework presented herein, several recent/representative CSO schemes have been used as baselines. Obviously, an exhaustive comparison is not feasible. However, the idea is to illustrate some *pros* and *cons* of different approaches and the impact of some design assumptions. The following benchmarks are considered:

✓ *Cell zooming*: It was proposed in [26]. The idea is to sequentially switch-off BS starting from the lowest loaded. The algorithm ends when a cell cannot be switched-off because at least one user cannot be re-allocated.

✓ *Improved cell zooming*: This scheme is presented in [22] and it is similar to the one in [26], but it includes a more flexible termination criterion that allows to check more cells before terminating, and so, more energy-efficient topologies

can be found.

- ✓ *Load-and-interference aware CSO*: The design of this CSO scheme presented in [17] takes into account both the received interference and load of each cell to create a ranking that is used to sequentially switch-off the cells whose load is below a certain threshold.
- ✓ *Set cover based CSO*: The CSO scheme proposed in [14] relies on the idea of switch-on BS sequentially according to a certain sorting criterion. In this work, the sorting criterion is based on the number of users a cell can served in the SNR regime.

The performance comparison is shown in Fig. 8. To make the comparison fair, the full-load ICI conditions are considered. Figs. 8a and 8b show the average number of active cells and QoS (for different service demand volumes), respectively. As it can be seen, the best energy saving is obtained by [17], although at the expense of QoS degradations. This is due to the fact that in [17], users can be easily put in outage. In contrast, CSO schemes such as [14, 22, 26] provide the desired QoS (as long as $V \leq V_{\text{Cap}}$) since CSO decisions require associating all users. However, this results in an increment in the average number of active cells with respect to [17]. The schemes labeled as 'MDA' and 'MOEA' correspond to the (infeasible) dynamic selection of network topologies from the sets $\mathcal{X}_{\text{FL}}$ obtained through Algorithms 1 and NSGA-II, respectively, which are shown as reference. The performance of the proposed MO CSO is indicated by red boxes and labeled as 'MO CSO'. As it can be seen, the proposed scheme provides an excellent tradeoff between the required number of active cells and the obtained QoS, especially when $V \geq V_{\text{Cap}}$ where the performance (QoS) of other CSO is compromised. However, the most significant enhancement in the proposed scheme is its feasibility. Fig. 8c shows four performance indicators: transitions, handovers, QoS, and NAC. Given that the network topologies are calculated offline, they can be evaluated extensively by means of system level simulations (under a wide range of coverage criteria and conditions) to further guarantee their real-time performance, i.e., the operator can select topologies with more active cells rather than the ones which strictly need to guarantee QoS. Therefore, the selected network topologies can be applied (without real-time complexity) during periods of time in which service demand is described by Γ; as a result, no transitions or handovers are induced due to CSO. Hence, feasible yet effective CSO performance is achieved. As it was shown earlier, the proposed framework is generic, flexible, and no assumption are made in regards to, for instance, the cellular layout or objective functions; as a result, the framework is also suitable for small-cell deployments where irregular topologies and heterogeneous demand conditions are expected.

F. Complexity analysis and calibration aspects

To close this section, a complexity overview of the optimization algorithms is provided. According to [41], the complexity of NSGA-II is $\mathcal{O}(M \cdot N^2)$, where M and N correspond to the population size and the number of objective functions, respectively. In our case, $N = 2$ and M can be set depending on the scale of the problem. However, there is a consensus about the size of the population when using genetic algorithms, such as NSGA-II, and it is considered that during calibration populations of 20 up to 100 individuals can be used. Values greater than 100 hardly achieve significant gains and the same global convergence is obtained [47]. Regarding Algorithm 1, it's complexity is $\mathcal{O}(L^2)$, where L is the number of cells in the network. In practice, $L^2 \ll M \cdot N^2$ which is a significant reduction in terms of complexity that comes at expense of some performance. In evolutionary algorithms, a termination criterion is usually defined/need. One metric used to measure the level of convergence is the the *hypervolume* indicator [40]. It reflects the size of volume dominated by the estimated Pareto Front. In this work, the seacrh is terminated if the improvement in the hypervolume is smaller than a threshold (0.001%) after a certain number of generations (in this study, 20). Finally, crossover and mutation probabilities are set to 1 and $1/L$ (one mutation per solution, on average), respectively, as indicated in Table III.

V. CONCLUSIONS AND RESEARCH DIRECTIONS

CSO is a promising strategy that allows significant energy saving in cellular networks where both radio access network (capacity supply) and service demand are heterogeneous. In this article 1) CSO has been carefully analyzed considering coverage criteria, ICI models, and practical aspects, such as network-initiated handovers and on-off/off-on transitions, and 2) a novel MO-based CSO scheme has been introduced. The proposed solution succeeds in minimizing the number of transitions and handovers caused by the CSO operation and it is able to operate without need for heavy computational burden as the core processing is done offline. In addition, a cluster based-operation have been proposed to allow for semi-distributed implementation. The results show that, when compared with previous proposals, the proposed solution provides competitive performance in terms of QoS and energy saving while offering clear advantages from the feasibility perspective as it reduces the number of handovers and transitions. The results also highlight the importance of considering coverage criteria (in downlink and uplink) and pay attention to the selection of operational parameters, e.g., the power allocated to PS (typically used as criterion for coverage).

A comparative analysis between ICI models (full-load and load-coupling) indicates that the full-load assumption is a *safe* approach in the context of CSO as it provides natural protection against deviations from average load values that are 1) used as input of the algorithm, and 2) inherent of real time operation, i.e., discrete realizations of users. The impact of CSO on the power consumption of UE has also been studied. The results indicate that sparse topologies (few active BSs) have a significant impact on uplink power consumption, and hence, CSO is not suitable for scenarios with energy-sensitive devices such as sensor networks.

Research on topology adaptation has still a long way until its maturity. Feasible and effective techniques for traffic pattern recognition to complement CSO are still in infancy. It is our strong belief that CSO, as a promising approach to *greener* networks, is a key piece of a more general set of capabilities

---

Algorithm 2: Iterative approximation of cells load.

Inputs : All relevant information and ϵ (termination).
Output : $\bar{\boldsymbol{\alpha}} \in \mathbb{R}_+^{|\mathcal{L}|}$: Load vector.

1   $\bar{\boldsymbol{\alpha}}^0 \leftarrow \mathbf{1}; \; \boldsymbol{\epsilon} \leftarrow \epsilon \cdot \mathbf{1}; \; k \leftarrow 0;$ /* Initializing */
2   **repeat**
3     $k \leftarrow k+1;$
4     $\bar{\boldsymbol{\alpha}}^* \leftarrow \bar{\boldsymbol{\alpha}}^{k-1};$
5   **for** $l = 1 : |\mathcal{L}|$ **do**
6       $\bar{\alpha}_l^k \leftarrow \texttt{Load}(\{\bar{\alpha}_j^{k-1} : (\forall j \in \mathcal{L}) \wedge (j \neq l)\});$
7       $\bar{\alpha}_l^{k-1} \leftarrow \bar{\alpha}_l^k;$ /* Fast update */
8   **end**
9     $\bar{\boldsymbol{\alpha}}^k \leftarrow [\bar{\alpha}_0^k \cdots \bar{\alpha}_{L-1}^k];$ /* Update: iteration k */
10 **until** $\boldsymbol{\epsilon} > \frac{|\bar{\boldsymbol{\alpha}}^k - \bar{\boldsymbol{\alpha}}^*|}{\bar{\boldsymbol{\alpha}}^*};$
11 **return** $\bar{\boldsymbol{\alpha}}^k;$ /* Return estimated load vector */

(a) Bounds of SINR distributions.

(b) Iterative load estimation.

Fig. 9. Impact of cell load coupling on SINR distributions.

that will appear in 5G networks, also including promizing and disruptive concepts, such as Downlink Uplink Decoupling (DUDe). DUDe, where user equipment can transmit and receive to and from different base stations, is indeed, a clear research direction from the perspective of CSO, where both uplink and downlink could be considered as *independent* networks.

ACKNOWLEDGMENT

## APPENDIX A
ITERATIVE APPROXIMATION OF CELL LOADS

In order to estimate the average load vector ($\bar{\boldsymbol{\alpha}}$), Algorithm 2 is proposed. Basically, the estimation of the average load at each cell ($\bar{\alpha}_l$) is refined through each iteration comprising Lines 3 to 9. In line 6, the function $\texttt{Load()}$ estimates each $\bar{\alpha}$, based on (10), from the values of previous iterations (where $\bar{\alpha}_l = \min\{1, \bar{\alpha}_l\}$) and the ones that have been just updated in the current iteration (this *fast* update is done in line 7). Fig. 9 illustrates the motivation and performance of Algorithm 2. Basically, the use of load coupling provides a more accurate estimation of ICI levels in the network as shown in Fig. 9a. Note that the use of Full ICIC represents a more *conservative* approach. Fig. 9b shows that Algorithm 2 only requires few iterations to converge and that this depends on the starting point, but in any case convergence is fast.

REFERENCES

[1] J. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, A. Soong, and J. Zhang, "What will 5G be?," *IEEE Journal on Selected Areas in Communications*, vol. 32, pp. 1065–1082, June 2014.

[2] A. Osseiran et al., "Scenarios for 5G mobile and wireless communications: the vision of the METIS project," *IEEE Communications Magazine*, vol. 52, pp. 26–35, May 2014.

[3] Z. Hasan, H. Boostanimehr, and V. Bhargava, "Green cellular networks: A survey, some research issues and challenges," *IEEE Comm. Surveys & Tutorials*, vol. 13, no. 4, pp. 524–540, 2011.
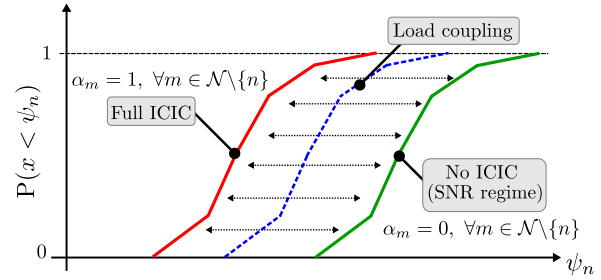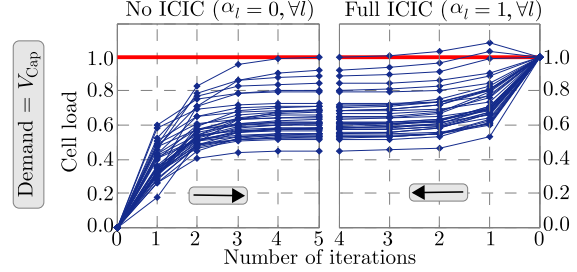
[4] I. Humar, X. Ge, L. Xiang, M. Jo, M. Chen, and J. Zhang, "Rethinking energy efficiency models of cellular networks with embodied energy," *IEEE Network*, vol. 25, no. 2, pp. 40–49, 2011.

[5] R. Schoenen and H. Yanikomeroglu, "User-in-the-loop: Spatial and temporal demand shaping for sustainable wireless networks," *IEEE Comm. Magazine*, vol. 52, pp. 196–203, Feb. 2014.

[6] M. Marsan, S. Buzzi, D. Ciullo, and M. Meo, "Switch-off transients in cellular access networks with sleep modes," in *2011 IEEE Int. Conference on Communications Workshops (ICC)*, pp. 1–6, 2011.

[7] A. Conte, A. Feki, S. Buzzi, D. Ciullo, M. Meo, and M. Marsan, "Cell wilting and blossoming for energy efficiency," *IEEE Wireless Communications*, vol. 18, no. 5, pp. 50–57, 2011.

[8] L. Saker and S. Elayoubi, "Sleep mode implementation issues in green base stations," in *2010 IEEE 21st Int. Symposium on Personal Indoor and Mobile Radio Communications*, pp. 1683–1688, 2010.

[9] S. Samarakoon, M. Bennis, W. Saad, and M. Latva-aho, "Dynamic clustering and on/off strategies for wireless small cell networks," *IEEE Transactions on Wireless Communications*, vol. 15, pp. 2164–2178, March 2016.

[10] C. Liu, B. Natarajan, and H. Xia, "Small cell base station sleep strategies for energy efficiency," *IEEE Transactions on Vehicular Technology*, vol. 65, pp. 1652–1661, March 2016.

[11] M. U. Jada, M. García-Lozano, and J. Hämäläinen, "Energy saving scheme for multicarrier HSPA + under realistic traffic fluctuation," *Mobile Networks and Applications*, vol. 21, no. 2, pp. 247–258, 2016.

[12] A. Antonopoulos, E. Kartsakli, A. Bousia, L. Alonso, and C. Verikoukis, "Energy-efficient infrastructure sharing in multi-operator mobile networks," *IEEE Communications Magazine*, vol. 53, pp. 242–249, May 2015.

[13] I. Aydin, H. Yanikomeroglu, and U. Aygolu, "User-aware cell switch-off algorithms," in *11th International Wireless Communications & Mobile Computing Conference (IWCMC)*, Aug. 2015.

[14] T. Beitelmal and H. Yanikomeroglu, "A set cover based algorithm for cell switch-off with different cell sorting criteria," in *2014 IEEE Int. Conference on Communications Workshops (ICC)*, pp. 641–646, June 2014.

[15] D. Gonzalez G., H. Yanikomeroglu, M. Garcia-Lozano, and S. R. Boque, "A novel multiobjective framework for cell switch-off in dense cellular networks," in *2014 IEEE Int. Conference on Communications (ICC)*, pp. 2641–2647, June 2014.

[16] C. Gao, W. Zhang, J. Tang, C. Wang, S. Zou, and S. Su, "Relax, but do not sleep: A new perspective on green wireless networking," in *2014 Proceedings IEEE INFOCOM*, pp. 907–915, April 2014.

[17] A. Alam, L. Dooley, and A. Poulton, "Traffic-and-interference aware base station switching for green cellular networks," in *2013 IEEE 18th Int. Workshop on Computer Aided Modeling and Design of Communication Links and Networks*, pp. 63–67, Sept 2013.

[18] H. Klessig, A. Fehske, G. Fettweis, and J. Voigt, "Cell load-aware energy

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2625743, IEEE Access

14

saving management in self-organizing networks," in *2013 IEEE 78th Vehicular Technology Conf. (VTC Fall)*, pp. 1–6, Sept 2013.

[19] C. Meng, X. Li, X. Lu, T. Liang, Y. Jiang, and W. Heng, "A low complex energy saving access algorithm based on base station sleep mode," in *2013 IEEE/CIC International Conference on Communications in China (ICCC)*, pp. 491–495, Aug 2013.

[20] Q. Fan, X. Chen, H. Yin, and R. Liu, "Smart switch: Optimize for green cellular networks," in *2013 IEEE Consumer Communications and Networking Conference (CCNC)*, pp. 701–704, Jan 2013.

[21] A. Yildiz, T. Girici, and H. Yanikomeroglu, "A pricing based algorithm for cell switching off in green cellular networks," in *2013 IEEE 77th Vehicular Technology Conf.*, pp. 1–6, June 2013.

[22] F. Alaca, A. B. Sediq, and H. Yanikomeroglu, "A genetic algorithm based cell switch-off scheme for energy saving in dense cell deployments," in *2012 IEEE Globecom Workshops*, pp. 63–68, 2012.

[23] T. Han and N. Ansari, "Optimizing cell size for energy saving in cellular networks with hybrid energy supplies," in *2012 IEEE Global Telecommunications Conference (GLOBECOM)*, 2012.

[24] C. Peng, S.-B. Lee, S. Lu, H. Luo, and H. Li, "Traffic-driven power saving in operational 3G cellular networks," in *ACM 17th Annual International Conference on Mobile Computing and Networking (MobiCom '11)*, pp. 121–132, 2011.

[25] T. Kang, X. Sun, and T. Zhang, "Base station switching based dynamic energy saving algorithm for cellular networks," in *2012 3rd IEEE Int. Conf. on Net. Infrastructure and Digital Content*, 2012.

[26] Z. Niu, Y. Wu, J. Gong, and Z. Yang, "Cell zooming for cost-efficient green cellular networks," *IEEE Communications Magazine*, vol. 48, no. 11, pp. 74–79, 2010.

[27] E. Oh and B. Krishnamachari, "Energy savings through dynamic base station switching in cellular wireless access networks," in *2010 IEEE Global Telecommunications Conference*, pp. 1–5, 2010.

[28] S. Zhou, J. Gong, Z. Yang, Z. Niu, and P. Yang, "Green mobile access network with dynamic base station energy saving," in *ACM 15th Annual International Conference on Mobile Computing and Networking (MobiCom '09)*, pp. 10–12, 2009.

[29] I. Siomina and D. Yuan, "Analysis of cell load coupling for LTE network planning and optimization," *IEEE Transactions on Wireless Communications*, vol. 11, no. 6, pp. 2287–2297, 2012.

[30] X. Zhou, Z. Zhao, R. Li, Y. Zhou, and H. Zhang, "The predictability of cellular networks traffic," in *2012 Int. Symposium on Communications and Information Technologies (ISCIT)*, pp. 973–978, Oct 2012.

[31] D. Lee, S. Zhou, X. Zhong, Z. Niu, X. Zhou, and H. Zhang, "Spatial modeling of the traffic density in cellular networks," *IEEE Wireless Communications*, vol. 21, pp. 80–88, February 2014.

[32] X. Wang, P. Krishnamurthy, and D. Tipper, "Cell sleeping for energy efficiency in cellular networks: Is it viable?," in *2012 IEEE Wireless Comm. and Networking Conf. (WCNC)*, pp. 2509–2514, 2012.

[33] Y. Sawaragi, I. Hirotaka, and T. Tanino, *Theory of Multiobjective Optimization*. Academic Press., 1st ed., 1985.

[34] D. González G., M. Garcia-Lozano, S. Ruiz, and D. S. Lee, "Optimization of soft frequency reuse for irregular LTE macrocellular networks," *IEEE Trans. on Wireless Communications*, vol. 12, pp. 2410–2423, May 2013.

[35] A. Simonsson and A. Furuskar, "Uplink power control in LTE - overview and performance, subtitle: Principles and benefits of utilizing rather than compensating for sinr variations," in *IEEE 68th Vehicular Technology Conference (VTC 2008-Fall)*, pp. 1–5, 2008.

[36] H. Holtkamp, G. Auer, V. Giannini, and H. Haas, "A parameterized base station power model," *IEEE Communications Letters*, vol. 17, pp. 2033–2035, November 2013.

[37] T. Weise, *Global optimization algorithms - Theory and application*. Self-Published, second ed., June 26, 2009. Online available at http://www.it-weise.de/.

[38] D. Gale, "Linear programming: The Simplex method," *Notices of the AMS*, vol. 54, pp. 364–369, March 2007.

[39] Gill, P. E. and Wong, E., "Sequential Quadratic Programming Methods," Tech. Rep. NA-10-03, Department of Mathematics, University of California, San Diego, La Jolla, CA, Aug 2010.

[40] C. A. Coello, G. B. Lamont, and D. A. Van Veldhuizen, *Evolutionary algorithms for solving multi-objective problems*. Springer: Genetic and Evolutionary Computation Series, 2nd ed., 2007.

[41] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, pp. 182–197, Apr 2002.

[42] D. M. Rose, J. Baumgarten, and T. Kurner, "Spatial traffic distributions for cellular networks with time varying usage intensities per land-use class," in *2014 IEEE 80th Vehicular Technology Conference (VTC2014-*

Fall), pp. 1–5, Sept 2014.

[43] S. Zhou, D. Lee, B. Leng, X. Zhou, H. Zhang, and Z. Niu, "On the spatial distribution of base stations and its relation to the traffic density in cellular networks," *IEEE Access*, vol. 3, pp. 998–1010, 2015.

[44] D. Lee, S. Zhou, X. Zhong, Z. Niu, X. Zhou, and H. Zhang, "Spatial modeling of the traffic density in cellular networks," *IEEE Wireless Communications*, vol. 21, pp. 80–88, February 2014.

[45] ITU-R, *M.2135: Guidelines for evaluation of radio interface technologies for IMT-Advanced*. ITU, 2008.

[46] D. González G. and J. Hämäläinen, "Topology and irregularity in cellular networks," in *2015 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1719–1724, March 2015.

[47] J. C. Spall, *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*, vol. 65. John Wiley & Sons, 2005.